BACnet® for Video Surveillance

By Zhou Ning, Student Member ASHRAE, **Zhou Man-li** and **Xu Yi-ping**

Integrating video surveillance and building automation systems can improve security, ensure timelier maintenance, and use the potentially vast amount of video information for other whole-building automation schemes. In the latter applications, the video data needs to be further processed and filtered before being presented to human operators or other computer systems. This processing may allow the monitored video scene, regardless of venue or time of day, to be analyzed to detect suspicious objects or movement and then generate an appropriate alarm. The analysis even can be extended to detect the presence of small amounts of fire or smoke in large spaces before the danger can be detected by traditional flame or smoke sensors.

Using BACnet® to integrate multiple types of building systems, such as HVAC, lighting, and fire detection, is well-known.[1] The protocol was designed to meet the needs of building automation and control applications. It conveys data including analog and binary values related to hardware and software, schedule information, alarm and event information, and files containing control logic or other information.[2] BACnet does not prescribe the internal configuration, data structures or control logic of a controller. Manufacturers can implement desired processing algorithms in their devices.

BACnet represents the information and functionality of any device by defining collections of related information called "objects," each of which has a set of properties that characterizes it. The object-oriented structure also provides a way to add new application functionality to BACnet by defining new objects. New application services also can be added.

'This processing may allow the monitored video scene…to be analyzed to detect suspicious objects or movement and then generate an appropriate alarm.'

To implement our BACnet-based video surveillance solution, we believe that new BACnet functionality is needed. We have proposed the addition of new object types to represent the features of video surveillance points and a new application service to provide a mechanism to start/stop video streaming and to control video camera preset positions. These proposals are under deliberation within SSPC 135, the ASHRAE BACnet Standing Standards Project Committee.

In this article, we present a study motivated by the advantages of processing video data at the camera level. To do this, we have engineered a working implementation of a BACnet-based camera, developed by augmenting a conventional camera with data processing capabilities. Through an alarm mechanism similar to that of the Life Safety Change algorithm in the BACnet standard, such a camera can convey a content-based description of the surveillance scene.

BACnet-Based Camera

Traditional video surveillance has two main requirements: the first is the ability to control the operation of the camera, and thus, the video "scene;" the second is the transmission, capture and interpretation of the image data. The first requirement is addressed by the operator's ability to pan, tilt, zoom (PTZ) and focus the camera. Depending on the application, this may be performed by an automatic scanning process or by slewing the camera to preset PTZ settings. In terms of image interpretation, the user may observe the scene in real time or store it for later processing.

In the case of early surveillance systems, all processing was done in the analog world — all the information was represented by the modulated intensity of the camera's output signal. Today, due to rapid progress in digital technology and the development of fast, economical analog-to-digital conversion, a new generation of digital cameras has appeared, opening the door to new and innovative applications.

Depending on the targeted application, the video data can be represented by a specific content-based description.[3] To avoid the transmission or storage of irrelevant information that may be present in the data, we perform the extraction of the content-based description at the acquisition stage, i.e., at the camera. Furthermore, the use of a content-based description of the data allows more flexibility in the design of complex and powerful surveillance systems as well as the potential for greater interoperability. This idea is also consistent with today's trend to transform "data" into "information" and "information" into "knowledge."

A block diagram of the BACnet-based camera is shown in *Figure 1*. The Image Analysis, Scene Description, Control and BACnet Coder/Decoder (CODEC) are required modules. The Video Coder only is required if the video data are to be transmitted digitally.

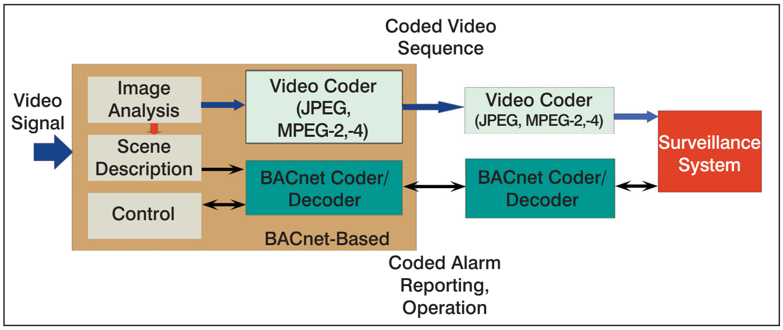


Figure 1: Structure of BACnet-based camera.

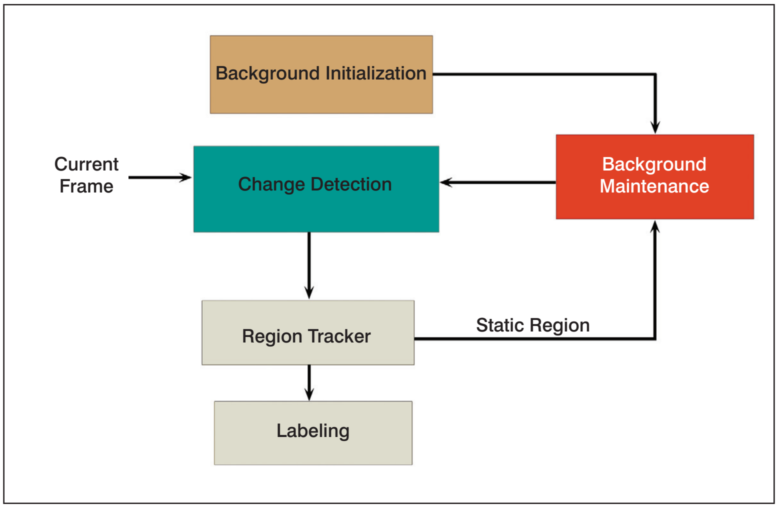


Figure 2: Image processing scheme.

Network-Visible Functionality

The functions that must be carried out using BACnet are camera control and control of the video transmission. The control module adjusts the camera's view of the monitored scene by panning, tilting, zooming or focusing. Pre-established PTZ settings also can be set by this module. We have chosen to represent the PTZ and focus settings as writable properties of a new BACnet object type, the "Video Point Object." To adjust the scene symmetrically up or down, left or right, the "pan" and "tilt" properties would range from –50 to +50. "Zoom" and "focus" would range from 0% (unzoomed/near) to 100% (fully zoomed/far). The current PTZ and focus coordinates of the camera can be read using the BACnet ReadProperty service.

One application is that of a large space with multiple access control sensors located such that a single camera must auto-scan back and forth to view all of possible entry points. When a sensor goes into alarm, the camera should slew to the correct PTZ coordinates to record what is happening. While the number of presets, the control logic and priority are outside the scope of the communication protocol, both the storage and activation of the PTZ coordinates for a given sensor location, or other point of interest, can be performed by a new BACnet application service that we call the "Video Point Operation" service.

The Video Coder module is optional. Although compressed video data uses less bandwidth and is more secure than an analog video signal, it may still represent excessive traffic on some of the control networks in common use today. While this situation is likely to change when more control networks make use of high-speed fiber technology, and are thus able to carry BACnet traffic and video streaming at the same time, we do not recommend embedding the video information in BACnet messages. The options are to carry the analog video signal on coaxial cable or to send the digital video data, appropriately compressed by the Video Coder module to remove redundant information, over a broadband LAN. In the case of digital transmission, the starting and stopping of the video streaming would be another function of the proposed Video Point Operation service.

Video Data Processing

The Image Analysis module consists of a set of algorithms for moving object detection and object tracking. These algorithms should only extract the features of the image that are relevant to the particular application. In surveillance systems such features as an object's position and trajectory must be considered.

Determining the position of moving objects in a video scene and then tracking them is challenging because physical objects are rarely homogeneous with respect to low-level features such as color, texture or optical flow.

In our work, we have been using an algorithm that performs automatic video object segmentation and scene tracking including the ability to deal with partial occlusions and interacting objects. "Video object segmentation" is the process of identifying and labeling the pixels (picture elements or dots) in the surveillance image associated with an object of interest. The algorithm's functionality can be subdivided into two main parts: the object-oriented segmentation and the object tracking.[4]

The purpose of object-oriented image segmentation is to distinguish objects of potential interest from the image background. To do so, the algorithm uses a change detection mask: the current frame is compared with the background frame. Pixels that change their value between these frames are considered to be part of an object of interest and appropriately labeled. Each closed region of the change detection mask is then given a unique number. Having a known background image is critical in surveillance applications.

Figure 2 shows the main image processing schemes.

Object tracking ensures that the same label is applied to a particular object over the entire sequence of frames in which it is present, detects and processes new objects, and detects disappearing objects. This is not as simple as it might seem. First, an object may grow, shrink or otherwise change its shape over time. In addition, the splitting of an object may form new regions in the object mask which must be differentiated from the case where a new object enters the scene. Finally, objects may "disappear" as they merge with other objects or are partially occluded. We have chosen a region-based tracker for the tracking function, the use of which will be shown later.

Scene Description

One of the greatest advantages of a BACnet-based camera is the ability to convey a description of the detected video objects via BACnet alarm and event services. Features of the image extracted by the image analysis module are passed to the scene description module. The manner in which the resulting descriptions are used depends on the application.

In the case of surveillance systems the most important information relates to an object's motion, including its position and trajectory. The most relevant low-level feature directly related to an object's physical properties, and is easily extractable from the video signals, is a moving object's position, describable in terms of its x and y coordinates. Other low-level features are object number and trajectory, which are also derived from the object's position. Additional high-level descriptions are used to uniquely label objects to provide some information about the object's past (whether it has been subject to merging or splitting) and to describe its perceptual motion. *Table 1* shows the descriptors that can be applied to moving objects in a surveillance image.

Descriptor	Purpose
Object Number	Identifies an Object
Activity States	Motion Description and History
Position	Specifies Position Using x, y Coordinates
Trajectory	Single Point Direction and Speed

Table 1: Descriptor set for a surveillance image.

The first three descriptors can be combined to form an object "label." The function of a label is not only to identify the object but also to provide some information about the object's past, i.e., whether it is the result of a merging with, or splitting of, other objects. When a new object appears, for the first time as well as after merging/splitting, it is assigned an object number which is the smallest number that has not already been used. The format of the object label is: L_i = {object-number, [appearing | disappearing | moving | stopped | out-of-scene | merging | splitting $\{L_j, L_k, ...\}$], (x-coordinate, y-coordinate)} where the square brackets enclose optional information, "|" indicates a choice of activity states, and the "$\{L_j, L_k, ... \}$" is present only when the object has been derived from the merging or splitting of other labeled objects.

These rules may result in rather long descriptions, particularly in the case of multiple merge/split cycles, but they allow the full reconstruction of each object's history and the merge and split detection part of the description can be omitted when searching a database of an object's history, which can be referred to simply by its object number.

The activity states provide a perceptual description of an object's motion and include such things as *appearing, disappearing, merging*, *splitting*, *stopped*, *moving*, and *out-of-scene*. The specific rules for determining the activity state may depend on the details of a particular scene, for example, information about door position or camera calibration, including view-field angle.

The trajectory descriptor is the set of moving object position coordinates in successive time slices. It can be used to compute an object's velocity and acceleration to predict subsequent motion. The standard BACnet Trend Log object type can be used as a buffer to monitor a property of a referenced object. We have chosen to use Trend Log objects to periodically store the position and activity states of the moving objects that we have detected. When a buffer in a Trend Log object is ready to be emptied, the camera sends a BACnet ConfirmedEventNotification message to the surveillance operator station. On the station side, a BACnet ReadRange service request is used to read the moving object's coordinate data and activity states from the Trend Log object.

Although the video image itself can be captured and stored when an incident occurs, scene descriptions can be used for further scene analysis to make a surveillance system more "intelligent." In a real-time distributed surveillance application, for example, a moving object's speed and direction can be computed from trajectory information. When such a moving object is detected, even several objects in different places at the same time, the number of objects and their direction may be used to decide which video image is most significant and to switch the surveillance personnel's monitor to that image. In addition to their use in real-time surveillance systems, the scene descriptions also allow video images to be quickly indexed according to the details of the scene description.

Alarm Mechanism

Whether or not a given scene or object activity state should generate an alarm depends on the particular application. The set of alarmable scene states we have defined include *quiet*, *suspicious-object-alarm*, *flame-alarm*, *fog-alarm*, and *abnormal-motion-alarm*. The *quiet* state denotes the unlikely situation wherein the camera input may have been "hijacked" by the insertion of a static image in which no moving object can be detected. We have also defined an algorithm that we call CHANGE_OF_VIDEO_SCENE, which parallels the CHANGE_OF_LIFE_SAFETY algorithm in the BACnet standard.[5] The algorithm produces one of three results: normal, off-normal, and video-scene-alarm where "normal" denotes a lack of detected motion or no change in any of the objects' activity states within the scene.

Whenever the activity state of a detected object changes, the CHANGE_OF_VIDEO_SCENE algorithm evaluates to "off-normal." The object activity state is obtained by extracting the moving object from the static background, tracking the object, and applying the analysis rules. For example, when a detected object cannot be found in the current object list, the object is considered to be *appearing*. Since this is a new activity state for a newly detected object, i.e., a change in the video scene, an off-normal event is generated. The *disappearing*, *merging*, *splitting*, *stopped*, *moving*, and *out-of-scene* activity states are similarly treated. In our image processing, a moving object occluded by other objects and a moving object leaving the scene should be distinguished. The off-normal event causes a notification to the operator that a changed condition exists and should be investigated.

When the changing of object states causes a scene state of *quiet*, *suspicious-object-alarm*, *flame-alarm*, *fog-alarm*, or *abnormal-motion-alarm*, the CHANGE_OF_VIDEO_SCENE algorithm generates a "video-scene-alarm." The scene state depends on the particular application. For example, a subway station surveillance system might detect someone jumping onto the tracks or a crowd disturbance, and produce an *abnormal-motion-alarm*. Particular scene knowledge, such as the physical layout of the subway station, may be used to facilitate the video analysis and the implementation may vary from one device manufacturer to the next. When a video-scene-alarm event is detected, the camera should notify the surveillance operator station by sending a BACnet ConfirmedEventNotification message. Once the operator verifies the alarm by whatever means, appropriate measures could be taken, such as activating an audible alarm or securing a particular access or egress route.

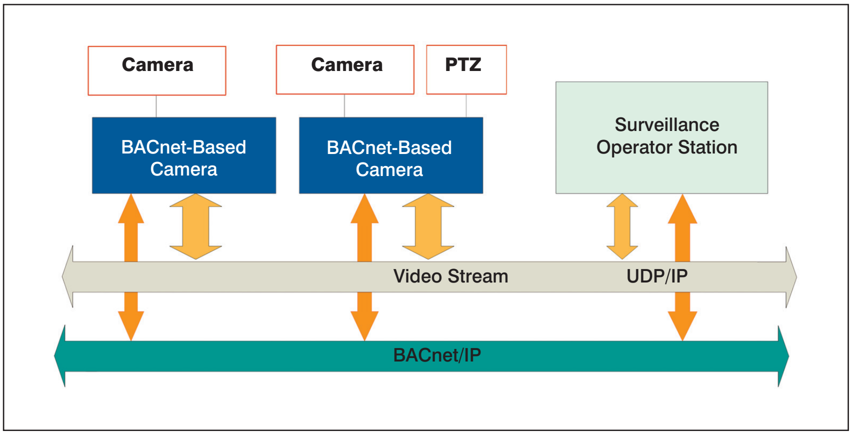


Figure 3: BACnet-based smart video surveillance.

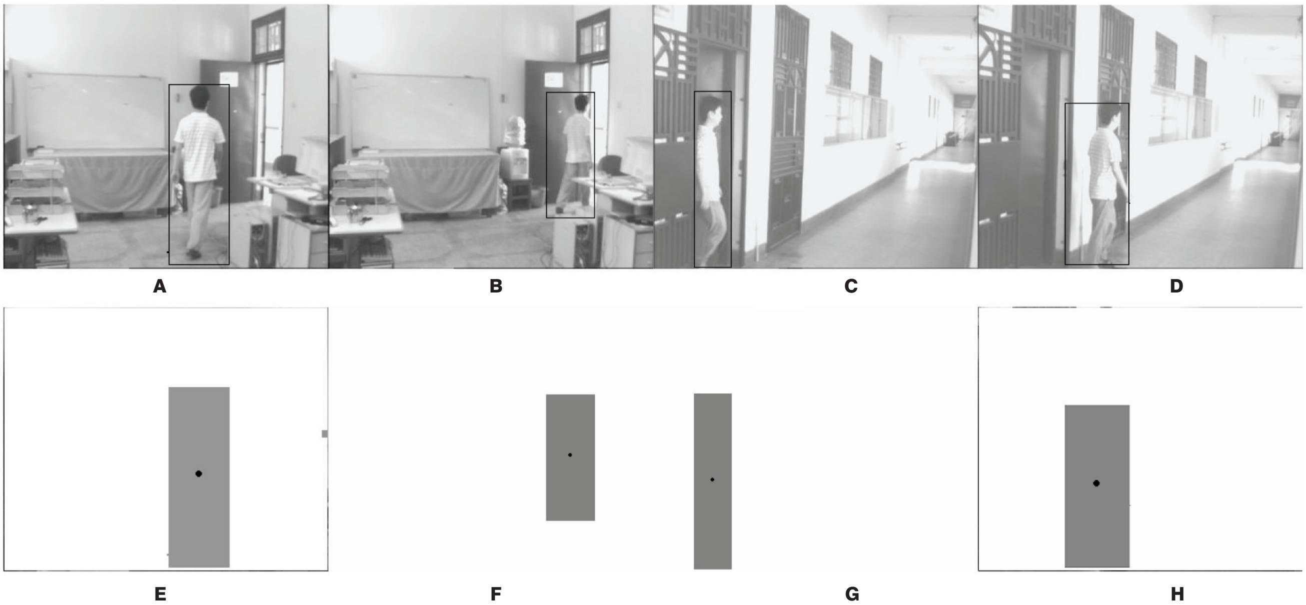
The Results

Taking advantage of the BACnet-camera described above, we developed a prototype BACnet-based video surveillance system. The system architecture is shown in *Figure 3*. In our prototype, one camera with PTZ control is in a room, a second camera without PTZ control is in the adjoining corridor. The compressed video data from the cameras is transmitted to the operator station via the User Datagram Protocol over IP. When either camera detects an abnormal condition, a video scene alarm is sent to the operator station using BACnet/IP.

Our goal in the prototype is to be able to detect and track a person who moves between the room and the corridor without losing the corresponding label. Only the motion of the objects is of interest.

To extract the positions of the moving objects, we implemented an efficient change detector[6] and tracking algorithm on a PC connected to the analog camera. The positions of the moving persons and their labels from the different cameras are then sent to the surveillance operator station. In the surveillance operator station, the virtual objects (represented in *Figure 4* by the dots) perfectly follow the movement of the persons, and the set of points representing the trajectory of the person is read from a BACnet Trend Log object in the camera controller by means of the BACnet ReadRange service.

In our experiment, the information on the door positions in the room and corridor background images is built into the surveillance scenarios. When a moving object in the room scene moves towards the door position and then disappears (see *Figure 4b*), the application on the surveillance operator station is programmed to index to the next surveillance point in the corridor scene related to the door. If an object appears at this door position in the corridor scene within a given time interval (see *Figure 4c*), the object in the corridor scene is considered to be the same object that just left the room scene and tracked without interruption.



A–B — Room Original Image; **C–D** — Corridor Original Image; **E–F** — Corresponding Room Virtual Object Image; **G–H** — Corresponding Corridor Virtual Image

Figure 4: Results of the experiment. Moving persons positions and their labels from the different cameras are sent to the surveillance operator station, where the virtual objects, represented by the dots, follow the movement of the persons.

Conclusions

We have proposed a new BACnet object (Video Point Object) and a new BACnet service (VideoPointOperation) to the ASHRAE BACnet committee and developed a BACnet-based camera prototype that potentially can be used in a number of distributed surveillance applications. One of the main advantages of such a camera, compared with a conventional camera, is its ability to extract and deliver some scene information in a digital format. Future enhancements include taking account of the details of an object's shape and making use of the content-based descriptions from the intelligent video surveillance system to assist the decision making process in security systems.

From our work, we believe that BACnet has the potential to be applied to a variety of problems beyond the simple monitoring and control of heating, ventilating and air-conditioning systems.

Acknowledgments

The authors would like to thank Steven T. Bushby, Member ASHRAE, and H. Michael Newman, Fellow ASHRAE, both past chairmen of ASHRAE SSPC 135, for their constructive comments and editorial help in the preparation of this article. We are also grateful for their encouragement.

References

1. Bushby, S.T. 1997. "BACnet™, a standard communication infrastructure for intelligent buildings," *Automation in Construction* (6):529–540.

2. Bushby, S.T. 2001. "Integrating fire alarm systems with building automation and control systems." *Fire Protection Engineering* (11):5–11.

3. Steiger, O., A. Cavallaro, T. Ebrahimi. 2002. "MPEG-7 description of generic video objects for scene reconstruction." *Proc. of SPIE Electronic Imaging 2002 - Visual Communications and Image Processing* 947–958, 21–23.

4. Tsaig, Y., A. Averbuch. 2002. "Automatic segmentation of moving objects in video sequences: a region labeling approach." *IEEE Transactions on Circuits and Systems for Video Technology* 12(7):597–612.

5. ANSI/ASHRAE Standard 135-2001, *BACnet - A Data Communication Protocol for Building Automation and Control Networks.* pp. 239–247.

6. Cavallaro, A., T. Ebrahimi. 2002. "Accurate video object segmentation through change detection." *Proc. of IEEE International Conference on Multimedia and Expo* pp. 445–448, 26–29.

Zhou Ning is a Ph.D. candidate at the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, P.R. China. Zhou Man-li is a professor at the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, P.R. China. Xu Yi-ping is a lecturer at the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, P.R. China.