Video Privacy in the Age of Automated Surveillance

Lisa M. Brown, Senem Velipasalar,* Andrew W. Senior, Sharath Pankanti,
Arun Hampapur, Chiao-fe Shu, Ying-li Tian

IBM T.J Watson Research Center
Yorktown Heights, NY 10598

*Princeton University
Princeton, NJ 08544

Abstract

Automated digital video surveillance has come of age and a new threat to personal privacy is posed. Traditional physical security coupled with information systems has slowly been eroding personal privacy. But the advent of automated video analytics and the latest integrated information systems will truly open Pandora's Box. The possibility that your employer will know when you arrive at work, when you loiter in front of your workplace, and when you leave with a package is already here. The possibility that your local retail merchant will know when you shop, with whom you shop, and what you do when you shop is around the corner. The potential for the misuse of this information is of increasing concern.

However, the same technology which can be used to compile personal information about you can also be used to protect your privacy. In this paper, we present several examples of automated surveillance techniques which can be used to ***limit*** *the information garnered from digital video surveillance. Exploiting user-defined alarm conditions which can be arbitrarily complex, we show how privacy can be preserved using vision technology in situations where privacy concerns might otherwise be overlooked. Specific examples are given at locations such as industrial entry points, parks and schools, and a retail environment.*

1. Introduction

Several efforts to ensure video privacy have been proposed. These include embedding video analytic capabilities on camera platforms and limiting the right to information based on access control lists[Senior 05]. Surveillance video can be selectively scrambled where the selection may be based on facial identity, body actions, or even location information. Considerable effort is underway to develop video encryption schemes to conceal identifying features and secure network protocols for safely delivering privacy sensitive video information to the appropriate hands [Ponte 05][Fialeao 04][Avidan 06].

These methods are predicated on the ability of the underlying computer vision techniques to detect, track and recognize human activity. While this technology is what makes automated surveillance possible, it also opens people and their actions to automated scrutiny, disregarding their privacy and exposing them to the potential misuse of their personal information. In this paper, we propose a methodology to protect public privacy by limiting information available to system users to relevant events thereby ignoring actions and events which might be over scrutinized, misused, or exploited. Unlike previous work in which all human identity is concealed or all human behavior is "blurred" to limit identifying characteristics, this methodology is designed to minimize the flow of extraneous information to people without the need to know, and optimize the flow of pertinent information to the appropriate personnel.

In particular, we propose a smart surveillance system which is based on complex alerts designed to detect specific behavioral infractions. This system is explained in detail in [Velipasalar 06]. In this paper we show how this methodology can be used to design access restrictions to ameliorate public concerns about privacy. In the next section we describe this methodology in more detail and discuss its capabilities and ramifications. In Section 3 we describe the composite spatio-temporal event detection approach and implementation. In Section 4 we show several examples of single and composite events which can be useful in improving the relevance of surveillance data and simultaneously limiting the access of private information. We end with conclusion in Section 5.

2. Limiting Access to Event Detection

Most surveillance systems today store their continuous video streams for all relevant cameras on DVR systems.

This information is backed up on tape and is stored for a week, month or longer. The live video streams are ostensibly watched on closed circuit television (CCTV) by security personnel. The American Civil Liberties Union (ACLU) describes five types of abuse related to CCTV: criminal, institutional, abuse for personal purposes, discriminatory targeting, and voyeurism.

A British study of video surveillance found that young male or black subjects were systematically and disproportionately targeted for “no obvious reason.” The study also found that some operators used video surveillance for voyeurism, and that no one used it to watch over those at risk, saying that operators don’t look out for possible victims, but focus on stereotypical categories of those they think might be likely offenders. Women were also more likely to be objects of voyeuristic rather than protective surveillance. The ACLU isn’t alone in concluding that video surveillance’s benefits and risks are disproportionate.

The advent of automated surveillance will eventually make the need for CCTV obsolete. Although this will alleviate abuses related to the use of CCTV, it will also introduce a new array of privacy concerns. Automated surveillance data will be indexed, catalogued and eventually mined.

In earlier work, we propose a layered approach to granting access to the different kinds of data that the system extracts. Depending on the authorization level, the interface might let the user access the original raw video or even video enhanced with additional information, or it might present only reconstructed video with details deliberately obscured. It might only give the user statistical information, such as the number of people in a space. Determining what information to give which users depends on the situation and the types of users, but we provide a set of tools and basic algorithms that handle the most common cases. For example, law enforcement officials can subpoena the original video, whereas security guards can see only (identity-obscured) rerendered video, unless they use an override button, which logs the time and the video footage appearing at that time. Other registered users might be able to access other information, and anonymous users can inquire about statistics. Additionally, a device might register as a user—for instance, an elevator control computer might be able to access how many people are standing in front of the elevator doors.

In this work, we extend this approach to include further divisions of data type. In the spectrum of information ranging from original raw video, video with details obscured and statistical information events, we add another useful category of video information. This is information regarding specific types of events, some of these event being complex high level behaviors which are significant for security issues.

To this end, we introduce a spatio-temporal event detection system which lets the users specify multiple composite events of high-complexity, and then detects their occurrence automatically. This will enable the detection of events of interest such as a person abandoning an object and then leaving the area, someone entering a building from a side entrance and then entering a high security region or someone taking an object and then driving past a gate.

Events can be defined on a single camera view or across multiple camera views. Simple or primitive events are used as elements to build semantically higher level event scenarios. Primitive events are combined using operators. Finally, a hierarchical scheme allows events at one level to be composed of subevents at the level below. In this way, highly complex events composed of subsequences and combinations of simpler events can be created.

3. Spatio-Temporal Event Detection Method

The spatio-temporal event detection system is composed of a user interface based on an underlying event definition language implemented via XML files. These files are passed to the smart surveillance engines which detect the primitive events. The interface is shown in Figure 1. The user selects the camera and defines the primitive events and their parameters to be detected in this view. At the bottom of the interface, the user combines the primitive events which may occur in different cameras selecting from a set of operators.

This composite event detection system is designed to be general, flexible and extensible. The combination of primitive events in a layered scheme allows the user to define an open-ended set of behaviors and actions whose complexity is unlimited.

Currently, there are six primitive events which can be detected by the system. The parameters of these events can be specified in the user interface. These primitives are: motion detection, directional motion, abandoned object, object removal, and trip wire. Additional feature attributes of track data can be constrained such as color, duration and size. As the system matures and composite event detection is realized in real applications, additional primitive events can be added to the system

There are five operators available in the current system: SEQUENCE, AND, OR, REPEAT-UNTIL and WHILE-DO. These operators specify the relationship needed between the primitive events to trigger the composite event. The most useful operator is SEQUENCE which specifies that one event must follow another event by a given minimum and maximum time period. The AND operator is useful for combining primitives which can occur in any order. Again as the system matures, additional operators can be added to enhance the system and fulfill the needs of specific domain problems. The system is particularly adept at allowing the user to visualize the composite event and build the composite event which satisfies both detecting the behavior of interest and leveraging the capabilities of the system which allow it to do so.

Finally, after the user has defined the composite event across cameras of interest, the system needs to detect the occurrence of this event. This involves passing the primitive events defined on the same camera view to the appropriate tracking engine. The XML files are parsed by the **Composite Event Manager** which then performs the transfer of this information to a new set of XML files which are sent to the corresponding tracking engines.

4. Examples

Three examples are given of complex composite events which could be used to improve security and limit the use of irrelevant video information. The first two examples involve tailgating, the first at a secure access door entry and the second at a vehicle gate entry. In both cases, entry monitoring of all events is not necessary for good security. The primary events of interest involve occurrences of tailgating. Enabling the detection of tailgating allows the security personnel to focus their efforts on significant security issues. At the same time, it can be used to limit their access to information which is not important to their duties.

In the first example, we show how the detection of tailgating at the indoor entry point is defined and detected. Figure 1 shows the user interface used to define the tailgating event interest. The event is defined as a sequence of four primitive events. First, the access to the badge reader is detected via motion detection. Then two events of a person entering the door detected by tripwires are defined. Finally, the door closing *after* two entries is detected using the abandoned object alarm which detects that the door is now stationary back at its original position. The bottom of the door is used because it is the nontransparent portion of the door. Figure 2 shows the detection of each primitive event in sequence, which then triggers the tailgating event alarm.

In the second example we show how the detection of tailgating at a vehicle gate is defined and detected. Figure 3 shows the composite event definition in the interface to detect vehicle tailgating. In this example, there are three primitive events which are detected in sequence. First, the gate opening is detected by motion detection at the closed-gate location. Second, two vehicles are detected in the region in front of the gate. Lastly, the closing of the gate is detected using motion detection at the open-gate location. Figure 4 shows an example of an instance of tailgating detected by the system.

The third example (Figure 5) is based on a retail scenario with multiple cameras In this case, a person removes an object from the store, is then followed through the store and then exits without paying. The event is defined as a sequence of three primitive events: object removal, directional motion and tripwire crossing. Our current implementation of object removal requires the user to specify the location of the object of interest. This would not be realistic in a retail environment. Also, the individual is followed based on a typical pattern of exiting from this location – this would need to be more general in the real case. However, the example is shown here to exemplify the types of events which could be detected using the composite spatio-temporal event detection system. Furthermore, we would like the reader to envision a system which detects shoplifting and other inappropriate behavior but leaves the normal shopping activity undetected.

Our last example (Figure 6) is based on the use of primitive event alarms which could be useful by themselves in restricting information to relevant scenarios. In this example, two primitive alarms (fighting and loitering) are detected on school grounds. The fighting alarm is not yet implemented in our system and is shown here as an example. The second alarm, loitering is based on track duration. This alarm also uses the system capability of constraining an alarm to specified days and a given time period. In this case, the alarm is only triggered before and after dusk when students should not be loitering on the campus.

5. Conclusions

In an age of automated surveillance, the potential misuse of information for criminal, institutional, personal, and discriminatory abuse is of growing concern. While the use of this technology promises to improve safety and security, it is important for its use to be appropriately limited to serve its purpose. The latest capabilities in complex event detection can be used to both improve the utility of surveillance data and to improve the protection of privacy.

We have introduced a novel multi-camera spatio-temporal event detection system to enable system users to define and detect complex hierarchical events of interest. We have shown several examples of pertinent events which can be detected in industrial, retail and school environments. The ability to detect specific events of interest will play an important role in changing the duties of security personnel and the types of problems they can address. At the same time, it is important to also use this ability to limit the access of information of more routine events to appropriate officials.

6. References

[Hampapur 05] A. Hampapur, L. Brown, J. Connell, Max Lu, Hans Merkl, S. Pankanti, A.W. Senior, Chiao-fe Shu, and Y.-L. Tian, “Multi-scale Tracking for Smart Video Surveillance,” IEEE Transactions on Signal Processing, vol. 22, No. 2, March 2005.

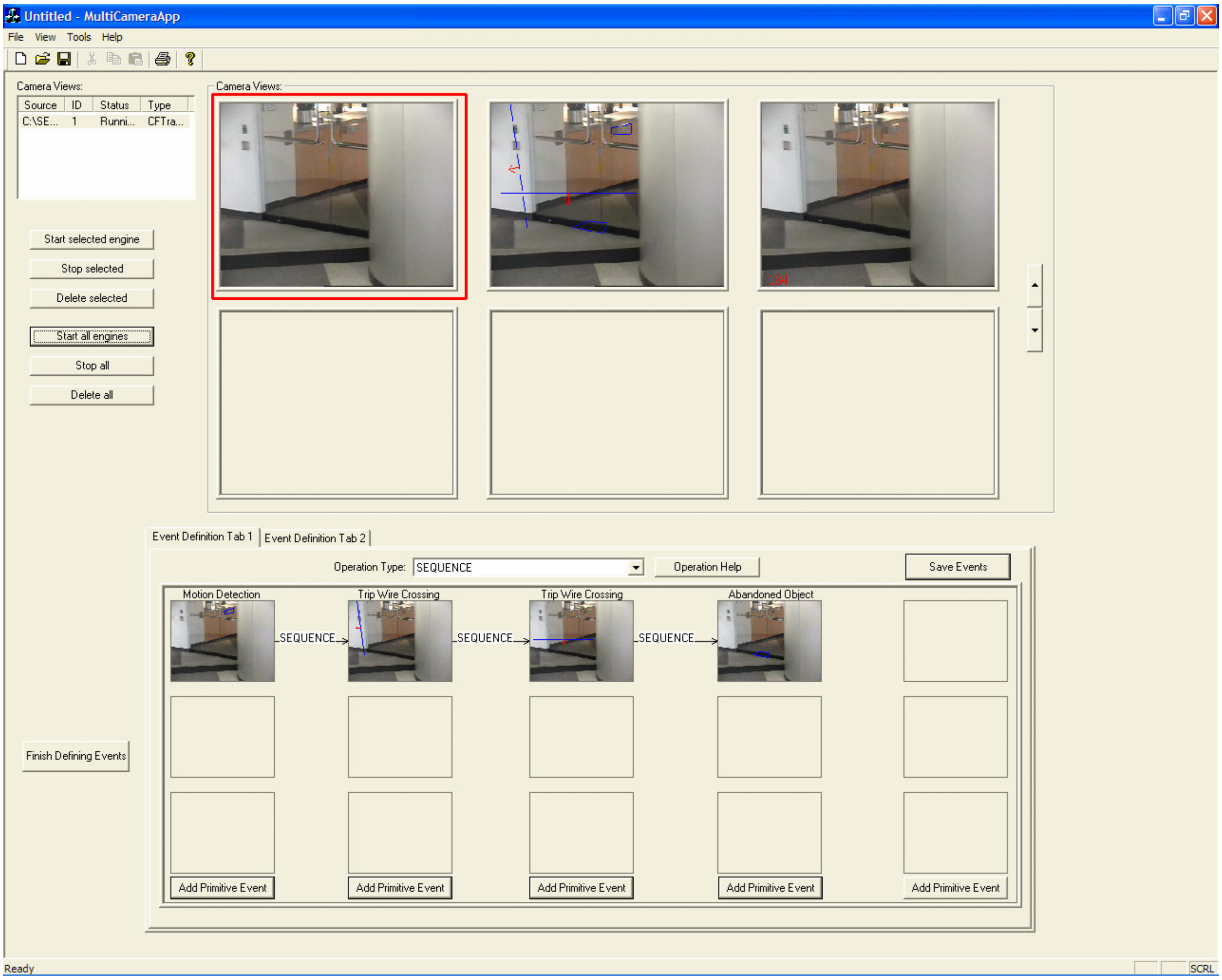
[Senior 05] A. W. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, J. Connell, Chiao-fe Shu, and Max Lu, “Enabling Video Privacy through Computer Vision,” IEEE Security & Privacy – Infrastructure Security, vol. 3, No. 3, May/June 2005.

[Ponte 05] I. Martinex et al., “Robust Human Face Hiding Ensuring Privacy,” Proc. of Int’l Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Montreux, Switzerland, April 2005.

[Fidaleo 04] D. A. Fidaleo et al. “The networked sensor tapestry (NeST): a privacy enhanced software architecture for interactive análisis of data in video-sensor networks,” Proc. of the ACM 2nd Int’l Workshop on Video Surveillance & Sensor Networks, New York, NY, 2004.

[Avidan 06] S. Avidan and B. Moshe, “Blind Vision,” Proc. of the 9th European Conf. on Computer Vision, Graz, Austria, May 2006.

[Velipasalar 06] S. Velipasalar et al., “Specifying, Interpreting and Detecting High-level, Spatio-Temporal Composite Events in Single and Multi-Camera Systems,” IEEE Workshop on Semantic Learning Applications in Multimedia, New York, NY, June 2006.



Source	ID	Status	Type
C:\SE...	1	Runni...	CFTra...

Operation Type: SEQUENCE

Figure 1. User Interface showing composite event detection of tailgating at door with badge reader access at industrial entry point

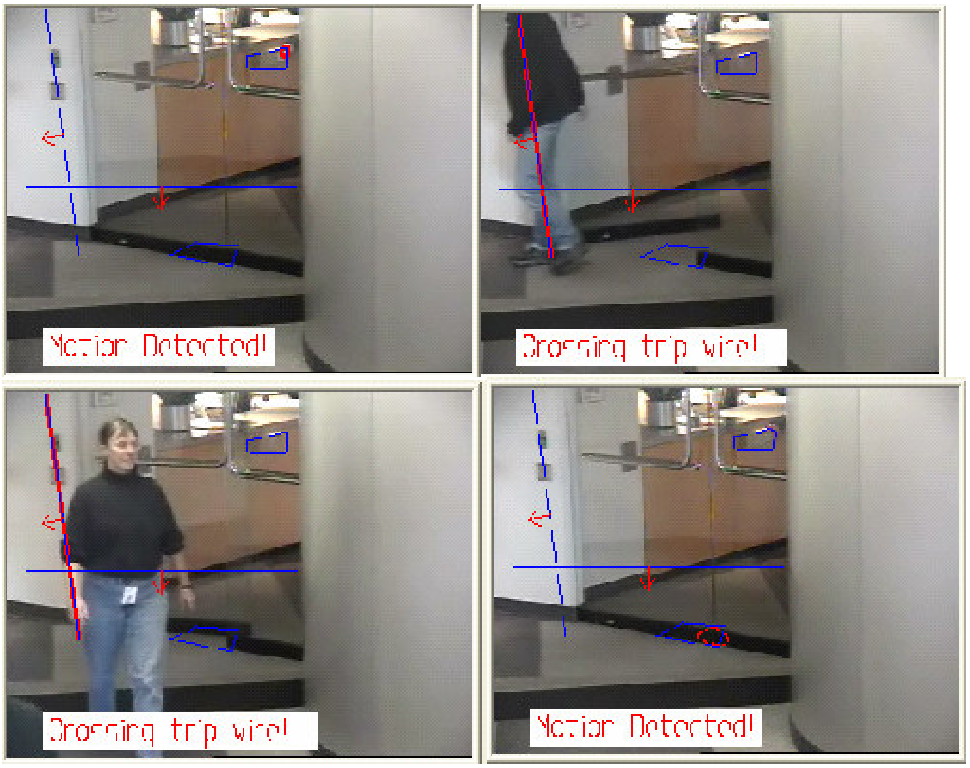
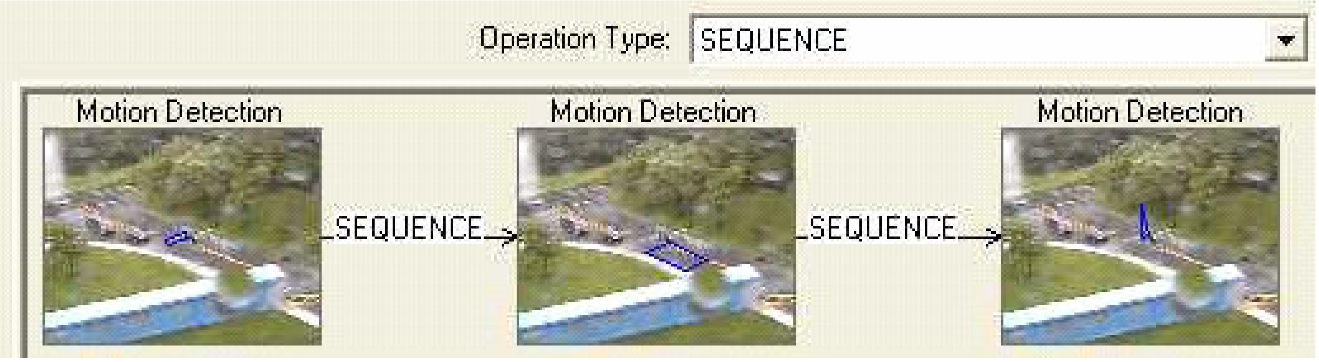


Figure 2. Detection of four stages of tailgating at door: (upper left) badge reader access – hand at badge reader in upper right of this image (upper right) first person enters (bottom left) second person enters (bottom right) door closes –black portion of door at original location at bottom of image.



Operation Type: SEQUENCE

Figure 3. Composite event detection of tailgating at industrial vehicle gate entry

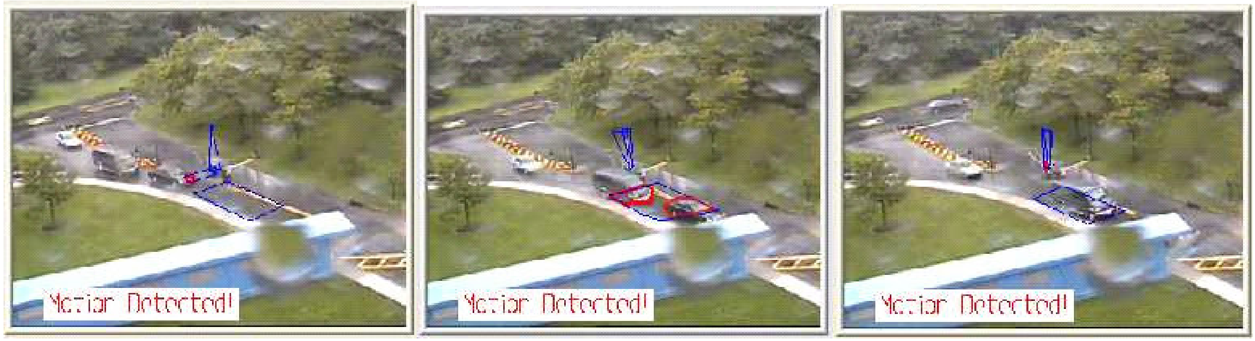


Figure 4. Detection of three stages of vehicle tailgating: (left) gate opens (center) two cars detected in region in front of gate (right) gate closes.

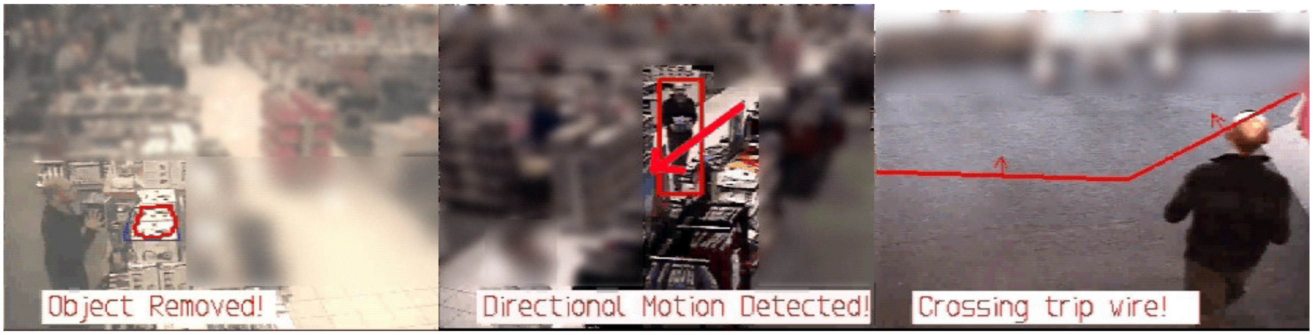


Figure 5. Multi-camera composite event detection of shoplifting event at retail environment

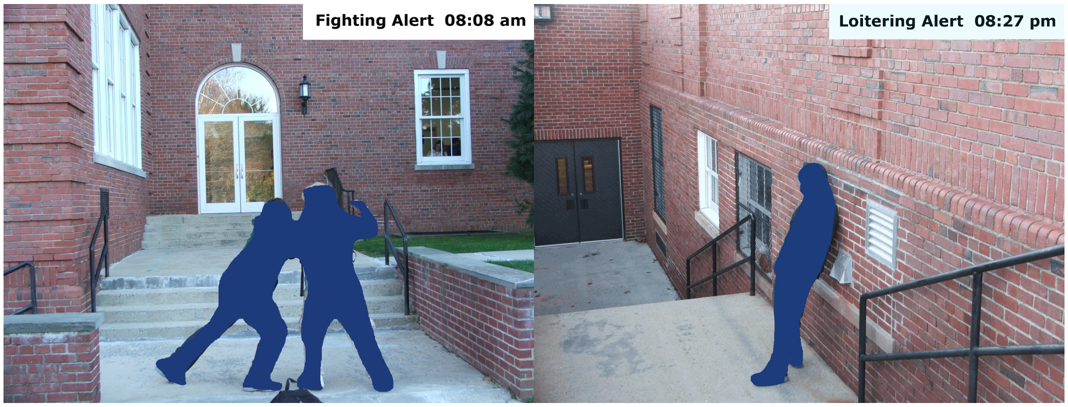


Figure 6. Detecting improper behaviors at a school environment such as fighting or loitering on school grounds after dusk.