Interface Components for Monitoring Security Video

Andreas Girgensohn[1], Frank Shipman[2], Thea Turner[1], Lynn Wilcox[1]

[1]FX Palo Alto Laboratory, Inc.
3400 Hillview Avenue
Palo Alto, CA 94304, USA

[2]Center for the Study of Digital Libraries
Texas A&M University
College Station, TX 77843-3112, USA

{andreasg, turner, wilcox}@fxpal.com, shipman@cs.tamu.edu

ABSTRACT
Video surveillance requires keeping the human in the loop. Software can aid security personnel in monitoring and using video. We have developed a set of interface components designed to locate and follow important activity within security video. By recognizing and visualizing localized activity, presenting overviews of activity over time, and temporally and geographically contextualizing video playback, we aim to support security personnel in making use of the growing quantity of security video.

ACM Classification: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – video.

General Terms: Algorithms, Design, Human Factors.

Keywords: Video surveillance, video analysis, multicamera playback, security cameras, video interfaces.

INTRODUCTION

The decreasing cost of installing video surveillance systems has led to increasing numbers of video streams per installation and their use in contexts with limited personnel. Much of the current research in video surveillance focuses on algorithms to analyze video and other media from multiple sources to automatically detect significant events [3]. However, automatic algorithms do not always correctly identify events so that keeping the human in the loop is crucial.

We have developed a set of interface components designed to improve the ability of security personnel to locate and follow important activity within security video. The components include the recognition and visualization of localized activity in a video feed and its presentation as a single still image. We have also developed event timelines and storyboards to present activity over a number of cameras over a period of time. Finally, we have developed a multicamera video viewer including a timeline and map to provide temporal and geographic context for the video being shown.

RECOGNIZING AND VISUALIZING ACTIVITY

Advanced techniques for analyzing video are generally computationally expensive. Our requirement is that the analysis can be performed in real-time for tens of video cameras on just a few computers. For determining activity, we use a standard foreground-background separation [2]. The background is a continuously computed average of pixel values in a time window to cope with changing lighting conditions. Relative importance of video segments is based on the amount of activity or on activity near hotspots.

In many instances, it is not possible to watch all activities. Thus, there needs to be less time-consuming presentations of activity. We visualize segments with activity as static images using overlays and time-lapsed images based on the region of activity. Figure 1 (left) shows a form of keyframe overlay visualizing the results of object tracking. The trajectories of moving objects are shown as a series of points taken at regular time intervals. This approach indicates motion without creating too much visual clutter. An alternative approach uses time-lapse presentations from the actual video (see Figure 1 right). This approach relies on the separation of foreground from background and overlays foreground objects among different frames onto the computed background. Ryall et al. [4] presented a simpler approach that alpha-blends different frames instead of performing a foreground-background separation first. Sample rates of 0.5 to 2 seconds generate visualizations with appropriate continuity without too much blurring. This time-lapse image can be enhanced by emphasizing foreground pixels periodically.

LONGER TERM EVENT SUMMARIES

Once local periods with activity are determined, longer-term visualization of activities is needed. We use a variation of Manga video summaries [1] that present the relative importance of video segments using a comic strip layout with varying image size. Figure 2 shows this approach applied to security video. Using the activity segments identified as described above, the storyboard includes keyframes for each period of activity. To create a storyboard for multiple synchronized video streams, the algorithm groups segments of simultaneous activity across video streams. For each group, the keyframes are placed in a pile such that the most impor-

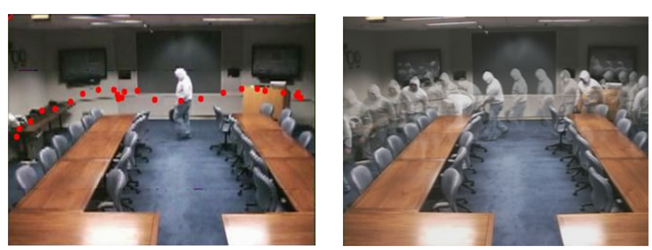

Figure 1: Trajectory of tracked object shown by marking positions at regular intervals and as overlays.

UIST'06, October 15–18, 2006, Montreux, Switzerland.

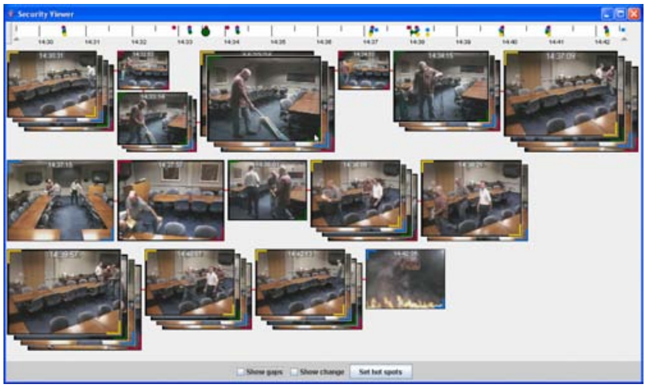

Figure 2: Storyboard of activity in multiple cameras.

tant keyframe is completely visible and other keyframes can be accessed by mouse movement. Storyboards cannot provide much detail when used with many cameras or over long periods of time. Hence, the map and timeline interface components allow the selection of the cameras and time period to be included.

MULTI-STREAM VIDEO PLAYER

To support the playback of video, we developed three elements: a multi-stream video player, a multi-scale timeline, and a map of the camera positions. The multi-stream video player presents video displays at different resolutions and frame rates. The multi-scale timeline permits easy navigation through recorded video or real-time video (see Figure 3). The player provides modes for automatic walkthroughs and manual or activity-based camera selection.

The left side of the player interface shows a traditional security camera interface at low frame rates. The main player area displays one or more video streams at higher frame rates and resolutions. The size of a video stream display indicates its relative importance. Skipping to a different position in the timeline will synchronously move all video displays to the same playback position.

The timeline provides access to the recorded video and lets the user switch back to live video. A non-linear scale transitions between a detailed linear scale for the video near the current playback position and a coarse linear scale for the video far from the playback position. Color indicates the density of the timeline. This timeline provides access to hours or days of video with second accuracy around the playback position. Because the current playback position does not stay under the mouse after a click, the change of the timeline from the old to the new position is animated and the new position is highlighted during the animation.

A map interface component provides security personnel with the location and field of view of each camera. Cameras being shown in the main player are color coded. Since many regions of buildings look similar, this is important for determining the location being shown and the geographic relations between different video feeds. The map is also used for selecting video streams to include in the storyboard or the video player. Clicking on a map position selects all the

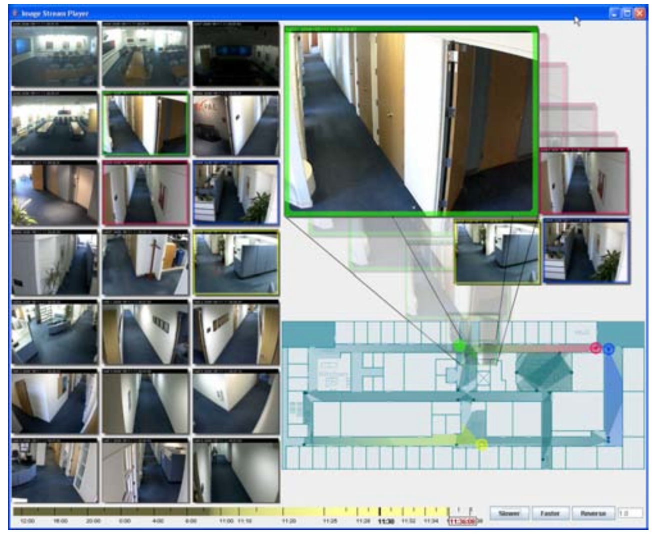

Figure 3: Video player with map and timeline.

cameras that can see that location. When video streams are selected for playback, the video stream displays are animated from the map position to the video player area (shown as time-lapse in Figure 3).

CONCLUSIONS

Video surveillance requires keeping the human in the loop. Interfaces are needed that make the overwhelming quantity of video more meaningful and direct the attention of security personnel to important video content. We provide activity-highlighting video summaries in the form of enhanced keyframes, timelines and storyboards to give users quick access to interesting events in recorded video. For live video, we automatically draw the users attention to video streams with activity by enlarging them and animating them into the center of view. We support users in seamlessly switching between live and recorded video and in synchronizing the playback of many video streams. For both live and recorded video, we connect the video streams to a map of camera locations for better orientation.

REFERENCES

1. J. Boreczky, A. Girgensohn, G. Golovchinsky, S. Uchihashi. An Interactive Comic Book Presentation for Exploring Video. *CHI 2000 Conference Proc.*, ACM Press, 185-192, 2000.
2. S.-C. Cheung and C. Kamath. Robust techniques for background subtraction in urban traffic video. Video Communications and Image Processing, SPIE Electronic Imaging, Volume 5308, pp. 881-892, 2004.
3. R. Cucchiara. Multimedia surveillance systems. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 3-10, 2005.
4. K. Ryall, Q. Li, A. Esenther. Temporal Magic Lens: Combined Spatial and Temporal Query and Presentation, Human-Computer Interaction INTERACT ’05, pp. 809-822, 2005.