Effects of Presenting Geographic Context on Tracking Activity between Cameras

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]Department of Computer Science &
Center for the Study of Digital Libraries
Texas A&M University
College Station, TX 77843-3112

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

ABSTRACT

A common video surveillance task is to keep track of people moving around the space being monitored. It is often difficult to track activity between cameras because locations such as hallways in office buildings can look quite similar and do not indicate the spatial proximity of the cameras. We describe a spatial video player that orients nearby video feeds with the field of view of the main playing video to aid in tracking between cameras. This is compared with the traditional bank of cameras with and without interactive maps for identifying and selecting cameras. We additionally explore the value of static and rotating maps for tracking activity between cameras. The study results show that both the spatial video player and the map improve user performance when compared to the camera-bank interface. Also, subjects change cameras more often with the spatial player than either the camera bank or the map, when available.

Author Keywords

Video surveillance, activity tracking, multiple video streams, security cameras, geographic context.

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems – *video*. I.4.8 [Image Processing and Computer Vision]: Scene Analysis – *tracking*.

INTRODUCTION

Video surveillance systems are common in commercial, industrial, and residential environments. A common surveillance activity is to keep track of people as they move from camera to camera; particularly tracking important people or people exhibiting suspicious behavior. With the decreasing cost of video hardware, the number of video streams per installation is increasing [9]. However, the limits of human attention and the number of video streams constrain the cost efficiency and effectiveness of such systems. Consequently, security personnel have great difficulty tracking activity across video streams.

One solution could be to automatically track activity or objects moving between camera views. While significant advances in automatic tracking have occurred, approaches that run in real time for dozens of cameras with reasonable computational resources are highly heuristic. Given this, the most effective and realistic solution is for human and system to work synergistically. Thus, interfaces are needed that aid people in tracking activity among cameras. One method of helping people track activity is to provide geographic cues of the relative positions of the cameras to make it easier to keep the activity in view by switching cameras.

In most surveillance installations, people can move directly from one camera view to a limited number of other camera views. These potential destination cameras tend to be geographically near the original camera. Presenting activity from nearby cameras as keyframes or video streams can facilitate the tracking of people as they move within the monitored space.

The Spatial Multi-Video (SMV) player is an interface for browsing and watching video from several cameras simultaneously while conveying spatial proximity. Rather than displaying all available camera views, only views in close proximity are shown. More importantly, multiple smaller players surrounding the central main player are placed around the video player such that a person walking out of the field of view of one camera will likely appear in the camera view adjacent to the direction that they walked out. In addition to the geographic cues from the video stream layout and animation, the SMV player includes a map depicting the locations and view directions of cameras.

A user study evaluated the relative value of different techniques for providing geographic context for tracking activity and to determine which user interface elements best support users in keeping track of a person walking from camera to camera. We compared the SMV player to a traditional multi-channel video player with a bank of small video displays and a single large video display. Both players were tested with and without a map showing camera positions and view angles. For the SMV player, we also tested a map that rotated to line up with the selected camera view.

The next section examines prior work and highlights new innovations in the SMV player and the use of geographic cues. This is followed by descriptions of the design of the SMV player and of the study and its results. The paper concludes with a discussion of the results and their implications for the next round of design for the SMV player and other interfaces that aid tracking activity among cameras.

Figure 1. Video player surrounded by cameras in spatial proximity. The location of each smaller video approximately indicates the direction of that camera. The map on the right shows camera locations and parts of the camera views.

RELATED WORK

Currently the surveillance security industry is undergoing a major shift from CCTV technology to digital video systems. The move to digital video makes camera output available for advanced processing, provides random access availability, and enables more flexible and advanced user interfaces. However, as systems and the activities they support become more complex, so does the user interface. The focus of this paper is to explore user interface options for the surveillance task of tracking people across cameras. While research system can perform this task automatically in controlled conditions, they are not reliable enough to take over from the user. The user interface designs investigated in this paper aim to support security personnel in situations where automatic approaches do not provide sufficient performance.

There are a large number of commercial systems for video surveillance, and a growing number of them offer some form of intelligent video solutions. Security industry analyst Freeman identified almost 40 commercial systems that involve some form of video analytics [9], a number that is growing quickly as demand for these systems grows. Some systems are under development by companies who specialize in public safety and military solutions [10, 13] or companies that have started with facilities management and engineering [8]. Others have emerged from the technology sector [1, 17, 20]. Some systems are beginning to include some level of geographic context [15, 21].

A primary focus of surveillance research has been the automatic tracking of activity across video surveillance cameras [3, 6, 23, 24]. Such mechanisms could be used to improve the selection of cameras in the SMV player but the current state-of-the-art does not remove the need for human involvement in the tracking process.

Work on interfaces for security video has emphasized the post-event task of video forensics [11]. Such systems have incorporated interactive visualizations of video content [7], the use of timelines [4], and virtual environments [19]. In comparison, there is limited research into user interfaces that aid real-time tracking. Iannizzotto and colleagues report on the use of gestures to control video selection and playback in a surveillance interface including a camera bank and a map [12]. Wang et al. describe techniques for identifying and visualizing the salient aspects of a video feed [22]. Most of the research results are in the area of system design and instantiation. There is a need for evaluating the success of these interfaces.

Considering video interfaces related to the SMV player more generally, many systems provide keyframes for navigation. The Rframes system [2] provides a list of keyframes that a user can scroll through to find a segment of interest. Clicking on a keyframe starts video playback at the selected segment. While these systems use keyframes as navigation aids, they do not synchronize the display of multiple video streams from several cameras.

We are not aware of other video players that arrange video displays spatially in an abstract fashion, i.e., other than placing keyframes or video displays on a map. The selection of geographically proximal cameras to the main camera with consideration of the fields-of-view of those cameras and the arrangement of video displays by camera location are novel aspects of our system.

THE SPATIAL MULTI-VIDEO PLAYER

Security video is normally observed and interacted with via a camera bank that shows available cameras. A player may be provided that presents a larger view for a selected video stream. In such a situation, the security guard must select

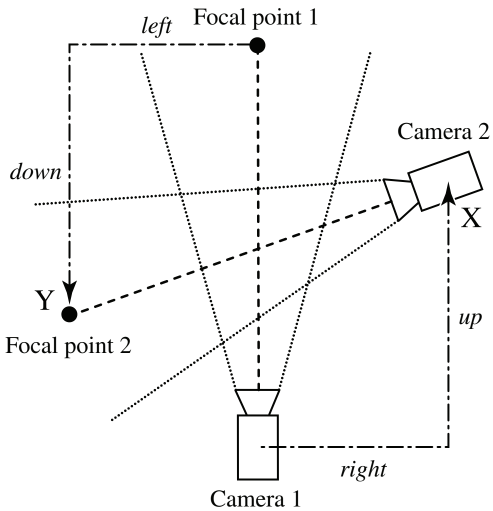

Figure 2. Camera Positions.

cameras from the camera bank to track activity from one view to another. Even if the views in the camera bank are grouped by spatial proximity, it is difficult for users to predict in which camera view a tracked person might appear after walking out of the main camera view. With many cameras, the images in the camera bank tend to be small so that it is difficult for the users to locate and recognize a tracked person in those images.

To address this problem, the Spatial Multi-Video (SMV) player selects and organizes its contents primarily based on geographic relations between the playing camera and those of the other camera views. Rather than displaying all available camera views, only views in close proximity are shown. More importantly, multiple smaller players surround the central player such that a person walking out of the field of view of one camera will likely appear in the camera view adjacent to the direction that they walked out. While the player does not automatically switch views, it makes it easier for the user to decide which of the surrounding views to select by clicking with the mouse on it.

Figure 1 depicts the view of a camera and the arrangement of nearby camera views that can be reached by walking down either of two hallways. The map in Figure 1 illustrates where those nearby cameras are located and shows partial views from those cameras. The person at the left of the main camera view in Figure 1 (hall13) already has one foot visible in the camera view to her left (hall15) and then could turn right to appear in the view in the top left (hall14) or continue in a straight line to appear in the center left (hall16). The other person in the main view is already partially visible in the distance of the view in the top right (hall08). The remaining view (hall17) is placed in the correct direction but it could only be reached by walking through two interim camera views because of walls blocking a more direct path. Figure 4 shows the complete configuration of the SMV player partially depicted in Figure 1.

Camera View Arrangement

Our layout algorithm arranges the peripheral camera views based on camera locations and view directions and angles.

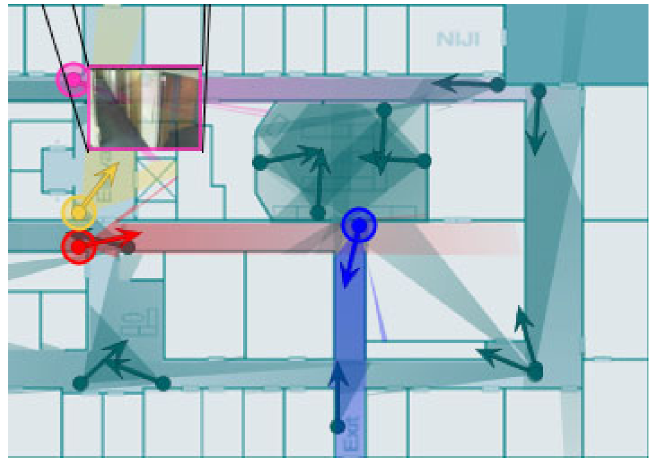

Figure 3. Map showing camera positions and directions. Selected cameras are visualized by fading keyframes.

Naively, one could just place the cameras on a map and determine the angle from the selected camera to another camera and place that camera view in that direction. However, this does not take into consideration that cameras show activity some distance away from the camera position. Instead of using the angle between camera positions on the map, we use the angle between their focal points. Figure 2 illustrates how the use of focal points can lead to very different camera view placements. When using just camera positions, Camera 2 would be placed to the right and up from Camera 1 (X). With the use of focal points, Camera 2 is placed to the left and down from Camera 1 (Y). The latter placement is correct because a person walking out of the left side of the field of view of Camera 1 will still be visible in Camera 2. We use focal lengths between 10 and 20 feet such that the relative size of an object at a focal point remains constant for cameras with different view angles. Eqn. 1 determines distances d_1 and d_2 between cameras and focal points with respect to the camera view angles of α_1 and α_2, respectively.

$$d_1 \tan\frac{\alpha_1}{2} = d_2 \tan\frac{\alpha_2}{2} \qquad \text{Eqn. 1}$$

Cameras closest to the selected camera are included in the display. We found that using a Euclidian distance measure does not properly reflect the situation in an office building with respect to how long it takes to walk from one camera to another. Instead, a city block distance measure (sum of horizontal and vertical distances) produces better results.

Ideally, the angle from the center of the player to the center of a camera view would be the same as the angle between the views of those two cameras. To fit more camera views, the angles are adjusted up to a certain amount by the SMV player. A least square measure of the angle errors can determine the best camera view placement. If a camera view cannot be placed within a certain error (e.g., 45 degrees), the number of displayed camera views needs to be reduced to maintain the spatial orientation.

Map with Camera Locations

We also display a map adjacent to the player (see Figure 3). The selected cameras are color-coded on the map. The same

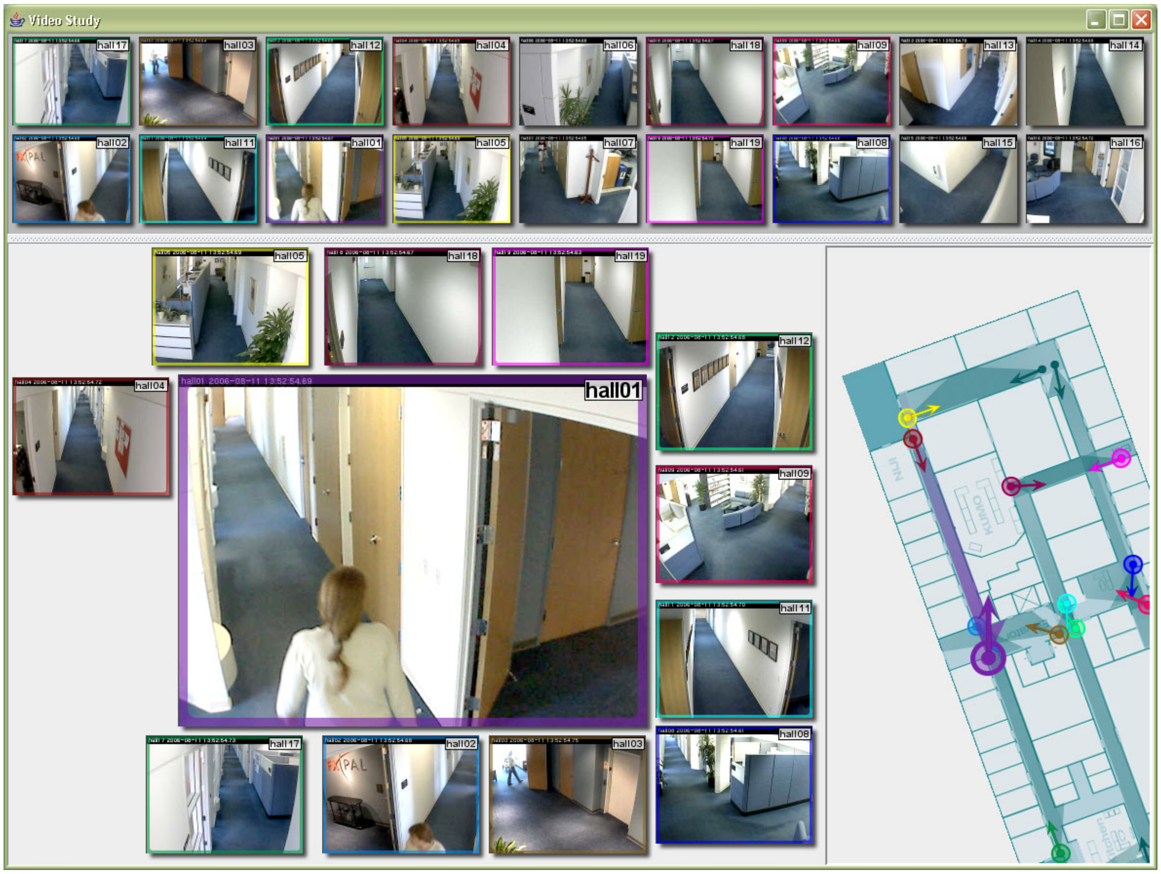

Figure 4. Complete SMV player interface.

colors are used as the borders of the camera views to tie the two displays together. A camera's location is shown as a dot on the map while the camera's view-direction is indicated by an arrow. The fields of view of the cameras are shown as shaded areas. We added the arrows after observing pilot study participants who could not reliably determine camera view directions just from the shaded areas. The dot and arrow of the main camera is enlarged and its field-of-view emphasized to indicate the location of the current main camera on the map.

Interaction with the Camera View Display

Figure 4 shows the complete SMV player interface with a camera bank at the top, the player area in the lower left, and the rotating map in the lower right. Normally, the camera bank would not be needed for a tracking task because the tracked person should appear in one of the camera views surrounding the main view. However, if the tracked person completely disappears from view, either by entering an area not covered by a camera or because the user selected the wrong camera, the camera bank might be the only means to reacquire the tracked person.

Users may click on any of the displayed video streams to select a new camera view for the main player. Using this technique, users can follow activity from camera view to camera view. Users may also select cameras using the map or the camera bank.

When changing the camera view, the movement of camera views to their new positions is animated to keep users oriented. Rather than animating camera views along straight lines, they are animated around the perimeter of the main view to indicate the view rotation. Camera views are predominately rotated in the same direction (clock-wise or counter-clock-wise) to indicate the view rotation. The animation duration is dependent on the distance camera views have to travel to avoid very fast or slow movement. On occasion, a camera view might be rotated in the opposite direction from the majority of camera views to reduce the travel distance; forcing all cameras to move in the same direction resulted in cameras moving too quickly or the animation taking too long in an earlier prototype.

When using a map in combination with the spatial display (see Figure 5), the map is centered to the newly selected camera and rotated such that the camera's view direction is to the top. The rotating map is intended to keep the user oriented and to simulate the rotation of a paper map as in the situation of a hiker rotating a trail map to line up with a trail at a junction. The map movement is animated and the animation duration is synchronized to that of the smaller camera views in the spatial display.

Infrastructure

The different versions of the multi-channel video player enable users to synchronously view several video streams. Our system supports viewing live or recorded video and can

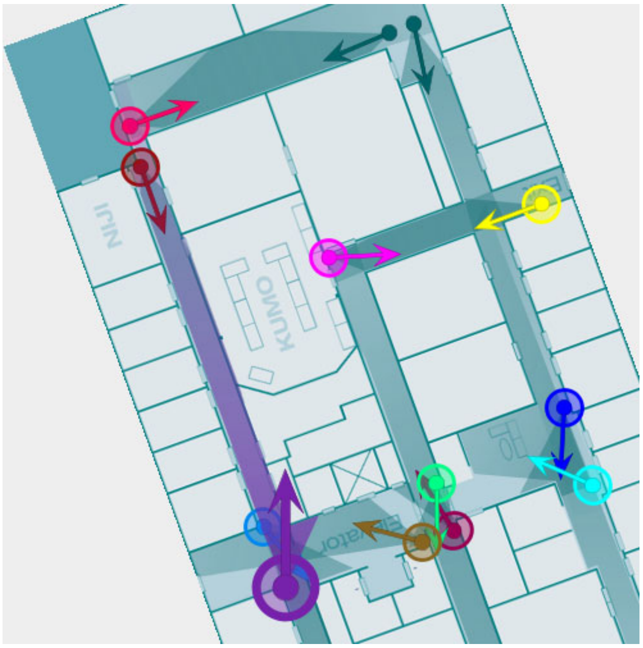

Figure 5. The rotated map indicates the position and direction of the selected camera. The position and view area of the main camera is emphasized over the other shown cameras.

seamlessly switch from one to the other. Recorded video can be watched at different speeds and even in reverse. We use Axis IP cameras [5] that provide access to video as Motion JPEG via HTTP or as MPEG-4 via RTP. We decided to focus on Motion JPEG because it is simpler to process and better supports seeking to different times. Also, Motion JPEG does not require a dedicated codec on the client side so that we were able to build a Java client that animates video player windows.

We record two different resolutions (640x480 and 320x240) from each camera at 15 frames per second. The cameras could provide 30 frames per second in a single resolution but we decided against the higher storage and network bandwidth requirements. This results in 1.7 GB worth of images per camera per hour (70% for the larger resolution). Our current setup consists of 22 cameras; 18 of the cameras were used in the study. We store the last 24-hours of video for each camera. The total storage and bandwidth required are 900 GB of disk space and 85 Mbps of network bandwidth.

Both live and recorded video are served from a digital video recorder as Motion JPEG via HTTP. The server can support several video players and is mostly limited by hard disk latency when different video players access recorded video at different times. We are currently working on an approach that can record video at high frame rates and quality and later reduce the quality during less interesting periods. This is similar to approaches described by Korshunov and Ooi [14] and Pillai et al. [18].

EVALUATION

Underlying the design of the SMV player is the hypothesis that providing geographic context will improve peoples' ability to track activity from camera to camera. The SMV player includes three techniques for providing geographic context to aid in tracking activity: the layout of video streams around the main player based on their fields of view, the map, and the rotation of the map to coincide with the field of view of the selected camera. This study evaluates the effect of these techniques on user performance and satisfaction.

Designs Examined in the Study

Two variations of a traditional camera-bank interface and three variations of the SMV player were created to evaluate the relative value of these techniques. Each player was presented in a window of 1184x888 pixels. Both player interfaces had a main display of 480x360 pixels (10 fps) that the participants used to focus in on the activity being tracked. All variants presented a bank of all 18 cameras used in the study that were grouped such that cameras in close proximity were shown near each other. Participants could select new cameras to display in the main viewer by clicking on a camera view in the camera bank. Some of the designs included a map where participants could click on a camera icon in the map to select a camera.

The first traditional camera-bank interface used in the study (Traditional No Map, TNM) consists of the main display and the camera bank with views of 160x120 pixels (4 fps). The second traditional camera-bank interface is shown in Figure 6 (Traditional Static Map, TSM). It adds a map that shows the location of the cameras and identifies the camera being shown in the main player.

Three variations of the SMV player were used in the study. In all variations, the main display was surrounded by a variable number of camera views of 160x120 pixels (10 fps) on which the participants could click to switch to that camera. All variations also had a bank of 18 camera views in a smaller size of 120x90 pixels (4 fps) to keep the total number of video view pixels similar across the two interfaces. The first variation consisted of the camera bank and the main player with surrounding smaller views (Spatial No Map, SNM). The second variation added a map similar to that in the traditional player (Spatial Static Map, SSM). In the third variation shown in Figure 4, the map rotates such that the view direction of the selected camera always points to the top of the map (Spatial Rotating Map, SRM).

Task and Video Data

Subjects were given the task of following a particular person from the moment they entered the building until they exited the building by keeping the best view of that person in the main viewer. Six different segments of activity between 161 and 244 seconds were prepared. Each segment was scripted to include about the same amount of activity in the building and similar challenges for tracking, e.g., the same number of times the target disappeared from view by walking into an office or through an area that lacked video coverage. We used three different actors to be tracked in the segments. The difference in segment lengths is due to the different walking speeds of the actors. Segment lengths with the same actor differ at most 6 seconds. The order of the six video recordings was kept the same for all subjects.

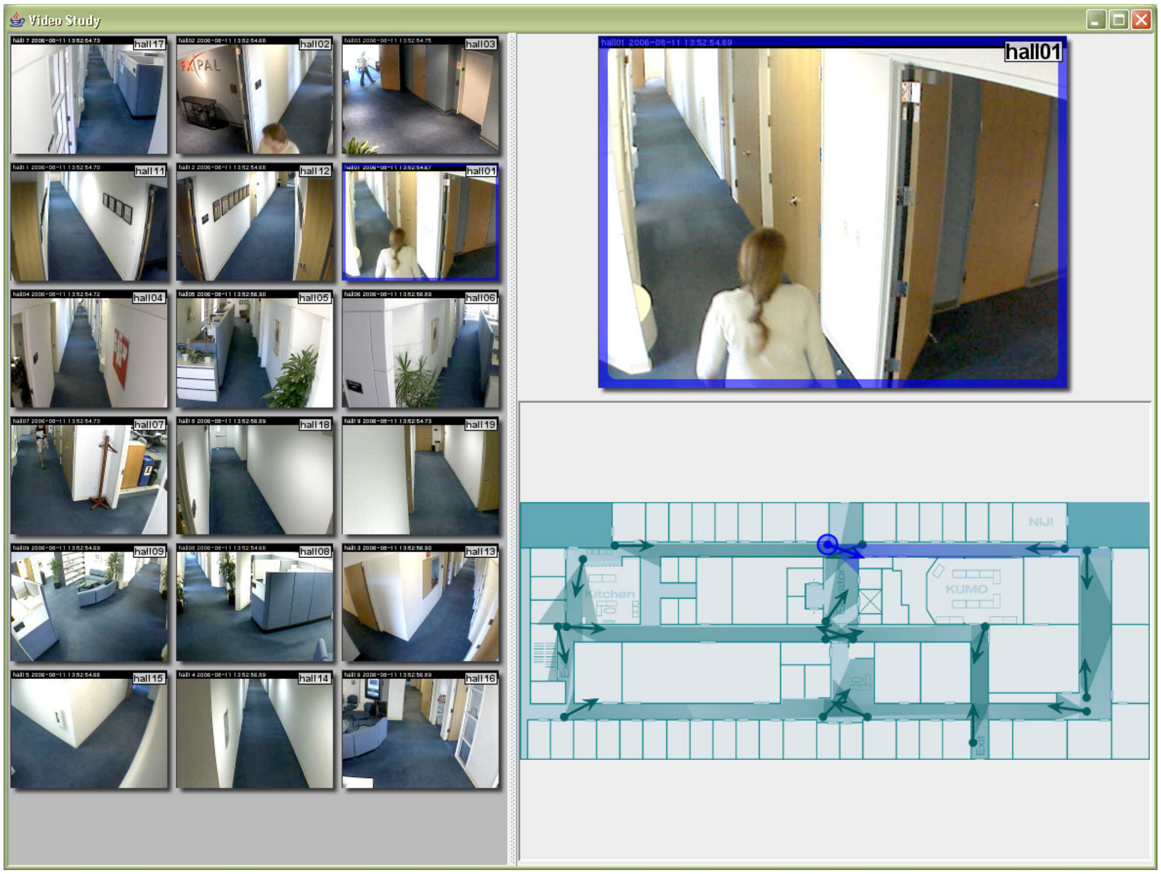

Figure 6. The traditional interface with map.

Study Procedure

These interfaces were used by 16 participants (8 female) between the ages of 23 and 53 (mean 36.7) recruited from regular employees, contractors, and interns in the building. Eight participants were long-time employees and eight were recent hires or interns with only a few months of experience in the building. All participants used computers regularly and none had worked in a job that included video tracking.

The first portion of the study compared the use of the more traditional camera-bank interface with the SMV player (TNM, TSM, SNM, SSM). Subjects were divided into eight groups of two using a modified latin-squares design. Each group included one long-term and one short-term employee to remove effects caused by differences in knowledge of the building. We decided to have subjects use both versions of the camera-bank interface first or both versions of the SMV player first in order to allow training once with the full version of an interface just prior to its use.

The second portion of the study compared the SMV player with rotating and non-rotating maps (SSM, SRM). The SSM condition was the same in both parts of the study. The order of SSM and SRM was balanced with respect to employee type and the order in the first part of the study.

Questionnaires

Subjects filled out demographic questionnaires prior to the task. One question asked whether subjects generally rotated paper maps during use. Following each tracking task, subjects reacted to a series of seven-point Likert-scale statements (from strongly disagree to strongly agree) concerning their experience and success and the interface's ease of use and design. After all six tracking tasks were completed, subjects were given a questionnaire asking about their preferences with regards to the different designs and the value of different interface components as well as open-ended questions asking for additional comments and suggestions.

Spatial Ability Test

People vary in their ability to extract and interpret geographical cues and to effectively transform map or direction information into a 3D representation. These spatial abilities may be factors in performance on the tracking task. To determine if participant spatial ability is correlated with their use of a map to follow the person being tracked, we asked all participants to complete a survey testing several related spatial abilities after the tracking tasks. We adapted several subtests from [16] on mazes and map reading. Because almost all participants completed the survey without errors, we used the survey completion time as an approximation for their spatial ability.

Data Analysis

To compare performance on the tracking task, ground truth for each of the six time periods was encoded with each camera being assigned one of four values at all times: 0 if the person being tracked was not in the camera view, 1 if the person was barely visible in the camera view, 2 if the person

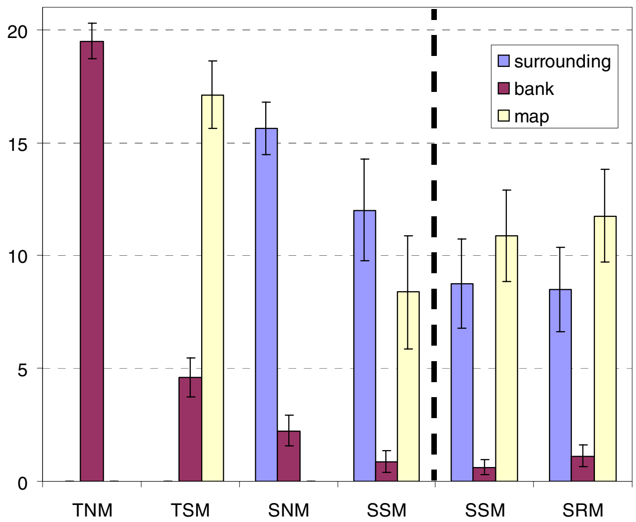

Figure 7. Mean number of mouse clicks (high spatial ability).

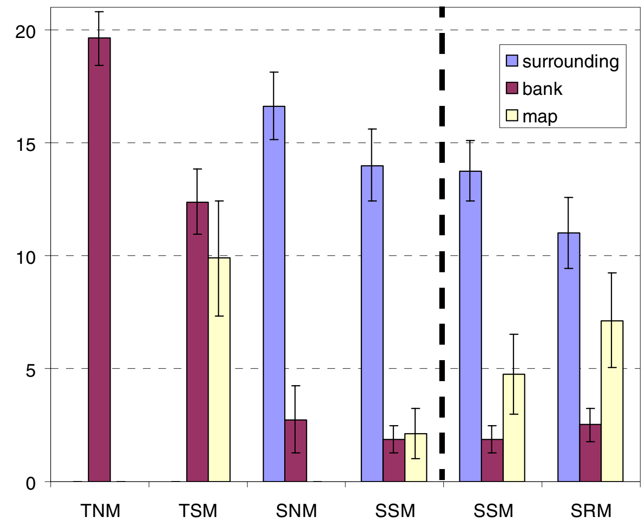

Figure 8. Mean number of mouse clicks (low spatial ability).

was visible but no details could be determined from the view, and 3 if the camera had a fairly detailed view of the person. Based on these assignments, the best possible score was determined by summing the best view for each second across the length of time the person was in the building. Upon completion of each task, the system logged the cameras that were selected, the time, and the method of selection (from camera bank, from map, or from spatial player). A subject's score was the sum of the scores of the camera in their main view; also at second boundaries. The ratio of the subject's score divided by the maximum possible resulted in a tracking performance score.

The survey responses, tracking performance scores and click patterns were analyzed with mixed model, repeated measures ANOVAs. Participants were divided into two groups according to their length of residence in the building (Employee type). The main effects and interactions for the analyses on questions about maps and animations and the map click data (Between-subjects variable Employee, within-subject variable Map) and for the main window click data (Between-subjects variable Employee, within-subject variable Interface) had Fs with degrees of freedom (1,14). All other analyses had main effects and interactions (Between-subjects variable Employee, within-subject variables Map and Interface) with Fs of (1,42) degrees of freedom. In the first part, we tested the following hypotheses:

- Availability of a map is better than having no map;
- The SMV interface is better than the traditional interface;
- The traditional interface without a map, lacking any form of geographic context, is worse than the other three.

In addition, we wanted to examine the effects of spatial ability on performance and attitudes. We categorized the participants based on their time to complete the spatial ability test into two groups of equal size. Although we term these two groups High and Low spatial ability, this categorization is relative to our set of participants, not to a standard population measure of spatial ability.

In the second part of the study, we compared the use of static or rotating maps.

RESULTS

The first portion of the study compared the use of a traditional camera-bank interface with and without a map to the SMV player with and without a map to determine the effect of the geographic context provided by a map, the spatial layout of selected smaller camera views around the main camera view, or both. The second portion of the study compared the use of the static and rotating maps with the SMV player.

Usage of Different Interface Components

While there are some statistically significant differences in the number of mouse clicks depending on the condition, most of these differences appear to be related to the availability of components. For example, when there is no map, clicks that may have been directed toward the map are instead directed toward the bank of cameras or the spatially-arranged camera views around the main window, if it is available (see Figures 7 and 8). Switches in camera view were mostly due to the tracked person leaving a camera view.

The use of the map is strongly correlated with the completion time in the spatial ability test discussed above. Thirty percent of the variance in number of clicks on the map is accounted for by this factor ($F(1,14)=21.20$, $p<0.001$). The eight "High Spatial" participants used the map frequently for switching to a different camera (see Figure 7). They used the surrounding camera views in the SMV player at a similar frequency and avoided the use of the camera bank. In contrast, the "Low Spatial" participants used the map less than half as often as "High Spatials" did (see Figure 8).

Neither group made much use of the camera bank when other alternatives were available. In observing the participants, we noticed that participants often used the camera bank as a means to recover after they lost the tracked person by switching to the wrong camera view.

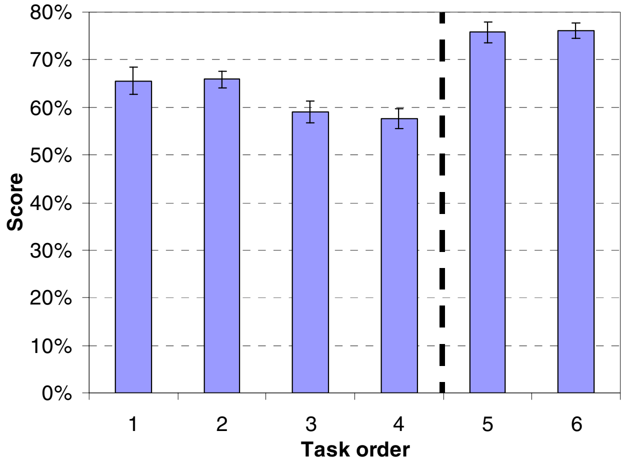

Figure 9. Mean scores by order of task completion.

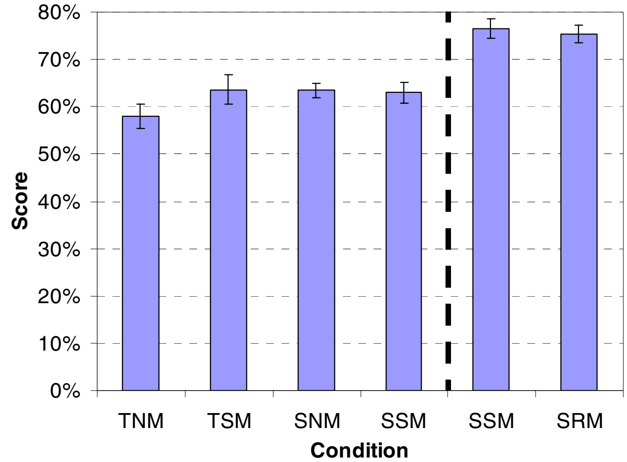

Figure 10. Mean scores by condition.

Effect of Actors' Walking Speed

Walking speed of the actor turned out to be a major (and unexpected) factor on the participants' performance. As described earlier, the routes were scripted to have similar lengths and tracking difficulties. This worked well within the two routes of each actor with a difference between 1 and 6 seconds. However, the sums of the two durations were 406, 325, and 486 seconds, respectively, for the routes of each actor. Because we did not vary the order of actors and grouped the two tracks of each actor, the effect can be seen in Figure 9 where the scores were best for the slowest actor (5 and 6) and were the worst for the fastest actor (3 and 4). It is unlikely that a training effect is responsible for the improved performance in the last two tasks because there is no noticeable improvement in performance in the second task in each actor group compared to the first task. There is no effect of walking speed in the second part of the study as only the slowest actor was used. Walking speed is a source of variance in the first part of the study but two actors were evenly distributed across the first four conditions because the order of all conditions was counter-balanced.

Value of Map and Spatial Player

The central hypothesis of the SMV player design is that geographic context aids tracking. This is strongly supported by the result that subjects using the spatial viewer, a map, or both, performed significantly better than those provided with the traditional camera bank ($F(1,42)=4.15$, $p<0.05$) in planned post-ANOVA orthogonal contrasts. All forms of geographic context aided the subjects and there were no significant differences in subject performance among the other three conditions, as shown in Figure 10. In the first part of the study, the traditional interface without a map performed worst and the other three interfaces were about equal.

The survey results showed similar results (see Figure 11; for the complete question text see Table 2). The traditional camera-bank interface without a map was viewed as the worst on all questions related to tracking ability and player functionality. The spatial viewer without map was generally viewed as better than the traditional interface without a map but not as positively as the two interfaces with maps. ANOVAs showed a significant positive impact of the map on a number of items. Long-time employees performed on

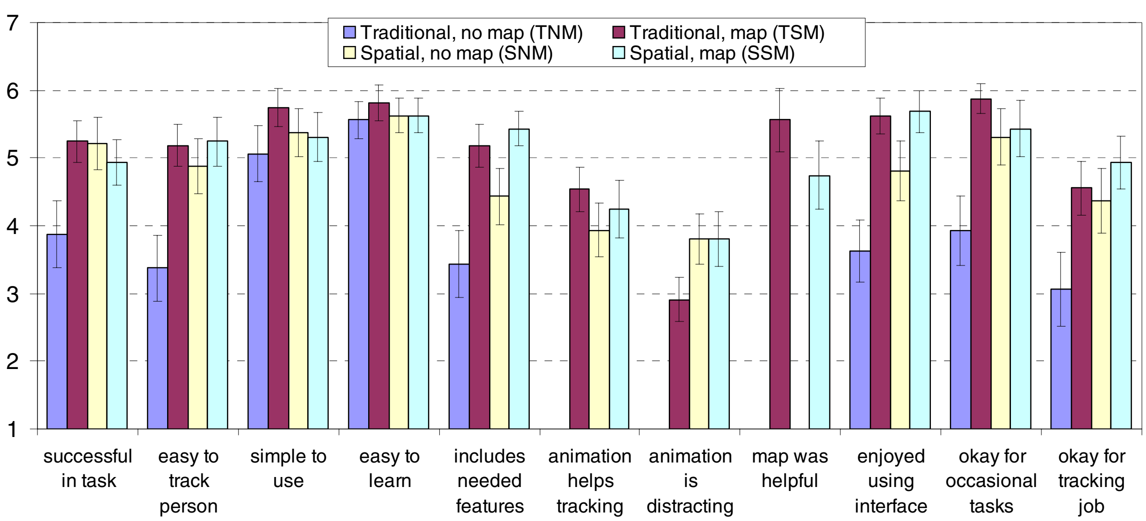

Figure 11. Survey results from first part of the study.

Measure	Factor	F	p
I was successful in tracking activity from camera to camera.	I	4.04	~
	M	4.54	*
	I x M	10.41	**
	M x E	5.64	*
The interface made it easy to follow a person from camera to camera.	I	6.18	*
	M	12.11	**
	I x M	5.23	*
	M x S	5.64	*
The interface included the necessary information and features.	I	5.04	*
	M	24.40	***
	M x S	10.42	**
I enjoyed using this interface.	I	4.02	*
	M	21.26	***
	I x M	3.26	~
If I occasionally tracked activity, this interface would be acceptable.	M	13.57	**
	I x M	10.48	**
	M x S	10.54	**
If my job was tracking activity, this interface would be acceptable.	I	6.84	*
	M	10.22	**
	I x S	3.370	~
	M x S	8.50	**
Performance score	M x S	7.91	**
	I x M x S	13.43	**
Clicks on camera bank	I	205.53	***
	M	51.67	***
	I x M	34.36	***
	I x S	4.98	*
	M x S	8.10	**
	I x M x S	6.23	*
Total number of clicks	I	5.23	*
	M	5.63	*
	I x M x E	3.38	~
Clicks on main window: F(1,14)	M	4.08	~
Clicks on map: F(1,14)	I	21.35	***
I=Interface, M=Map, S=Spatial ability, E=Employee type Significance levels (p): ~ 0.08, * 0.05, ** 0.01, *** 0.001			

Table 1. Results from first part of study. Most are F(1,42).

average slightly better in every condition although this was not statistically significant. See Table 1 for a summary of the results.

Static and Rotating Maps

The second portion of the study compared the use of static and rotating maps in the SMV player. The pre-task questionnaire showed that 13 of the subjects occasionally rotated paper maps when navigating. Nine subjects reported they do this regularly (more than 1/4 of the time they use a map).

In the second part of the study, there is no significant performance difference between the rotating and the static map. Figures 7 and 8 show that the rotating map was used more frequently. However, this difference is not statistically significant. Although this may indicate that some participants found it more useful, its increased use could be due to increased attention on the map resulting from its rotation.

Statement	SSM	SRM	p
I was successful in tracking activity from camera to camera.	5.75	5.56	.38
The interface made it easy to follow a person from camera to camera.	5.63	5.31	.14
The interface was simple to use.	5.63	5.25	.16
The interface was easy to learn.	5.88	5.13	**.01**
The interface included the necessary information and features.	5.19	5.44	.48
The animation in the interface helps in tracking activity between cameras.	4.56	4.25	.43
The animation in the interface is distracting.	3.53	4.47	**.02**
The map was helpful.	4.88	5.00	.77
I enjoyed using this interface.	5.44	5.00	.11
If I occasionally tracked activity, this interface would be acceptable.	5.63	5.13	**.04**
If my job was tracking activity, this interface would be acceptable.	4.88	4.56	.24

Table 2. Survey results from second part of study (paired t-test).

The post-task surveys showed users had a slight preference for the static map. Paired t-tests showed significant advantages for the static map in the perceived ease of use, how distracting the animation was, and whether the system was acceptable for occasional tracking tasks (see Table 2). One subject indicated a strong preference and three others a slight preference for the rotating map.

DESIGN IMPLICATIONS

Clearly, spatial information is helpful in tracking tasks. The lack of a map in the traditional interface leads to worse performance and lower satisfaction. The small camera views surrounding the main view in the SMV player were the most frequently used means of switching to a different camera. Also, the SMV player without the map scored significantly higher than the traditional player without the map in user satisfaction and performance. This leads us to conclude that the main features of the SMV player, the selection of geographically proximal cameras and the spatial arrangement of smaller camera views around the main view, is suitable for tracking tasks. Those features support the user in deciding which of the surrounding views to select to continue tracking a person

In the SMV player, the map is not used much when available and there is no difference in performance between the map and the spatially arranged camera views. A likely explanation is that the spatially-arranged camera views were sufficient for the tracking task. However, the participants indicated that they missed the map, especially with their answers to the questions “includes needed features” and “enjoyed using interface.”

The rotating map received mixed opinions. It did not affect performance but it scored lower in most of the survey questions. It appears to have been used more than the fixed map but this trend is lost in the variance of overall map usage. The use of maps is strongly correlated with the completion time of a spatial ability test. Because of individual differences, both the map and its orientation need to be user-selectable options.

The animation in the SMV player was rated fairly neutral with respect to helpfulness and distraction. Some participants commented that the animation was too slow and thus made it more difficult to correct mistakes after selecting the wrong camera.

Overall, the results are promising and indicate that our approaches to presenting spatial information are helpful for tracking tasks. Because several of the study participants asked for visualizations on the map depicting the locations of people, we plan to add such a feature to our system soon. Also, we are working on visual overlays to the video displays that would draw the user's attention to the display that might contain the tracked person.

CONCLUSIONS

Video surveillance installations are growing in number and size. Tracking activity between cameras is an important but difficult task. Typical systems in use today do not include integrated geographical aids. This contributes to difficulties in cross-camera activities such as tracking. The SMV player provides spatial layout of proximal cameras, maps, and map rotation to provide geographic context to aid in this task. Our studies show that this information contributes to more successful tracking activities.

The SMV player displays related video streams such that their placement indicates the spatial relationships of the cameras that captured the video streams. The map marks a camera location with a color-coded dot, provides an arrow indicating each camera's view direction, and highlights the main camera with an enlarged dot and arrow. The rotation of the map provides additional context regarding the direction of the field of view of the currently selected camera.

Requests for visualizations on the map depicting the locations of people indicate that orientation information may be more effective if it includes the objects moving in the space as well as the space itself. We are currently working on such a system.

REFERENCES

1. 3VR Security. http://www.3vr.com/
2. F. Arman, R. Depommier, A. Hsu, M.-Y. Chiu, Content-based browsing of video sequences. In Proc. ACM Multimedia 94, San Francisco, CA, pp. 97-103, 1994.
3. E. Ardizzone, M. La Cascia, G. Lo Re, and M. Ortolani. An integrated architecture for surveillance and monitoring in an archaeological site. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 79-86, 2005.
4. P.K. Atrey, M.S. Kankanhalli, and R. Jain. Timeline-based information assimilation in multimedia surveillance and monitoring systems. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 103-112, 2005.
5. Axis Communications. http://www.axis.com/
6. R. Cucchiara. Multimedia surveillance systems. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 3-10, 2005.
7. G. Daniel and M. Chen. Video Visualization. Proc. IEEE Visualization 2003, pp. 409-416, 2003.
8. Energy Control, Inc. http://energyctrl.com/security.htm
9. J.P. Freeman, B. McChesney, M. Denari, and S. Thompson. The explosion in intelligent video, International Security Conference West, 2006. http://www.iscwest.com/
10. Guardian Solutions. http://www.guardiansolutions.com/about.asp
11. L. Huston, R. Sukthankar, J. Campbell, and P. Pillai. Forensic video reconstruction. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 20-28, 2004.
12. G. Iannizzotto, C. Costanzo, F. La Rosa, and P. Lanzafame. A multimodal perceptual user interface for video-surveillance environments. Proc. Conference on Multimodal Interfaces, pp. 45-52, 2005.
13. Johnson Controls. http://www.johnsoncontrols.com/security/Default.htm
14. P. Korshunov and W.T. Ooi. Critical video quality for distributed automated video surveillance. Proc. ACM Multimedia, pp. 151-160, 2005.
15. L3 Communications. http://www.l3praetorian.com/
16. Mechanical and Spatial Aptitude, LearningExpress, LLC:NY, 2001. http://booklearning.tripod.com/Mechanical_and_Spatial_.pdf
17. ObjectVideo. http://www.objectvideo.com/products/
18. P. Pillai, Y. Ke, and J. Campbell. Multi-fidelity storage. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 72-79, 2004.
19. I.O. Sebe, J. Hu, S. You, and U. Neumann. 3D video surveillance with augmented virtual environments. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 107-112, 2003.
20. Vidient. http://www.vidient.com/
21. VistaScape Security System. http://www.vistascape.com/
22. G. Wang, T.-T. Wong, and P.-A. Heng. Real-time surveillance video display with salience. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 37-44, 2005.
23. J. Wang, M.S. Kankanhalli, W. Yan, and R. Jain. Experiential sampling for video surveillance. Proc. ACM Workshop on Video Surveillance and Sensor Networks, pp. 77-86, 2003.
24. G. Wu, Y. Wu, L. Jiao, Y.-F. Wang, and E.Y. Chang. Multi-camera spatio-temporal fusion and biased sequence-data learning for security surveillance. Proc. ACM Multimedia, pp. 528-538, 2003.