

## Visualizing the History of Living Spaces

Yuri Ivanov, Christopher Wren, Alexander Sorokin, Ishwinder Kaur

TR2007-068 August 2007

### Abstract

The technology available to building designers now makes it possible to monitor buildings on a very large scale. Video cameras and motion sensors are commonplace in practically every office space, and are slowly making their way into living spaces. The application of such technologies, in particular video cameras, while improving security, also violates privacy. On the other hand, motion sensors, while being privacy-conscious, typically do not provide enough information for a human operator to maintain the same degree of awareness about the space that can be achieved by using video cameras. We propose a novel approach in which we use a large number of simple motion sensors and a small set of video cameras to monitor a large office space. In our system we deployed 215 motion sensors and six video cameras to monitor the 3,000-square-meter office space occupied by 80 people for a period of about one year. The main problem in operating such systems is finding a way to present this highly multidimensional data, which includes both spatial and temporal components, to a human operator to allow browsing and searching recorded data in an efficient and intuitive way. In this paper we present our experiences and the solutions that we have developed in the course of our work on the system. We consider this work to be the first step in helping designers and managers of building systems gain access to information about occupants' behavior in the context of an entire building in a way that is only minimally intrusive to the occupants' privacy.

*Information Visualization, October 2007*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



# Visualizing the History of Living Spaces

Yuri A. Ivanov *Member, IEEE*, Christopher R. Wren *Member, IEEE*,  
Alexander Sorokin *Member, IEEE*, Ishwinder Kaur *Member, IEEE*

**Abstract**— The technology available to building designers now makes it possible to monitor buildings on a very large scale. Video cameras and motion sensors are commonplace in practically every office space, and are slowly making their way into living spaces. The application of such technologies, in particular video cameras, while improving security, also violates privacy. On the other hand, motion sensors, while being privacy-conscious, typically do not provide enough information for a human operator to maintain the same degree of awareness about the space that can be achieved by using video cameras. We propose a novel approach in which we use a large number of simple motion sensors and a small set of video cameras to monitor a large office space. In our system we deployed 215 motion sensors and six video cameras to monitor the 3,000-square-meter office space occupied by 80 people for a period of about one year. The main problem in operating such systems is finding a way to present this highly multidimensional data, which includes both spatial and temporal components, to a human operator to allow browsing and searching recorded data in an efficient and intuitive way.

In this paper we present our experiences and the solutions that we have developed in the course of our work on the system. We consider this work to be the first step in helping designers and managers of building systems gain access to information about occupants' behavior in the context of an entire building in a way that is only minimally intrusive to the occupants' privacy.

**Index Terms**—Sensor networks, user interfaces, surveillance, timeline, spatio-temporal visualization.

---

◆

## 1 INTRODUCTION

On July 8, 2005, a day after the tragic events of London bombings, a member of the intelligence community addressed the audience of computer vision scientists at a meeting that we attended. He said: “It is most likely that right now dozens of people in London are watching thousands of hours of videotapes while whoever they are looking for is getting away.”

That statement emphasized an important problem that, paradoxically, has been created by advances in technology. Vast amounts of video data can be collected and stored cheaply, but when it is needed, it is ultimately a human operator who needs to look at it and decide if it is relevant.

In our view, most current approaches are *sensor-centric*: they present the world as it is seen through a sensor – say, a camera. However, people do not live their lives so as to allow a better view of themselves to a video camera. Consequently, in computer vision, before an event can be analyzed, it is often a challenge to normalize pose, scale, or illumination, or to find a video that contains some event in its entirety.

Understanding that the hopes for “better data” will never materialize, we have developed an alternative approach to building monitoring systems that we hope is not only applicable to surveillance but to a larger group of problems, generally utilizing awareness of large spaces that are populated by people. We propose a *space-centric* approach, which gives some information about the entire space as a unified representation, allowing analysis and search for events that might span a large set of sensors and extend for long periods of time. To this end, we have built a system that uses a sparse array of video cameras but keeps track of the global context with a large network of motion sensors. This context makes it significantly easier to solve problems of large-scale monitoring and search.

- 
- Yuri A. Ivanov is with Mitsubishi Electric Research Labs, E-mail: [yivanov@merl.com](mailto:yivanov@merl.com).
  - Christopher R. Wren is with Mitsubishi Electric Research Labs, E-mail: [wren@merl.com](mailto:wren@merl.com).
  - Alexander Sorokin is with UIUC, E-mail: [sorokin2@uiuc.edu](mailto:sorokin2@uiuc.edu).
  - Ishwinder Kaur is with MIT Media Lab, E-mail: [kaur@media.mit.edu](mailto:kaur@media.mit.edu).

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 27 October 2007.

For information on obtaining reprints of this article, please send e-mail to: [tvcs@computer.org](mailto:tvcs@computer.org).

We have previously presented the hardware used and a number of individual components of the interface [6, 7] while focusing on the technical aspects of managing the data. In this paper, we focus on the design considerations for the entire project, seamlessly bringing together these and other components within a unified view of the interface, visualization, and interaction. We detail the user interface and visualization problems, present the principles that guided us, and elaborate on the solutions.

While focusing largely on the problems of inference in sensor networks and computer vision, the main principles that we tried to follow in designing the system were simplicity of use, responsiveness, and clarity of visualization. In fact, the relative success of the system is largely due to the painstaking process of trial and error used to discover exactly how to make the information that we displayed be understood pre-attentively, i.e., at a single glance. Even though we have not conducted a formal user study, we believe that we have been reasonably successful in our solutions, perhaps maybe even reinventing the wheel now and again, as a visualization professional would probably point out.

## 2 RELATED WORK

The main visualization challenge that we address in our system is simultaneous presentation of time and space. This problem has long been known in the visualization community, in particular in map-based applications. Some examples can be found in [3, 8, 10], where an extra dimension is added to a two-dimensional map display to show temporal order of events while offering no interaction. However, this method of visualization cannot be directly used on our data, as its spatial component is generally three-dimensional and the number of simultaneous events causes visual clutter.

More relevant to our applications, is the work of Andrienko *et al.* on the visualization of changing map-based information [1]. They suggest augmenting a map with a graph of temporally changing parameters pertaining to the information displayed on the map. Our approach is similar in spirit, but we give more importance to the interaction of the two components, using them for display as well as for interaction and query building.

Geldon and Bouthemiy [4] use motion descriptors obtained from the recorded video to summarize the shots and index them for future retrieval. In contrast, our approach uses other sensors and logical as well as temporal conditions on them to index into large collections of video data.

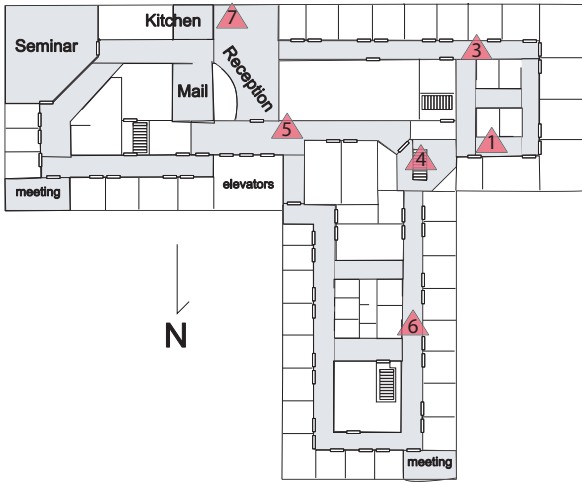


Fig. 1. Map of one floor of the office space. Shaded areas show public spaces where the sensors are installed. Locations of the six cameras are marked by small triangles.

There has been quite a lot of work done on analytic techniques to extract index features from video data, e.g. [2, 12]. These index sets are used to facilitate future searches. The problem is knowing what index data to create, i.e., correctly predicting what information will be useful a day, a week, or years in the future. We chose to index very general data in the form of sensor activations. Each individual index datum is not very meaningful, but together they represent the context succinctly and can be visualized in a way that is accessible to human operators. Such representation of context lends itself to efficient searches that are defined by the interaction with the operator at the precise time of need.

The primary representation of time in our system is accomplished by a *timeline* control, which records the activations of a large array of sensors in order of their arrival and is similar in appearance to a piano roll [5]. The effectiveness of this class of controls for multi-stream data has been demonstrated more recently [9, 11, 14].

### 3 VISUALIZATION OF SPATIO-TEMPORAL DATA

A frequent problem with sensor networks involves the visualization of large volumes of spatial data that have a temporal order. Several solutions have been proposed in the past, including the work by Kapler and Wright [8]. A common approach to displaying the temporally varying spatial data is to add a visual dimension representing time. Unfortunately for our applications, this approach is not generally acceptable due to the higher dimensionality of the spatial data, as we explain below.

#### 3.1 Problem domain

The system that we describe here is built for the purpose of long-term monitoring of indoor spaces occupied by large number of people. The system addresses the problem of flexible search of historical data that is collected from a variety of generally heterogeneous sensors. The problem is very common in surveillance. For the sensing part of the system we use a sparse array of Pan-Tilt-Zoom (PTZ) cameras and a large network of simple wireless motion sensors (detailed in our earlier papers, e.g. [6, 7]). A typical target application for such a system might include a theft taking place in a busy office space for which users would like to quickly reconstruct the sequence of events and identify possible perpetrators and witnesses from whatever partial information might be available.

#### 3.2 Data

The data in our system was recorded from a network of 215 wireless ultra-low-power motion sensors over the course of about a year. We also collected videos from six PTZ video cameras for one month

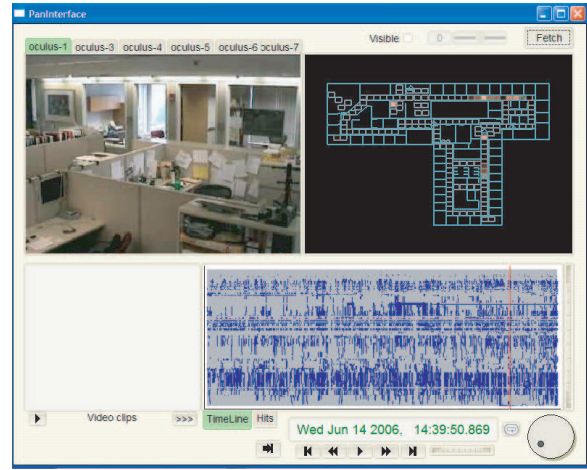


Fig. 2. User interface for the search module of the system. Clockwise from top left: camera view, map, timeline, clip bin.

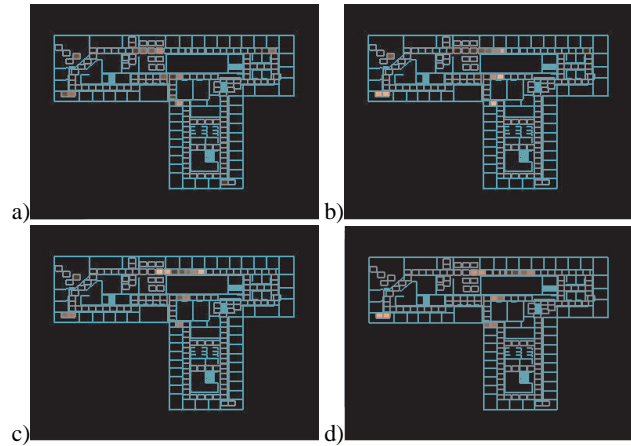


Fig. 3. Time-lapse capture of the map control while displaying movements of several people in the office space. Location and approximate number of people can be estimated instantaneously for the entire space.

of this year. Motion sensors were attached to the ceiling at regular intervals of approximately 2 meters covering the area of 3,000-plus-square-meter office space. The motion sensors report the presence of motion in their field of view by sending a single bit to the server, which stores the time stamp of the activation in the database. For the experiments we chose a randomized Pan-Tilt policy for the video cameras, which causes them to foveate to a random location every few minutes. One of the cameras is omni-directional and is located at a central junction between the north and east wings of our building. In all, we have approximately 20,000,000 sensor activations and 150,000,000 video frames, totalling 1.7TB of storage space. The map in Figure 1 depicts the test area.

The data collected in our lab spans two floors of the office building. In general, the occupants of the space travel freely between the floors via the staircase: therefore, the data represents a temporal evolution of three-dimensional space. This fact makes it difficult to use any topographic-like visualization methods that represent time by adding another visual dimension, as it would inevitably increase clutter in the display. To avoid clutter, we chose to approach the problem of visualization as a set of linked two-dimensional views, as described in Section 4.

#### 3.3 Visualization principles

Typical surveillance and monitoring techniques rely heavily on applications of computer vision to detect interesting activity. As the

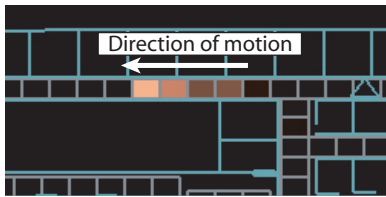


Fig. 4. Determining direction of motion from a sensor map. The sensor that is most recently activated is plotted in a bright color, which then slowly fades to black. The direction of the gradient helps to quickly identify the motion at a single frame.

very first glance at the volume and the rate of acquisition of the data should prove, we cannot use any but the lightest-weight computational processing. In fact, we opted to build a video monitoring system that uses practically no computer vision. That is not to exclude computer vision from this application entirely, but simply to use it sparingly, on much smaller subsets of data that we select by some other means.

The main guiding insight of the system design was the idea of a *space-centric* representation. We tried to avoid the traditional camera-centric approach and not show what is happening in front of the camera, but rather what is happening in the physical space. In designing the system we formulated the following requirements and then applied them to the interaction design (albeit with a varying rate of success):

1. All information presented on the display should be accessible pre-attentively. For example, all information about motion in the space should be evident from a single image.
2. The display should allow the assessment of the situation in the entire location, not only in front of a single sensor.
3. The interaction should be easily transferable to other interaction modalities – mouse, pen-based tablet, touch table (Diamond-Touch, see Figure 6). To this end, we tried to carefully map the manipulations available with each pointing device, making sure that all chosen manipulations are available on all devices. Additionally, we implemented simple control gestures for screen manipulations and query selection.
4. Unused parts of the data should be out of the way of the main interaction. We attempted to achieve this by using tabbed panels, which we will not detail further in this paper.
5. All display elements should be temporally consistent and respond to a single centralized clock. Any change in the global time reference should cause all visible display elements to show the data related to that time.
6. All data processing required for indexing and storage of the data should be much faster than real-time. If the processing takes longer than a small fraction of real-time, then the system will not scale well.

These principles helped us in the development of the system, but were not always easy to implement. The tracking module of the system was especially challenging, as it required simultaneous display of the paths of several people in correspondence with other parts of the interface. We describe the visualization design decisions in Section 6.

#### 4 SPACE-TIME VISUALIZATION IN MONITORING APPLICATIONS

The multidimensional character of the spatio-temporal data dictated the approach we took in displaying it to the user. Our solution was to use a dual view methodology by splitting a single display into a pair of synchronized panels. The first type of display corresponds to an instantaneous snapshot of the sensor field overlaid on the floor plan of the office space. We call it the *map*.

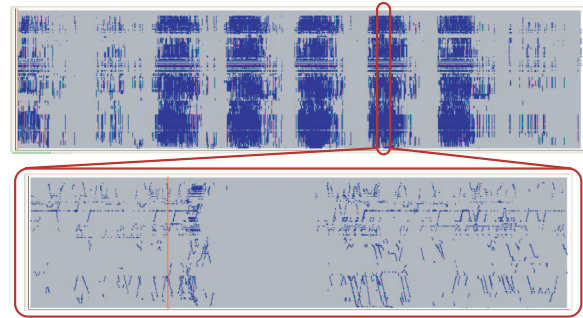


Fig. 5. Timeline control. The top example shows the timeline at its eight-day resolution. A user can zoom in on any portion of the timeline. The example on the bottom shows approximately two-hour range.



Fig. 6. One interactive installation of the system uses a DiamondTouch projection table. Interactions with the table make it useful to have a gesture-based interface that can allow the same manipulations as would normally be performed with a multi-button mouse.

The second type of display shows history of sensor activations over some period of time. We call this display the *timeline*.

The two main panels are shown on the right side of the screenshot of the search module of the system in Figure 2. The figure also shows the *camera view* panel on the top left and a *clip bin* on the bottom left.

##### 4.1 Visualization of space

The first type of display is shown in a series of screenshots taken at brief time intervals in Figure 3. Each image in the figure shows the floor plan with the positions of the sensors represented by gray rectangles. Once a sensor is activated, the color of the corresponding rectangle is set to light orange, which subsequently fades to black at a fixed decay rate.

If a person passes under a chain of sensors, they would generate a fading trail, which is easily interpreted visually, providing cues about the person's direction and speed of motion upon a momentary exposition. This is illustrated in Figure 4. From the direction of the fading pattern of the color we can easily judge the direction of motion of the person activating the motion sensors. Since the decay rate is locked with time, the perceived length of the trail directly corresponds with the velocity of motion.

##### 4.2 Visualization of time

Complementary to the *map* display is the *timeline*. The timeline shows the history of sensor activations for the entire space. The sensors are arranged along the vertical axis of the control while preserving some notion of neighborhood. The horizontal direction of the panel represents time in the left-to-right order.

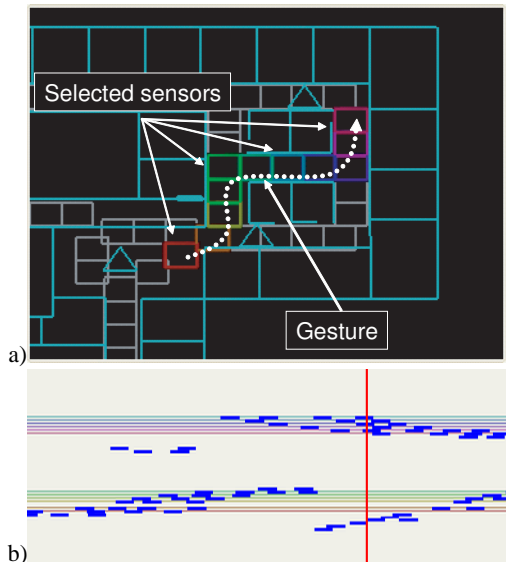


Fig. 7. Top: A query is specified by a gesture traversing a path across the map. The selected sensors are coded in unique colors. Bottom: Selected sensors are highlighted with the same color to establish visual correspondence.

The display is similar to a “piano roll,” where one horizontal row corresponds to a single unique sensor. Every sensor activation is marked on the row by a small blue line.

As can be seen in Figure 5, one can estimate very quickly temporal patterns of space use over time. For instance, the snapshot at the top of Figure 5 corresponds to a week’s worth of data. In this display, day and night are clearly identified. Weekends are seen as lower-intensity vertical blue bands.

A user can also assess anomalies in the use of the space. The bottom snapshot in Figure 5 shows a zoomed-in view of the middle of one day. The gap in the sensor activations corresponds to a fire alarm, with the flurry of activations preceding it showing the evacuation of the entire floor of the building. This data shows that the procedure took three minutes. Curiously, the majority of the people exited through a single fire escape, while two other exits remained mostly unused. This fact is clearly evident when the map and timeline visualizations are used jointly.

The timeline is used as a central time manipulation control that drives the system clock. Operators can use a number of manipulations to change the reference time as well as the time range. The time is changed by simple scrubbing<sup>1</sup> of the control in either autoplay or paused modes. The timeline can be panned and zoomed with simple familiar gestures and manipulations.

The timeline may be switched to display *hits*. Hits are time intervals that contain sequences of sensor activation in the order requested by the user (see the following section). In this display all such ranges are stitched together, while irrelevant time intervals are removed from the display, as shown in Figure 8a.

## 5 SPATIO-TEMPORAL QUERIES

Browsing the recorded history is only a small part of the functionality needed to enable a directed search for interesting events in the monitored space. To find these events, we currently rely completely on the set of motion sensors. One reason for that is the search efficiency. We can use the sensor network to define the context for the video. For instance, we may be interested in videos of people who traversed a certain path through space, or we may be interested in finding participants of activities that happened outside the camera view. To achieve these goals, we developed the query module for our system.

<sup>1</sup>Dragging the time marker across the timeline.

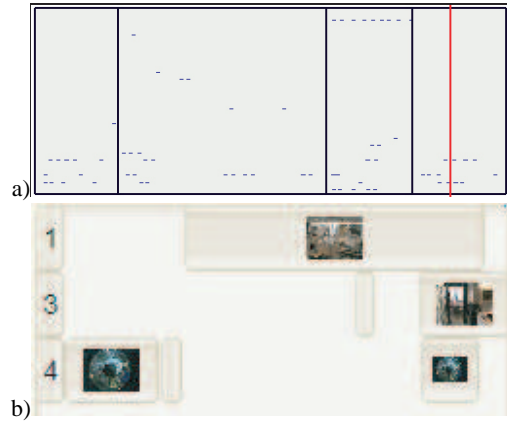


Fig. 8. Top: The timeline in the result display mode. Bottom: Selecting any of the hits causes the clip bin to fill up with video clips containing the accompanying video evidence.

### 5.1 Gestural query interface

The query interface in our system is implemented as a front-end to an SQL engine that maintains the database of sensor activations. The “path” queries, described above, are formulated by simply drawing a path of interest on the map. While drawing, the closest sensor is identified for each point in the trajectory and its ID is inserted into a query constraint set. Then all time intervals that contain the sensor IDs activated in the order traversed by the gesture are retrieved from the database.

An example of such query is shown in Figure 7a. As sensors are selected they are colored in a unique set of hues. These colors reflect the order in which they are selected and serve as visual integration cues. Sensors that have been selected are highlighted with the same colors on the timeline (Figure 7b), establishing visual correspondence between the controls.

The result of the query is displayed in the timeline as a series of time ranges that contain the specified sequences of sensor activations. As useful as it is, this range-style hit can still be improved. Indeed, the range result only suggests to the user that there might be video data from any of the cameras that might have observed the behavior in question, but it would still be up to the user to determine which camera has observed which part of the trajectory of the sensor activations.

We solve this problem by first calibrating the cameras to the sensor network. The calibration does not involve finding the extrinsic parameters of the camera but rather a much simpler calibration of the control parameters. For each sensor we calculate the range of PTZ control parameters of each camera that results in the sensor field of view being contained in the camera field of view. Then, if we keep track of the PTZ parameters of the camera at all times<sup>2</sup>, the query can be conditioned on *sensor visibility*. Consequently, for every sensor activation in the retrieved result set, we can determine whether any camera observed the participant who activated the sensor and retrieve the corresponding video clips.

A set of such clips is placed into the *clip bin* shown in Figures 8b, 2, and 14. Note that this is not a simple linear edit of the video – the resulting clip set contains the entire set of video evidence from all cameras that observed any part of the motion pattern returned by the query. Figure 9 shows several representative frames from the selected clips. This mechanism effectively implements a camera handover for even remote non-overlapping cameras.

## 6 DISCOVERY OF EMBEDDED STRUCTURES

Previous sections have focused on taking dimensional slices through the data, exposing primarily spatial or temporal information. This

<sup>2</sup>For example, in our cameras the exact values of Pan, Tilt and Zoom parameters are stored in the header of every image. Alternatively, a simple polling mechanism can be employed with most PTZ cameras.

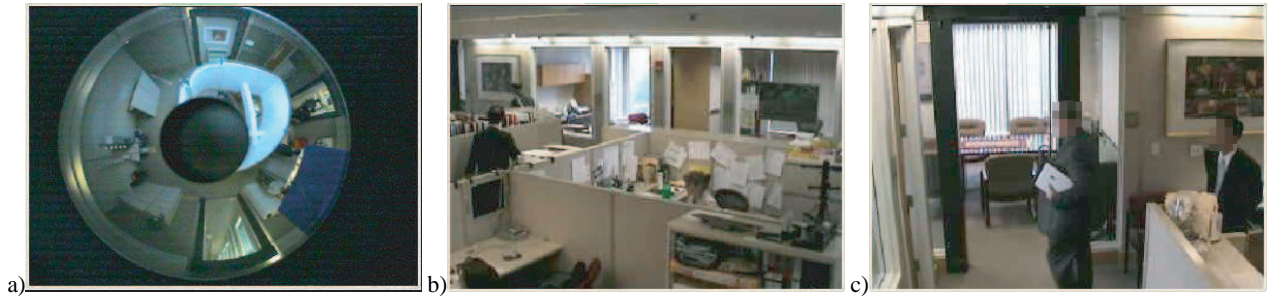


Fig. 9. Selected frames of the videos from the clip bin. The clips demonstrate automatic handover and tracking mechanisms.

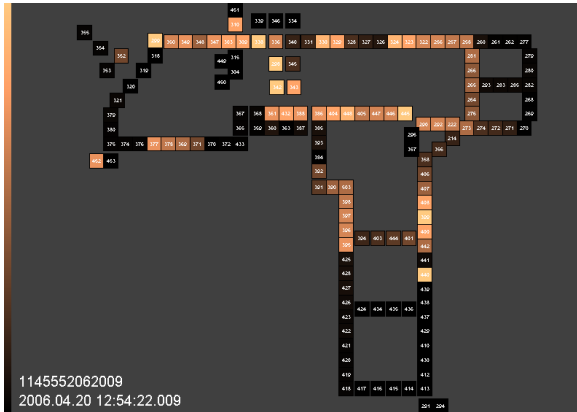


Fig. 10. Example of crowd movement during a fire drill.

dataset is interesting because it contains significant structures that exist across these visualizations. As people move through the building they create a spatio-temporal trace of motion activations. When they pass by or interact with other individuals their traces become tangled into a spatio-temporal graph of possibilities. In this section we show how the use of these graphs helps an operator quickly untangle these trace relationships and find all video- and sensor-based evidence relating to an individual occupant of the space.

### 6.1 Human-guided tracking

In this section we present the principles and challenges of information visualization used in the tracking module of our system. The technical background of track composition and analysis technique that we use here can be found in our earlier paper [6].

Due to the impoverished nature of the motion sensors, it is not possible to unambiguously track individuals through a building if they cross paths or otherwise interact with other individuals in the space. A massive crowd would generate an overwhelming mass of ambiguity, such as during a fire evacuation as depicted in Figure 10. However, the much more common case is that each individual interacts with a few others to create webs of ambiguity, such as those represented by the graph in Figure 11.

These graphs are constructed from nodes of ambiguity connected by unambiguous spatio-temporal traces called *tracklets*. The tracklets are depicted on the map as distinct lines tracing an unambiguous path through the space, as seen in Figure 13. At any moment there may be multiple tracklets under consideration: a series of selected tracks and several possible future continuations. These tracklets are distinguished from one another in several ways. First, each tracklet is coded with a unique color. Second, as is apparent in Figure 13, the tracklets are assigned spatially distinct channels on the map to reduce the possibility of overlap and improve the intelligibility of the display.

Finally, although tracklets may traverse a common space with other tracklets, they may do so in different directions and different times.

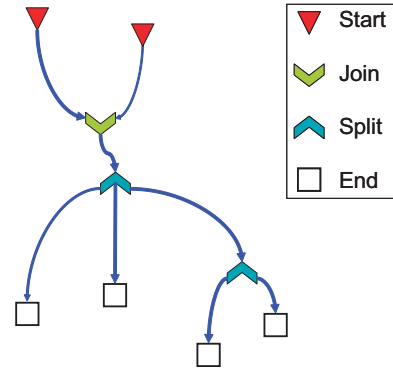


Fig. 11. Tracklet graph representation of the track bundle. Each edge, called a *Tracklet*, represents a contiguous sequence of sensor activations, while nodes represent ambiguities and endpoints.

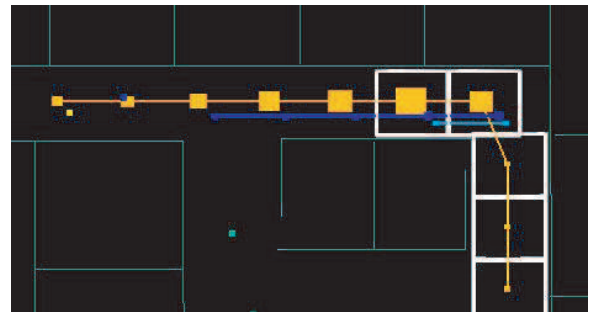


Fig. 13. Tracklet display. In order to achieve the pre-attentive assessment of the multitude of tracks and direction of motion we chose an asymmetric swell as a direction cue.

This temporal component is shown with an asymmetric swell that communicates both the direction and the current location of the individual. In Figure 13 we see that that the orange tracklet is active and the person is moving toward the right, while the blue and cyan tracks are currently not active.

As above, control over the temporal aspect of the visualization occurs in the timeline window. Scrubbing the time marker over the timeline simultaneously animates the swell of the tracklets on the map. This provides a very fluid mechanism of interaction with various tracklets over time.

The forensic surveillance system shown in Figure 14 allows the operator to build a story about the movements of a particular individual, presumably in response to an alarm or other event. In order to recover an unambiguous track of a particular person, the human has to traverse the graph and resolve all the ambiguities, selecting the correct continuation at each node.

The system supports this task by helping the operator navigate through space and time to quickly inspect each ambiguity, providing

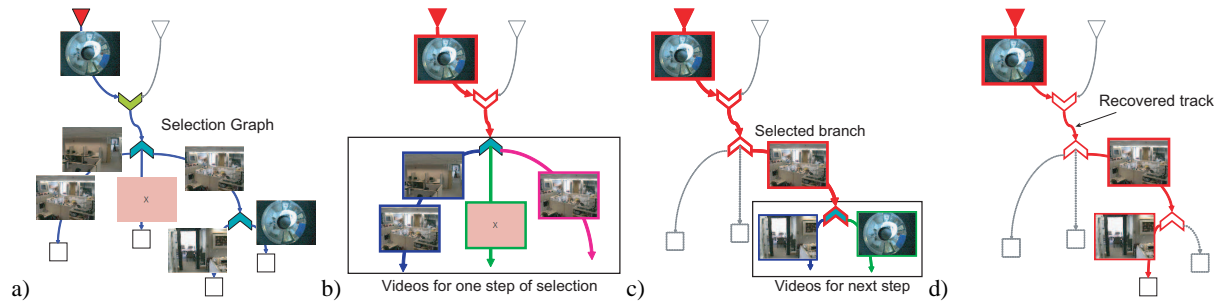


Fig. 12. Human-guided track selection process using tracklet tree representation. a) Example of the selection subgraph which includes camera views available for each tracklet, as well as split/join locations where track splicing occurs. Tracklets are shown as edges of the graph passing through the camera views. b) First step of the interactive graph pruning process. One step-lookahead tracklets are presented to the operator. c) Second step of the graph pruning. d) Final track recovered.

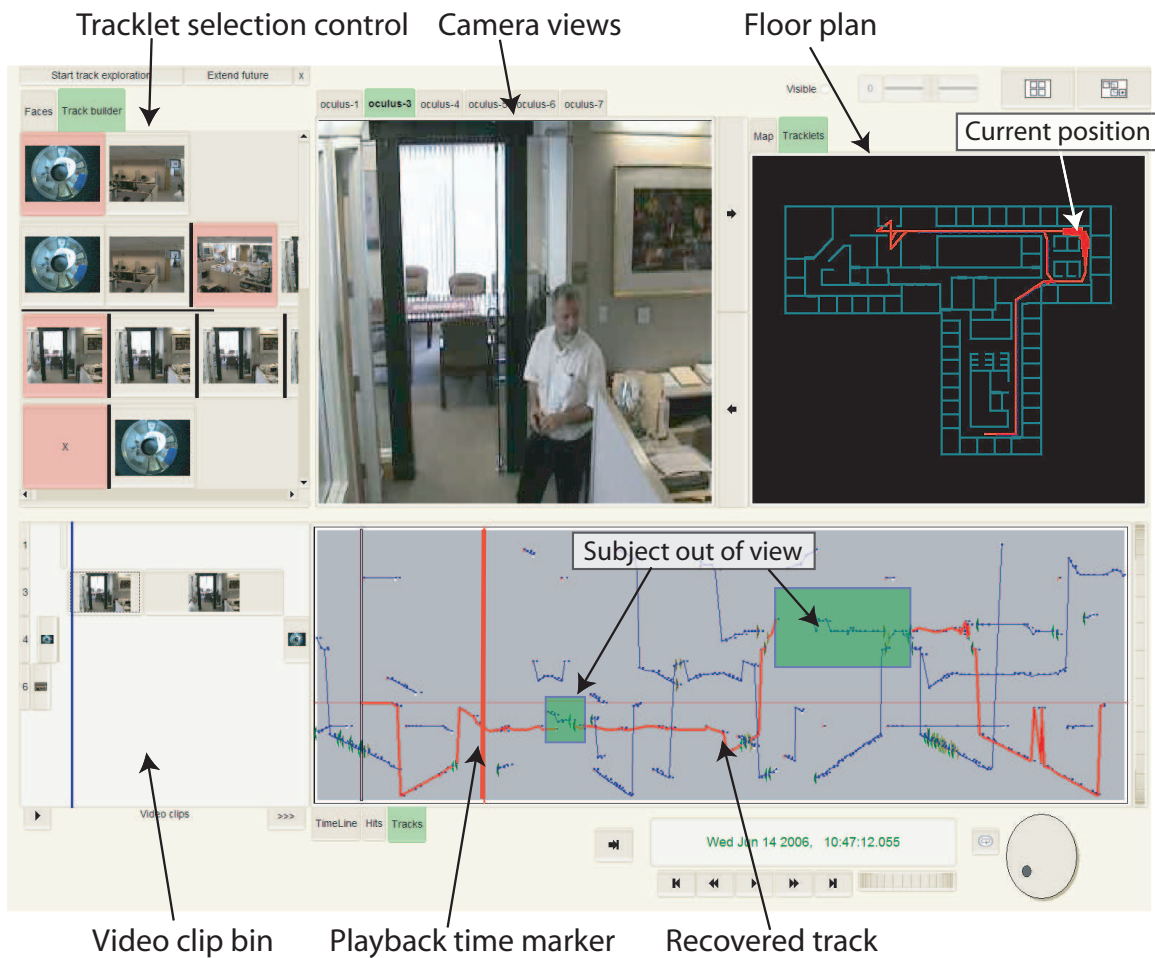


Fig. 14. User interface of the MERL forensic surveillance system. The interface includes an additional panel that allows for a visual graph traversal and track construction. Parts of the track where the subject is out of the view of the system are highlighted manually for illustration purposes.



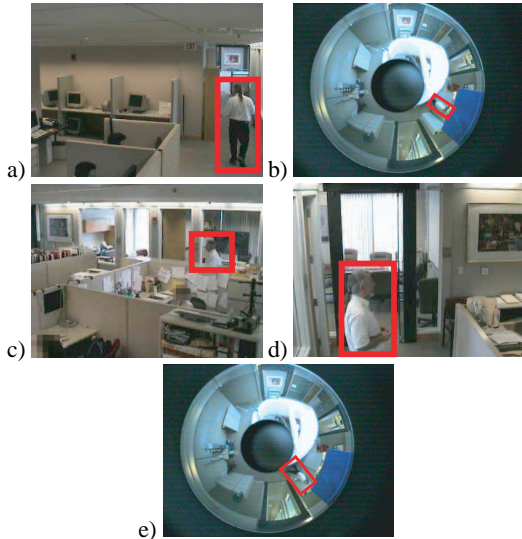


Fig. 15. Collected video clips illustrated here with selected frames. The system tracks the individual over the course of 15 minutes, which includes a large portion of the track that is not seen by cameras.

all the evidence surrounding each ambiguity in a way that makes it instantly available. For example, Figure 14 shows the spatial information displayed on the map (top right). Next to it (center top) a view from a video camera is shown, which can be selected either manually or automatically. On the top left the figure shows the current state of the graph traversal. Read top-down, it symbolically shows the current decision point in the context of past decisions (buttons outlined in red), along with embedded icons depicting any relevant video observations.

By playing or scrubbing through time, the operator can animate the tracklet swellings. In this way, we leverage the sophisticated perceptual system of the user, combined with their domain knowledge about the space to solve the difficult tracking problem. It is relatively easy for a human to see that one tracklet is “obviously” a continuation of another even when that would not be easy to detect with an algorithmic analysis.

Figure 12 illustrates the procedure of “walking a graph.” When the operator selects a starting event, that event links into the database to specify an entire graph and its associated evidence (Figure 12a). At each ambiguity the operator is presented with a set of choices, represented in Figures 12b and 12c by boxes. These choices are presented to the operator in the graph-walking tool on the left side of the interface in Figure 14. Note that if a mistake is made the operator can use this control to roll back the chain of selections to an arbitrary depth and traverse a different branch of the graph.

When the operator arrives at the termination of a tracklet, as in Figure 12d, then the story authoring is complete. The output of the process includes spatial, temporal, and evidential portions. The spatial part of the story is overlaid in red on the map, as in Figure 16. The spatial component in a single frame shows us where the story begins and ends and which parts of the building are involved. The temporal component is overlaid in red on the timeline and in the bottom of Figure 14. The temporal component reveals a snapshot of the temporal extent of the event, and if there happen to be gaps in the observations. Finally, the evidential component is a collection of video clips associated with the event, as shown in Figure 15. This video collection is automatically edited into a video summary of the event.

This set of visualizations and interface tools enables an operator to build an explanation behind an event quickly and easily. By providing the relevant information in the appropriate form, the operator can move from ambiguity to ambiguity, making only a small number of decisions.

The final assembled track, covering a 15-minute period, is shown in Figure 16 in red. The track begins outside the view of any of the cam-

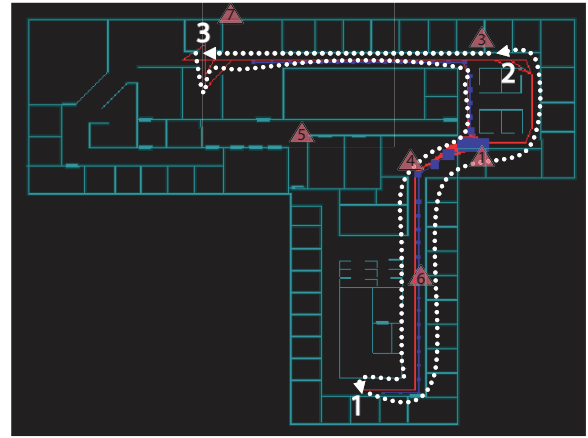


Fig. 16. Final assembled track covering a 15-minute period. Locations of the six cameras are marked by small numbered triangles.

eras on the bottom left tip of the line (point 1). It then follows through three cameras with IDs 6, 4, and 1 (Figure 15a-c, respectively) to the upper right corner, where the person stops to talk to another occupant (point 2, camera 3, Figure 15d). Then the track proceeds all the way to the left, where the person disappears out of the sensor view for about three minutes (point 3). Finally, after a short period of hovering in the left-most end of the track, the person retreats to his office via a different path, again passing in view of camera 4 (Figure 15e). Note that camera 6 did not observe the person on his return trajectory, as at that time it was pointing away from the sensors in the path.

## 6.2 Other applications

Context information from the sensors can be utilized algorithmically to automatically improve the effectiveness of building systems. However, there is also significant value in the visualizations. Anecdotally, the visualizations have already provided several insights into local behaviors around our lab.

In relation to safety, the map visualization in Figure 3 provided insight into evacuation procedures during a false fire alarm. Despite the flood of activations caused by so many people moving at the same time, it was still apparent from the visualization that most people did not exit from the nearest fire escape. Most occupants instead walked halfway across the building to a more familiar exit. This insight provides an opportunity for education and hopefully improved efficiency in the event of a real emergency. Without such a system it would be very difficult to obtain this kind of situation awareness during an actual evacuation.

The timeline visualization shown in Figure 5 has been useful in expanding general awareness of work habits. People who tend to come in late are surprised that many employees show up before their bosses. Similarly, at least one manager was surprised that people stayed so late, sometimes until 3 a.m. or 4 a.m. People are often surprised at the level of activity on weekends as well. This one visualization explains quite a lot about local work habits without the need to explore them interactively.

We expect building safety and the facilitation of emergency response to be early applications for these systems. Classical surveillance systems operate in forensic mode. After an undesirable event occurs, the surveillance system is used to recover evidence, but often hours later. Searching large databases of video data is extremely time-consuming. The automated visual routines available on the market are still quite primitive and require significant computational resources. Displays that provide seamless global context and support the sort of spatio-temporal query gestures described in Section 5.1 narrow the search domain and thereby render the whole process much more efficient.

Many environments have a heterogeneous population. For example, a health care facility may have staff who carry active identification

badges and patients who sometimes remove their tags. Intruders or wandering patients who have lost their tags are both potentially very interesting populations. Interfaces such as that described in Section 6.1 could be very useful in quickly focusing a search for a missing patient.

Similarly, in warehouse applications the heterogeneous network might combine RFID tags on crates with motion sensors to track the crates more efficiently. Instead of the dense network of RFID readers, the system could use motion sensors to provide additional information about the movement of the crates. The visualizations described in this paper could be used in systems that allow operators to quickly recover the history associated with a particular wayward crate.

These are just a few examples of ways that buildings could be safer, more secure, and better tuned to the inhabitants if there were a sensor network supplying context to such systems in the building.

## 7 CONCLUSIONS

In the course of working on the system we encountered numerous design and visualization challenges. We have addressed some issues of responsiveness, synchrony, and clarity of the display, but many other issues are still waiting for solutions.

One of the big challenges is building an intuitive interface for human-guided tracking, which we describe in Section 6. The interface for graph traversal (see top left of Figure 14) is usable but still not satisfactory. There are two related aspects of that challenge: visual and computational. The computational aspect has to do with the ordering of candidate branches in a graph for continuing a selected track. Currently, these candidates are presented in the order of retrieval from the database. However, to more efficiently utilize the limited real estate of the control, it is necessary to better rank the candidates. In order to accomplish that now (after observing the space for a year), we calculate what we call a *graph prior* – a probability that a person will take a particular route through the space. What remains to be done is to use the videos themselves to find better correspondence between a parent and child branches in the graph. We plan to do that with computer vision algorithms, using shape and texture statistics of the moving objects as well as face detection, tracking, and recognition algorithms [13]. Addressing this problem will help us design a better interaction with the graph traversal control.

The system described in this paper is the embodiment of the idea that, in reality, most information of interest is contained in the context of the entire building, for which the complexity and the bandwidth of camera-based applications are not needed. In fact, even to us it was surprising to find out exactly how much information the simple 1-bit sensors can deliver if viewed in an ensemble. It is an important validation to our idea that the tasks described here can be performed in our system flexibly, intuitively, and at a very fast pace. A central aspect of the success was the decisions made for design of visualization and interaction methods:

1. Joint synchronized use of specialized time and space controls.
2. Pursuing the pre-attentive visualization, or a “single glance” principle for estimating direction and speed of motion, as well as patterns of the temporal activity in the entire building.
3. Keeping unused elements of the user interface out of the way by using tabbed panels.
4. Removing all irrelevant data from the display. In particular, estimating the sensor visibility (see Section 5) and filtering of the video feeds to show only the frames that contain the persons of interest dramatically improved the quality of the search.
5. Automatic camera switching policy in displaying the results of the query.
6. Admitting that no matter how hard we try, we will not solve the problem of multi-camera tracking with 100 percent accuracy. This motivates a search for alternative methods to support the operator’s efficiency.

7. Using the concept of tracklets and the mechanism of human-guided graph traversal to reconstruct the actions of individuals.

We believe that we achieved a good balance in the level of visualization and control in our system.

## ACKNOWLEDGEMENTS

The authors are indebted to the anonymous reviewers for an outstanding quality of the reviews and many helpful comments that made the presentation of this paper better.

## REFERENCES

- [1] N. Andrienko, G. Andrienko, and P. Gatalisky. Towards exploratory visualization of spatio-temporal data. In *3rd AGILE Conference on Geographic Information Science*, Helsinki/Espoo, Finland, 2000.
- [2] R. Brunelli, O. Mich, and C. Modena. A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation*, 10(2):78–112, 1999.
- [3] P. Gatalisky, N. Andrienko, and G. Andrienko. Interactive analysis of event data using space-time cube. In *Information Visualization*, pages 145–152. IEEE, 2004.
- [4] M. Gelgon and P. Boutheymy. Determining a structured spatio-temporal representation of video content for efficient visualization and indexing. *Lecture Notes in Computer Science*, 1406:595, 1998.
- [5] H. B. Horton. Machine for registering music, 1855. US Patent Office 13,946.
- [6] Y. Ivanov, A. Sorokin, C. Wren, and I. Kaur. Tracking people in mixed modality systems. In *VCIP*, San Jose, CA, USA, February 2007.
- [7] Y. Ivanov and C. Wren. Toward spatial queries for spatial surveillance tasks. In *Pervasive: Workshop on Pervasive Technology Applied to Real-World Experiences with RFID and Sensor Networks*, volume EI123, May 2006.
- [8] T. Kapler and W. Wright. Geotime information visualization. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'04)*, pages 25–32, Washington, DC, USA, 2004. IEEE Computer Society.
- [9] R. Kincaid and H. Lam. Line graph explorer: scalable display of line graphs using focus+context. In *Working Conference on Advanced Visual Interfaces*, pages 404–411, May 2006.
- [10] M.-P. Kwan and J. Lee. *Geovisualization of Human Activity Patterns Using 3D GIS: A Time-Geographic Approach*, pages 48–66. Oxford University Press, 2004.
- [11] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: visualizing personal histories. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 221–ff., New York, NY, USA, 1996. ACM Press.
- [12] C. Snoek, M. Worring, J. Geusebroek, D. Koelma, F. Seinstra, and A. Smeulders. The semantic pathfinder: Using an authoring metaphor for generic multimedia indexing. 28(10):1678–1689, October 2006.
- [13] P. Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. In *IEEE International Conference on Computer Vision (ICCV)*, pages 734–741, 2003.
- [14] M. Wattenberg. Baby names, visualization, and social data analysis. In *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, page 1, Washington, DC, USA, 2005. IEEE Computer Society.