Determining Activity Patterns in Retail Spaces through Video Analysis

Andreas Girgensohn, Frank Shipman, Lynn Wilcox

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

{andreasg, wilcox}@fxpal.com

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

shipman@cs.tamu.edu

ABSTRACT

Retail establishments want to know about traffic flow and patterns of activity in order to better arrange and staff their business. A large number of fixed video cameras are commonly installed at these locations. While they can be used to observe activity in the retail environment, assigning personnel to this is too time consuming to be valuable for retail analysis. We have developed video processing and visualization techniques that generate presentations appropriate for examining traffic flow and changes in activity at different times of the day. Taking the results of video tracking software as input, our system aggregates activity in different regions of the area being analyzed, determines the average speed of moving objects in the region, and segments time based on significant changes in the quantity and/or location of activity. Visualizations present the results as heat maps to show activity and object counts and average velocities overlaid on the map of the space.

Categories and Subject Descriptors

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

General Terms

Algorithms, Design, Human Factors.

Keywords

Video surveillance, multiple security cameras, retail analysis, activity segmentation.

1. INTRODUCTION

Retail analysis involves the identification of patterns of shopper behavior that are then used to make better use of the retail space and staff. Accurate retail analysis has traditionally involved a human-intensive process of observing customers' movements to determine misused space or staff placement. RFID and other sensor technologies are currently being deployed in some environments to gain customer movement data without requiring human observation. Automatic analysis of video from installed cameras is another means for determining activity.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

Copyright 2008 ACM 978-1-60558-303-7/08/10...$5.00.

One goal of collecting data from sensors is to understand patterns of activity in the area being monitored. Such an understanding is useful for predicting future activity. For example, where activity is periodic (e.g. activity that varies in a daily or weekly pattern), identifying anomalous activity (e.g. activity outside of the norm for a given period), and for post-hoc analysis of activity.

We previously worked on a project for a large Japanese post office to determine traffic patterns of customers and employees. We have since applied our work on video monitoring for the analysis of video from multiple cameras in multiple locations [2] to that problem. We calibrated the cameras in the post office and projected

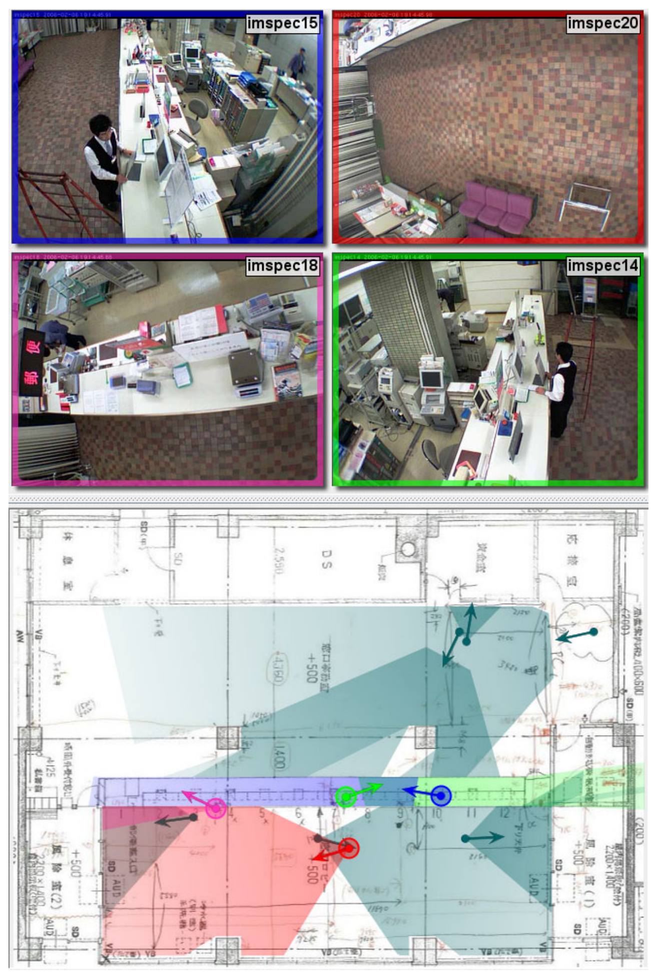

Figure 1. Camera views and floor plan of post office.

tracked people onto the floor plan (see Figure 1).

We use the object-tracking data from the video monitoring component to segment time based on activity. Once recognized, these time segments are used to generate visualizations to aid the comprehension of changes in activity over time and as indices into the source data (i.e. video). While others have visualized activity in the form of heat maps, our novel contribution is the automatic temporal segmentation based on the activity patterns detected by several cameras or sensors and the visualization of those segments.

Figure 2. The FXPAL DOTS system with temporally segmented activity heat maps.

2. RELATED WORK

Work related to supporting retail analysis using surveillance video falls into two categories: methods for segmenting video based on classification of object movement and methods for presenting overviews of activity in video.

Segmentation of video based on the motion of a single object is commonplace for video streams. Some of these segmentation algorithms classify the type of motion to segment the video [1, 5]. Analyzing quantities of stored video to learn common object motion patterns so that the motion of a specific object can be classified has also been examined [7, 9]. The goal of segmentation in the prior work is to locate and characterize activity of individual tracked objects, not to characterize the aggregate activity of all objects. As such, none of this work segments video content based on differences in aggregate behavior.

An exception is the work of Santini [6], who analyzes aggregate behavior of objects in video frames to help road traffic monitoring in metropolitan areas. There is no use of the analysis to segment time based on different aggregate behavior.

Work on presenting overviews of activity in video does have the goal of visualizing aggregate behavior. The visualization of the segmented video in our system is similar to Pingali's use of heat maps as a retrieval interface for video [4]. Pingali does not attempt to segment the video based on activity and does not present different visualizations for different time intervals, nor do their heat maps provide menus for selection among activity at a location. Additionally, our presentation of traffic counts and average velocity are not found in the above visualizations.

Prior work to support retail analysis has used the list of items at checkout to infer a path through the store for each customer [8]. Such inferred paths miss any information about items the customers looked at but did not buy and the paths are likely to differ from the actual paths followed. Larson and colleagues use RFIDs attached to shopping carts to get around such limitations and to cluster paths to identify several typical behaviors [3]. Instead of requiring additional infrastructure, our approach leverages existing infrastructure such as video cameras.

Our prior work on multi-camera video monitoring uses a geometric model of the space and object tracking in video to follow objects' movement across cameras [10]. This process results in a representation of the objects moving and their locations over time (see Figure 2). Given this, we have developed tools for supporting retail analysis without additional infrastructure.

3. TECHNICAL DETAILS

Segmenting time based on activity can be simple in a single-sensor or single-camera context. Motion in the field of view of the camera or other sensor can be quantized and that motion can be used to segment the data feed. In the case of sensor networks, such as video surveillance systems, where dozens or hundreds of cameras cover large spaces, measures of activity need to become more complicated. This is because having a single measure of activity in the space, such as the sum of activity across all sensors, misses distinctive patterns of activity, and will potentially equate 10 people moving around the area as a whole with 10 people moving in one specific area.

This system models the geographic space being monitored, in particular it divides the space monitored into a grid, and looks at changes in activity within each portion of the grid to determine time segments. In practice, this is comparing histograms of activity, where the activity value for each portion of the grid is one element in a histogram.

The approach to segmenting the period of analysis into distinctive time windows involves three stages: identifying activity in video, recording and spreading this activity in time and space, and recognizing boundaries in the activity patterns. Finally, the segments are the basis of user interface design for visualizing and accessing activity in the sensor data.

3.1 Identifying Activity in Video

Identifying and locating activity in the space covered by the surveillance video is the first step. Measures of activity could be collected from many sources, e.g. RFID events, heat or sound sensors, or pressure plates. While this system uses previously described algorithms for identifying moving objects in video and their location [10], the only requirement of the system is to have <location, time> pairs for activity in the space.

3.2 Spreading Activity in Time and Space

The activity in a space is highly variable. This variance makes the raw object counts and locations noisy in terms of trying to recognize periods of activity. To reduce the effect of this variance when trying to recognize distinctive periods of activity, the system spreads recognized activity across time and space.

To spread activity over time, the time dimension is divided into small chunks (e.g. 5 minute chunks) and any observed activity is spread over time in a one-dimensional distribution (e.g. Poisson distribution over the 30 minutes surrounding the observed event). Alternative chunk lengths, distribution functions, and distribution lengths could be used.

To spread activity in space, observed activity is allocated to the geographic elements near the observation. Our geographic model is a simple grid. To allocate activity to the grid, each observation of an object is spread over two-dimensional area with the center being the computed position of the object. If the area resides completely in one grid element then 100% of the object's activity being attributed to that element. Otherwise the portion of the area that is in each grid element is assigned to that element.

3.3 Boundaries in Activity Patterns

The recognition of boundaries can begin once the observed activity has been spread across time and space. The system identifies boundaries at intervals equivalent to the chunk size used to spread activity over time (5 minutes in our case). Multiples of this size could also be used.

The system first compares the activity in a fixed period prior to the potential boundary to the activity in the same length period after the potential boundary. This computation occurs for each grid element. The computed prior and after activity is a weighted sum of the activity allocated to a number of time windows prior and after the potential boundary, respectively. Currently, the system uses an even weighting of the six prior and following time chunks. The difference between these two values for a grid element is the change

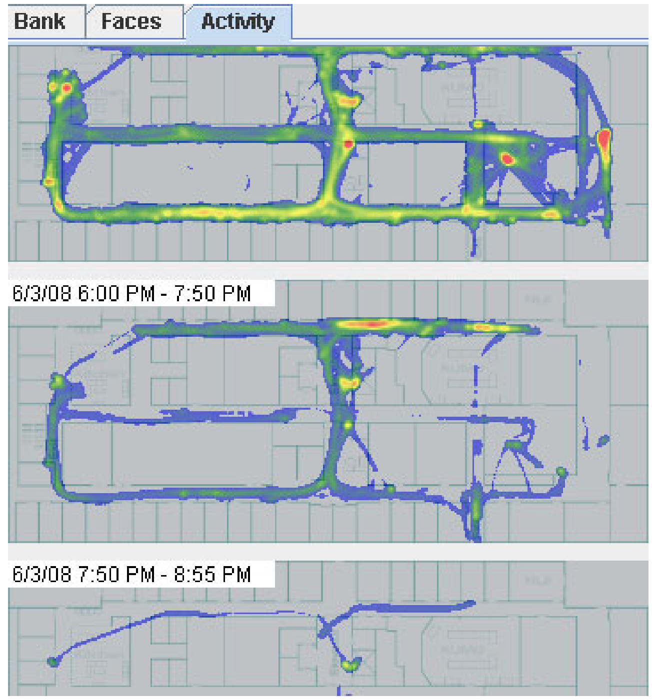

Figure 3. Temporally segmented activity heat maps.

in activity for that grid element at that time. The sum of the absolute differences across all grid elements is a measure of change in activity across the space as a whole at that time.

Once the change in activity of the whole space has been calculated at each potential boundary, we locate peaks in this measure to select boundaries. The selection of boundaries is such that no boundary can be within a predefined minimum segment length of another boundary, 60 minutes in our case. We ignore any peaks within the predefined minimum segment length of the selected boundary or within the predefined minimum segment length of a higher-level change peak that was not selected as a boundary. Once peaks have been filtered based on this rule, the time period is segmented based on the remaining peaks that are above a predefined minimum activity level.

3.4 Visualizing and Accessing Activity Patterns

Once time has been segmented into periods based on activity, these segments are used in an interactive visualization to facilitate comprehension of changes in activity over time.

The visualization consists of heat maps of activity in the different time segments. The visualization of activity in each heat map is normalized based on the length of the time segment. In this way, the same average level of activity over time between segments of different length will appear the same in the visualization. These heat maps are also used to provide indices into the video at the start of a segment and to activity at a location during a time segment.

Because of the easily accessible data from more than 20 cameras in constant operation, we performed most experiments with FXPAL's DOTS video monitoring system (see Figure 2). The heat maps are in the pane on the left. Figure 3 shows a close up of the heat maps. The time period is shown on the heat map and users can click on a point in the heat map to select cameras that can see that location

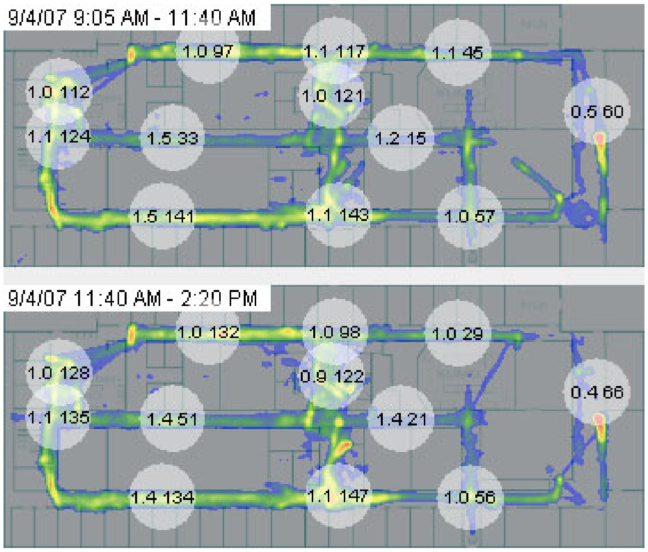

Figure 4. Overlays displaying average speeds and counts.

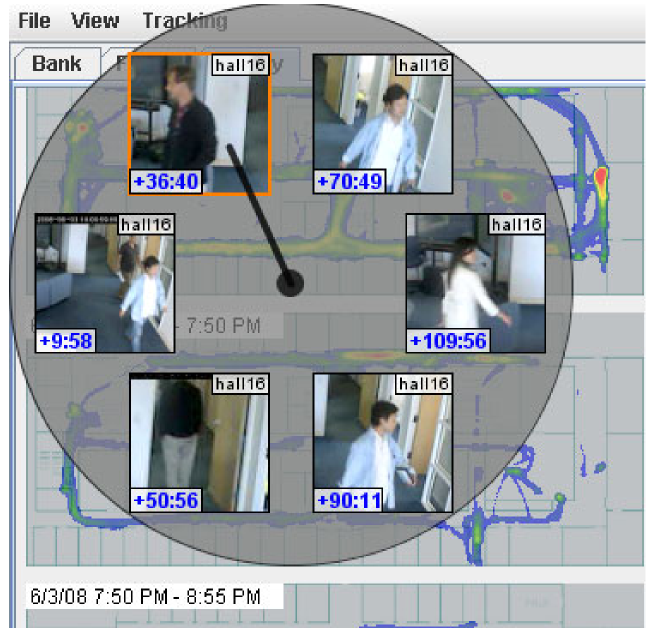

Figure 5. Activity at the location during the time segment.

and start playback of video at the beginning of that time segment. We also provide overlays for busy parts of the heat maps that display average walking speeds and the number of people passing by that point during the time interval (see Figure 4).

To understand locations of activity, users may select among the activity at a location during the time segment. To support this, the activity at that location in the time segment is clustered and presented as a hierarchy of pie menus. Figure 5 shows a menu for selecting activity at a particular location during the time period. Clustering of activity at a location can be based on time, as in this example, or based on characteristics of the activity (e.g. the person, the color of the object, the type of motion).

4. CONCLUSIONS

Presenting sensor data collected over long periods of time in a short run-time presentation requires techniques for selecting and visualizing data. By recognizing time segments with distinctive activity patterns, systems can better present overviews of activity over time. For example, the overview can be used to recognize the different patterns of motion between two time periods such as before work and lunch time. Our technique divides the area covered by sensors into regions and calculates a measure of activity for each region. When overall activity changes, or when a significant proportion of activity moves from one region to another, our approach will generate a segment boundary in the time period being evaluated. These time segments are used as both indices into the sensor data and to provide visualizations to support humans analyzing the activity. As an example visualization, we present a series of heat maps showing the activity patterns in the time periods identified by the segmentation approach. Clicking on a heat map sets the playback of the sensor data to the period of time selected and selects sensors that provide data for the location of the mouse click.

5. REFERENCES

[1] Demirdjian, D., Tollmar, K., Koile, K., Checka N., and Darrell, T. “Activity maps for location-aware computing”, Proceedings of IEEE Workshop on Applications of Computer Vision, 2002.

[2] Girgensohn, A., Kimber, D., Vaughan, J., Yang, T., Shipman, F., Turner, T., Rieffel, E., Wilcox, L., Chen, F., and Dunnigan, T. DOTS: Support for Effective Video Surveillance. ACM Multimedia 2007, pp. 423-432, 2007

[3] Larson, J., Bradlow, E., and Fader, P. “An Exploratory Look at In-Store Supermarket Shopping Patterns”, Wharton School of Business, University of Pennsylvania.

[4] Pingali, G., Opalach, A., and Carlbom, I. “Multimedia retrieval through spatio-temporal activity maps”, Proceedings of ACM Multimedia, pp. 129-136, 2001.

[5] Porikli, F. “Multi-Camera Surveillance: Object-Based Summarization Approach”, 2003. http://www.merl.com/papers/docs/TR2003-145.pdf

[6] Santini, S. “Analysis of traffic flow in urban areas using web cameras”, Proceedings of IEEE Workshop on Applications of Computer Vision, 2000.

[7] Stauffer, C., and Grimson, W.E. “Learning Patterns of Activity Using Real-Time Tracking”, IEEE Transactions on Pattern Matching and Machine Intelligence, 22, 8 (2000), pp. 747-757. (http://people.csail.mit.edu/stauffer/Home/_papers/vsam-pami-tracking.pdf)

[8] “Tag Team: Tracking the Patterns of Supermarket Shoppers”, Knowledge@Wharton, http://knowledge.wharton.upenn.edu, 2005.

[9] Xiang, T. and Gong, S. “Activity Based Video Content Trajectory Representation and Segmentation”, Proceedings of BMVC, 2004, pp. 177-186.

[10] Yang, T., Chen, F., Kimber, D., and Vaughan, J. Robust People Detection and Tracking in a Multi-camera Indoor Visual Surveillance System. ICME 2007, pp. 675-678, 2007.