Chapter 1

RECOGNIZING ACTIVITY STRUCTURES IN MASSIVE NUMBERS OF SIMPLE EVENTS OVER LARGE AREAS

Raymond D. Rimey
Lockheed Martin IS&GS

Abstract: Newly emerging kinds of sensors are able to observe movement of vehicles and people in large areas over long time intervals. While some sensors may appear very different, such as a ground sensor network and a downlooking wide-area motion imagery sensor, they have common problems, generally involving the recognition of activity structures in massive numbers of observations, usually movement-derived events. This paper presents a taxonomy of different kinds of activity structure, and illustrative examples of visualization techniques or automated exploitation algorithms to help expose and understand those activity structures.

Key words: Activity recognition; automated exploitation; persistent surveillance; ground sensor network; wide-area motion imagery.

1. INTRODUCTION

Consider a ground sensor network that contains a large number of simple sensor nodes spread over a large area. A single node contains a relatively simple sensor that provides minimal information in the immediate vicinity of the sensor, for example the sensor might detect when motion occurs anywhere within a few meter wide circle. The simplicity of the sensor data is compensated for by having a very large number of sensors, all networked together. The sensor network spans an area large enough to contain many people performing many independent activities, for example an area containing a few buildings or an area containing a small town.

A fundamental problem is knowing what quality and quantity of sensor network observations are necessary to discriminate activities of different complexity. It seems logical that simpler observations, such as a single bit for detected movement at one time instant, will require more and denser sensor nodes in order to recognize an activity from that data. And more complex observations, such as adding one bit (or more) describing attributes of the entity whose movement was detected, will require fewer sensor nodes to recognize the same activity. An additional issue is that the “complexity” of different activities needs to be described, enabling researchers to eventually relate the complexity of an activity with the quality/quantity of sensor observations needed to recognize it.

This paper makes two contributions. (1) A methodology is presented for describing the “structure” of complex activities that span large space-time volumes, including: spatial structure, temporal structure, event-linkage structure, short event-sequence structure, and network structure. These structural elements enable descriptions of complex activities, and suggest exploitation capabilities that a human analyst may find useful. (2) Examples of several techniques are presented for recognizing complex activity patterns given large numbers of simple events from a ground sensor network, while leveraging the different types of activity structure. These techniques are also relevant to events extracted from wide-area motion imagery.

The following types of activity structure are depicted in Figure 1.

- <u>Spatial Structure</u> refers to spatial distributions and spatial arrangements of events within an area, typically integrated over a time interval. Think of this as spatial arrangements of events within the disk shown in the figure.
- <u>Temporal Structure</u> refers to temporal distributions and spatial arrangements of events within a time interval, typically integrated over a small area. Think of this as temporal arrangements of events within the vertical timeline in the figure.
- <u>Event-Linkage Structure</u> refers to commonly occurring relationships between pairs of events (and/or locations) within a small space-time volume. Think of a complex activity as a node-link graph embedded in a space-time volume, and the building blocks for that structure are the links consisting of two events and a relationship. Think of this as a single link connecting two points in the figure (events at two locations at different times).
- <u>Short Event-Sequence Structure</u>: Many activities performed by humans involve sequential structure, meaning multiple short event sequences embedded within a larger more complex structure. We distinguish two cases: Sequences of events occurring essentially at the same spatial location (so they are embedded in a timeline), and sequences of events occurring in a small space-time volume (so spatial location must be encoded).

- Network Structure: Humans are creatures of habit, and in particular humans have regular transportation routines. Network structure refers to the fixed set of locations that are regularly visited (network nodes) and the regular movement between those locations (network links). Networks provide a framework within which some high-level activities occur.
- Loose Structure: Many higher-level activities lack a large-scale or highly regular structure, but often contain fragments of the different activity structures above within a large space-time volume.

Intelligence analysts today are just starting to think in terms of these kinds of activity structures. New research and new algorithms are needed to help a human understand these activity structures.

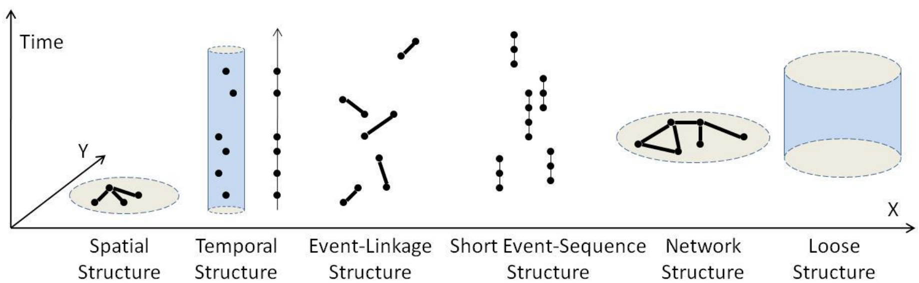


Figure 1-1. Complex activities that span large space-time volumes contain these six kinds of "activity structure". New automated exploitation algorithms and visualization techniques that expose such structure will enable analysts to better understand activities observed via wide-area persistent surveillance sensors.

2. SPATIAL STRUCTURE

When assigned a new unfamiliar area to study, an analyst needs to understand the spatial distribution of events and activities in the area, and the analyst needs to understand more detailed spatial arrangements of events that occur in the area. An understanding of historical spatial structure provides the foundation for recognizing activities in instantaneous sensor data. New automated exploitation algorithms and associated visualization techniques can form the basis for new tools that help an analyst to understand spatial activity structures.

Figures 2 and 3 show examples [5] to illustrate the value of automated exploitation algorithms to help understand the spatial structure of activity in an area. These experiments used simulated people within a real neighborhood near Al Mahmudiyah, Iraq, shown in Figure 2(a). The simulation involved 26 agents with 145 agent behavior rules, living within 52 distinct areas within the neighborhood over 56 days. The 178 nodes of a ground sensor network embedded within this area contained sensors that detect movement in their immediate vicinity and optionally several attributes of the moving entity. Figure 2(b) shows a heatmap, a color-coded display of a surface derived from the sensor network observations that covers the neighborhood. This example surface shows the total number of movement-related events across the neighborhood during a specified time interval. Algorithmic filters can optionally be applied to generate additional surfaces. The spatial distribution of events is typically a function of many factors, for example Figure 2(c) shows how the distribution changes over several time slices of one day.

An analyst can subtract heatmaps for two time intervals to reveal areas where activity patterns have changed. Figure 3 shows what happens during lunch. “Came From” are work locations. “Went To” are lunch locations. This kind of analysis begins to expose interesting locations in the neighborhood, and movement patterns between them, elements of the spatial structure of activity in this area. Given a select set of candidate locations, we have also developed algorithms [5] that detect deviations in event statistics for each area in order to flag them for further study by an analyst.

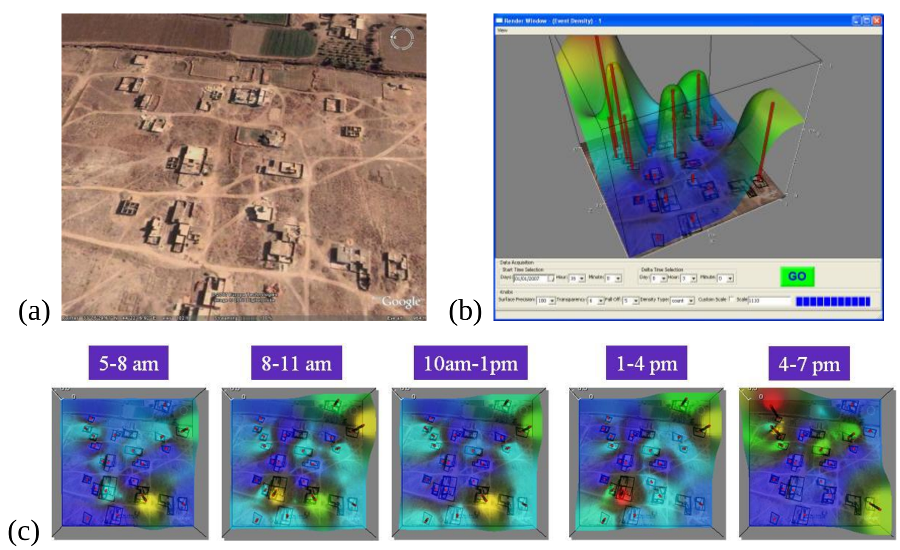


Figure 1-2. (a) Area covered by sensor network. (b) Visualization tool for creating 3D heatmaps. (c) Example showing spatial structure of activity during several time intervals.

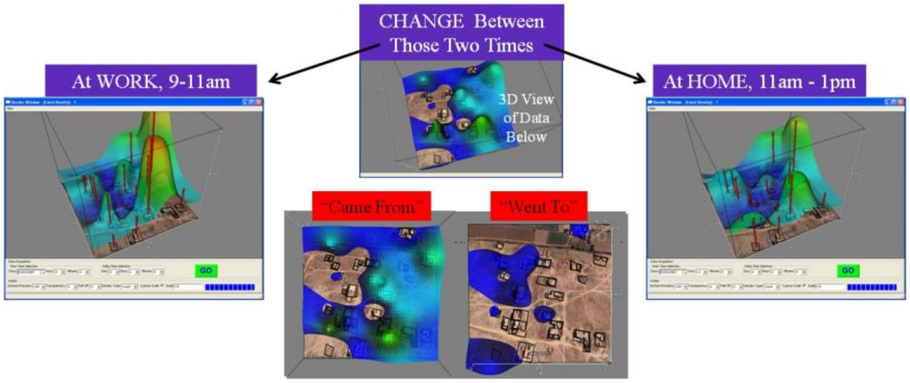


Figure 1-3. Heatmaps for two time intervals (shown on either side) are subtracted (shown in the middle) to reveal the structure of lunchtime movement.

3. TEMPORAL STRUCTURE

Analysts need tools that help them understand the temporal structure of activity in an area. For one interesting location identified from the above heatmap analysis, an analyst may want to see how activity levels vary over time. Figure 4 provides an illustration of the idea [3]. A sensor network with passive infrared motion detectors was installed inside a large campus building (this same idea extends to a sensor network outdoors covering an entire neighborhood). Figure 4 shows the amount of movement inside the building over an entire semester. Different time intervals are shown from left to right: 2 hours, 1 day, 1 week, and 1 semester. The density map contains many tiny columns that each denote the reading from a single sensor node, and the tiny rows are time steps, so the brightness of the tiny dots depicts the amount of movement in different areas over time. The point is that some fundamental activity patterns inside the building are visually obvious within a timeline visualization. A similar analysis can be performed for specific locations inside this building.

Algorithms can expose more complex structure than is visible in raw sensor data, helping an analyst to understand the higher-level or semantic structure of activity in a space-time volume. Figure 5 shows an example [6]. This experiment used 74 days of webcam video (one frame every 90 seconds) of a market square in Germany. The spatial area of the market square was manually populated with 18 polygonal areas, and change events (described as one of four sizes) were automatically detected in those polygons over 74 days. Probabilistic latent semantic analysis (PLSA) was applied to bags (histograms) of short event sequences to discover semantic categories of activity. This is an unsupervised technique to discover the semantic categories of activity (five categories were discovered), so a human studied the categories to assign textual names to them. Figure 5 shows the dominant semantic categories of activity assigned by the system for 14 days of activity in the market square. These days had been automatically assigned overall to the "Normal" (i.e., most common) activity category. The point is that higher-level temporal structure of activity is visible over the hours of each day. Algorithms that reveal such high-level temporal structure can help improve an analyst's understanding of any area being studied.

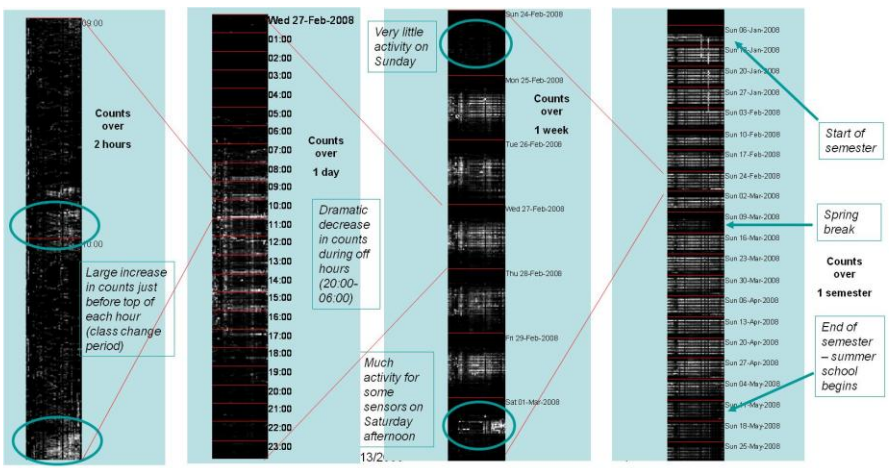


Figure 1-4. Visualization showing temporal structure of activity in a building. Rows are time steps. Each tiny column is one node in a network of passive infrared motion detectors.

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14
06-07						MF				MF	F		MF	
07-08					0	CC	0			0				
08-09				0	F		CC							
09-10		F		0			0			MF				
10-11							H							
11-12							0						H	
12-13							CC			H		MF		
13-14										CC				
14-15									CC					
15-16				CC						CC				
16-17				CC										
17-18										F				
18-19		CC		H						0				
19-20				0			0			0	0	0		F
20-21				CC						0	F	MF		

Normal · H Holiday · MF Mini-Fair · CC Crazy Coffee · F Fair · 0 No Events

Figure 1-5. Higher-level temporal structure of activity for 1-hour clips (rows) over several days (columns) in a market square. The color of each cell denotes the semantic activity category automatically assigned to that time interval.

4. EVENT-LINKAGE STRUCTURE

Analysts need tools that help them understand the "event-linkage" structure of activity in an area. Such tools are heavily dependent on algorithmic processing of large sets of events. The idea of event-linkage includes (a) when one event is followed by another specific kind of event and (b) when activity patterns in one area are correlated with activity patterns in another area. An example of the first case is adjacent events within the ngrams discussed in the next section. An example of the second case is shown in Figure 6. This example uses the same sensor network dataset as in Figure 2 for the neighborhood near Al Mahmudiyah, Iraq. The analyst has specified several areas of interest (buildings in the neighborhood) and a time interval, and the visualization in Figure 6 shows how correlated the activity is between those areas. The bottom of the 3D area inside the tool depicts a 3x3 table, because in this case the analyst selected 3 areas (buildings). The vertical axis is delta time. The size of each sphere indicates how correlated events are between two areas with a specific delta time between the activities in the two areas.

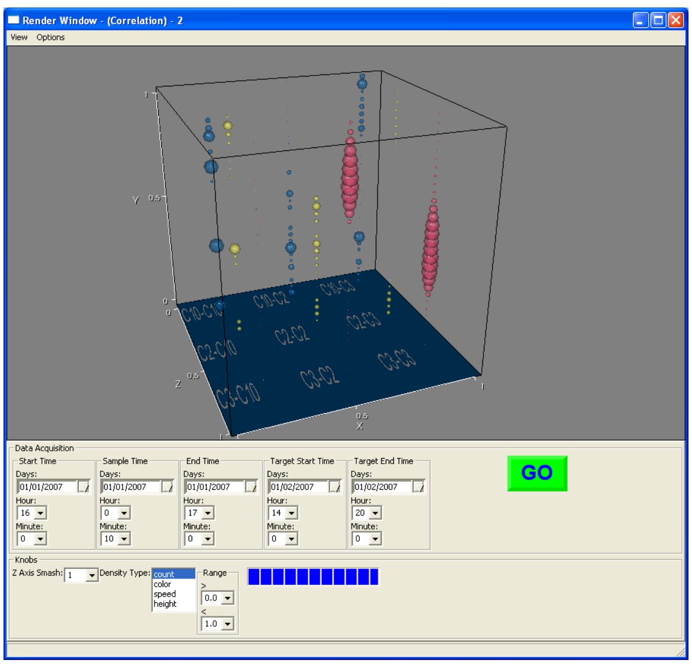


Figure 1-6. This visualization depicts time-shifted correlations between events at three different locations (buildings), in order to help an analyst understand the event-linkage structure of activity in the area (neighborhood) being studied.

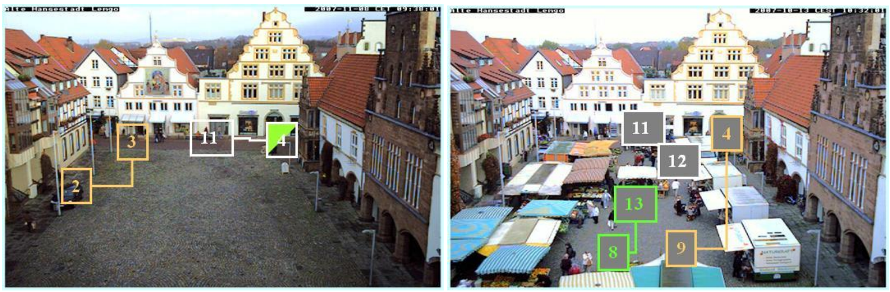
5. SHORT EVENT-SEQUENCE STRUCTURE

After binary event pairs, the next higher step is n-ary event structure, and this section focuses on n-ary structure where time dominates more than space. Two cases can be distinguished: (a) a sequence of events that occurs at a single location, so location does not need to be part of the description, and (b) a sequence of events that includes a small amount of spatial information but the temporal information is more important. Two illustrative examples follow.

The algorithms for the first example [6] are summarized in Figure 7. Change events are detected in the image sequences within pre-specified polygonal regions. A vocabulary of "visual words" or symbols is constructed from the change event attributes plus the polygon indices, e.g., the symbol G might denote event type T4 in polygon P7. The sequence of symbols generated over time is transformed into a stream of ngrams (subsequences of length n). Finally, a histogram (called a "bag") of the most common ngrams is constructed over some time interval, such as an hour or a day. The probabilistic latent semantic analysis (PLSA) method is applied to the bags of ngrams to discover semantically meaningful activity categories (also used in the example in Figure 5).

Figure 1-7. The "bag of ngrams" description is a histogram of length-n event subsequences.

An analyst needs tools to help understand the meaning of discovered activity categories, which initially are represented only by category indices. A natural way to describe activity categories is by using the event stream symbols themselves since their meaning is mapped to easily interpreted events. We adapted several methods from [2]. The original symbol streams in each category can be viewed as variable-length Markov chains, and a motif is one of the variable memory elements of the chain. The top motif description for a category is the sequence that is predictive for the category (meaning the next symbol in the sequence after the motif can be predicted with confidence) while not being predictive for other categories. Longer motifs are favored. Motifs are ranked for each category according to an optimization function. Other description methods are the top ngram (the ngram that has the highest frequency in a category) and the unique ngram (a top ngram that is also not seen in other categories). Figure 8 shows examples of these description techniques for two activity categories discovered in the market place.



(a) (b)

Figure 1-8. (a) The "Normal" activity category is described by the following short-event sequences: The top motif is 4L-4L, which may be vehicle traffic along a road across the top of the market square. The top ngram is 4M-11M, which seems similar to the top motif. The unique ngram is 3M-2L, which may describe a regular delivery truck to one store. (b) Descriptions for the "Fair1" category are: The top motif is 13M-8M, describing groups of people walking past a line of stalls. The top ngram is 11M-12M, similar seems similar. The unique ngram is 9H-4H, whose meaning is unclear. The numbers are polygon indices. M, L, H denote medium, large, huge sized change events.

The second example [3] of short event-sequence structure involves 2D binary templates that describe a local set of events (columns) in a short time window (rows). The data in this example (same as Figure 4) is from 50 passive infrared motion sensors within a large campus building. Local movement patterns appear as structures in the sensor hit matrix, Figure 9. For example, a person walking down the hall causes consecutive hits in adjacent sensors, which appears as a diagonal streak in the data. Of course, the pattern would only take the form of a diagonal streak if the sensors along the hallway are numbered in consecutive order; if the sensors are numbered in a different order, then the pattern would take a different but predictable form.

Correlation methods were applied to historical sensor network data to discover the neighborhood (2 to 5 sensors) that is adjacent to each sensor in the network. Clustering methods were applied to detect local repeated activity patterns. These are localized in space (in terms of the neighborhood of sensors) and time. Figure 10 shows the cluster results for neighborhood 50. Looking at the meaning of these clusters, we can interpret C1 as a person walking to the left in the hallway. C2 and C3 represent a person walking to the right (at slightly different speeds). Similar local activity primitives can be discovered using PLSA methods.

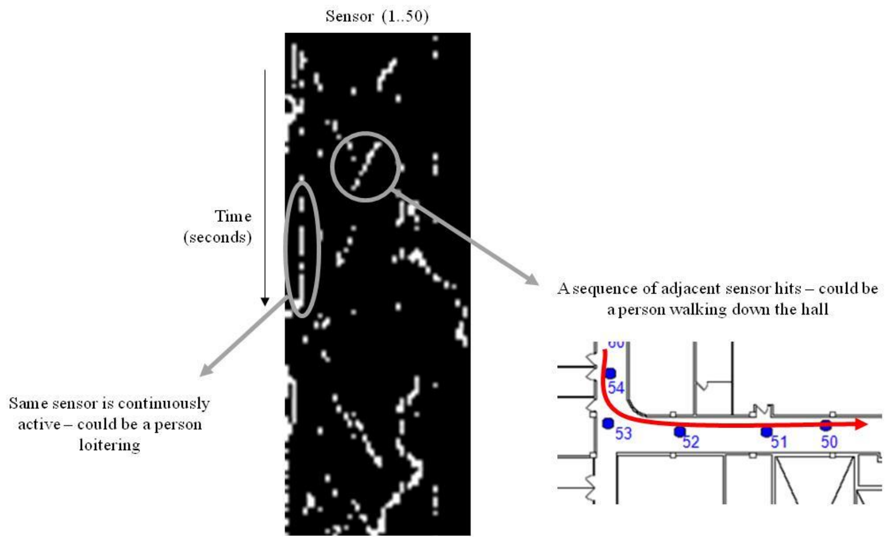


Figure 1-9. Example of sensor network data in the form of a two dimensional matrix.

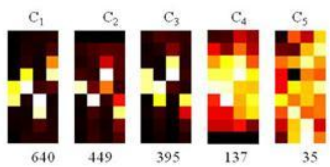


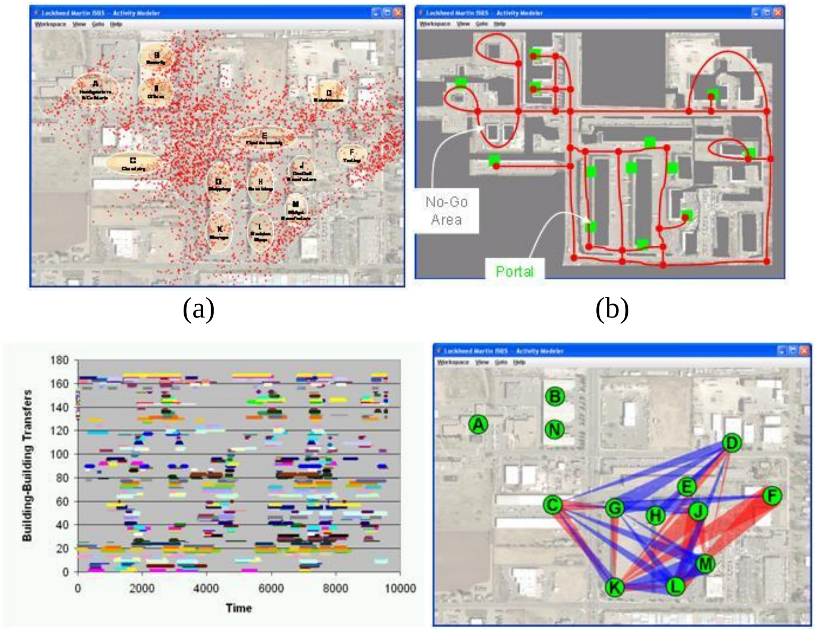
Figure 1-10. Short event-sequence structure discovered for the neighborhood centered at sensor 50, with number of occurrences in each cluster.

Short event-sequence structure, described for example via ngrams or 2D templates, provides a fundamental language to help an analyst study the details of complex activities in large areas.

6. NETWORK STRUCTURE

Human activity often involves regular transportation routines, which leads to network-based structure for describing complex activity. Nodes in a network are locations visited, and links represent movement between those locations. Separate links may represent separate routes between two locations, or a single link may denote all movement between those two locations. Network detection includes the following sub-problems. Automated exploitation algorithms and associated visualization techniques are needed so help analysts work on these problems: Node Discovery

(Discover new candidates for nodes), Node Linking (Describe significance of a link between two nodes), Node Monitoring (Reject candidate nodes; Detect interesting activity in a node's area; Detect vehicles departing / arriving; Characterize activity within each node's area).



(a) (b)

(c) (d)

Figure 1-11. (a) Short tracks are constructed from these raw GMTI dots (simulated data). (b) "no-go" graph used to constrain interpretation of low quality tracks. (c) Individual transits detected between buildings. (d) Links in network derived from repeated transits.

An example problem is to discover links between the buildings in Figure 11(a) from extremely noisy short tracks derived from a GMTI sensor (longer tracks extracted from wide-area motion imagery is an easier problem). Our method utilizes a "no-go graph", Figure 11(b), which is derived from the areas where a vehicle cannot drive through, so this provides a constraint on the interpretation of the raw movement evidence [7]. Figure 11(c) shows all the movements detected between pairs of buildings. Figure 11(d) shows the links discovered between buildings -- thicker lines denote a strong link, and the pointy direction of lines denotes the direction of the link.

Some high-level activities occur within the context of a network, so knowing the network structure is another building block toward understanding those higher-level activities. Many other high-level activities lack a single global structure, but instead contain local fragments of the other kinds of activity structures described in this paper. Visualization tools and automated exploitation algorithms that help an analyst see and understand such component activity structures will help the analyst to understand larger-sized and more loosely structured high-level activities of interest.

7. SUMMARY

Key work related to the methods illustrated in this paper include visualizations on movement inside a building [4], VLMMs for event sequences [1,2], and topic space descriptions such as PLSA [8]. Additional key work is referenced within [3,5,6,7].

This paper has presented a taxonomy of different kinds of activity structure, and illustrative examples of automated exploitation algorithms or visualization techniques to help expose and understand those different kinds of activity structure. This work helps suggest areas for future research and new capabilities that would benefit analysts.

ACKNOWLEDGEMENTS

This work was supported by Internal Research & Development projects within Lockheed Martin IS&GS.

REFERENCES

[1] A. Galata, A.G. Cohn, D. Magee, D. Hogg, “Modeling Interaction Using Learnt Qualitative Spatio-Temporal Relations and Variable Length Markov Models,” ECAI 2002.
[2] R. Hamid, et al., “A Novel Sequence Representation for Unsupervised Analysis of Human Activities,” AI Journal, 173(14), 2009.
[3] W.A. Hoff, J.W. Howard, “Activity Recognition in a Dense Sensor Network,” First International Conference on Sensor Networks and Applications, 2009.
[4] Y.A. Ivanov, C.R. Wren, A. Sorokin, I. Kaur, “Visualizing the History of Living Spaces,” IEEE Transactions on Visualization and Computer Graphics, 13(6):1153–1160, 2007.
[5] R.D. Rimey, D. Keefe, “Exploitation of Massive Numbers of Simple Events (Poster),” 37th Applied Imagery and Pattern Recognition Workshop, 2008.
[6] R.D. Rimey, D. Keefe, “Temporal Structure Methods for Image-Based Change Analysis,” 37th Applied Imagery and Pattern Recognition Workshop, 2008.
[7] R.D. Rimey, W. Hoff, J.Y. Lee, “Recognizing Wide-Area and Process-Type Activities,” International Conference on Information Fusion, 2007.
[8] S.-F. Wong, T.-K. K., R. Cipolla, “Learning Motion Categories Using Both Semantic and Structural Information,” CVPR 2007.