MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Ambient Intelligence As The Bridge To The Future of Pervasive Computing

Christopher Wren, Yuri Ivanov

TR2008-079 December 2008

Abstract

One prediction about this future of pervasive technology is that people will carry the tools needed to interface with technological resources sprinkled through out the environment. A problem with this vision is the dark side of the network effect: early adopters will end up carrying around interfaces for technology that largely does not yet exist, and building managers will question the value of installing technology with features that almost no one will be able to use. An intermediate solution is that certain buildings with specific needs for efficiency or security (such as hospitals) may become smart, with technology insinuated into particular spaces. Since many, or even most of the people in these spaces will not have the technology to interface directly with the new pervasive resources, we must think of the interaction idiom as initially being closer to the notion of smart environments. These environments will have to sense, interpret, and facilitate the actions of the inhabitants, possibly with very little help from technology attached to the people involved, or even their cooperation. We survey a body of work on perceptual tools for smart buildings, built on the sensor network model, and focused on the idea that statistical methods and population dynamics can provide valuable information even in situations where detection of individual instances of behavior may be difficult to detect. These are some of the tools which will fuel the building optimization applications that will justify the efforts of early adopters to build smart buildings studded with pervasive technology.

IEEE Int'l Conf. on Automatic Face and Gesture Recognition

Ambient Intelligence as the Bridge to the Future of Pervasive Computing

Christopher R. Wren

Yuri A. Ivanov

Mitsubishi Electric Research Laboratories
201 Broadway; Cambridge MA USA
http://www.merl.com/projects/ulrs

Abstract

One prediction about the future of pervasive technology is that people will carry the tools needed to interface with technological resources sprinkled through out the environment. A problem with this vision is the dark side of the network effect: early adopters will end up carrying around interfaces for technology that largely does not yet exist, and building managers will question the value of installing technology with features that almost no one will be able to use. An intermediate solution is that certain buildings with specific needs for efficiency or security (such as hospitals) may become smart, with technology insinuated into particular spaces. Since many, or even most of the people in these spaces will not have the technology to interface directly with the new pervasive resources, we must think of the interaction idiom as initially being closer to the notion of smart environments. These environments will have to sense, interpret, and facilitate the actions of the inhabitants, possibly with very little help from technology attached to the people involved, or even their cooperation. We survey a body of work on perceptual tools for smart buildings, built on the sensor network model, and focused on the idea that statistical methods and population dynamics can provide valuable information even in situations where detection of individual instances of behavior may be difficult to detect. These are some of the tools which will fuel the building optimization applications that will justify the efforts of early adopters to build smart buildings studded with pervasive technology.

1. Introduction

The first smart buildings already exist. Air conditioning systems, automatic window shades, and lighting respond automatically to solar load. Security systems annotate surveillance video with access control and sales data. Elevators account for patterns of demand to improve efficiency. Oddly, these buildings are so far largely insensitive to the fact that there are people in them. The people are sensed indirectly at best, through card swipes, elevator calls, or unexplained heat load.

The next generation of smart buildings will have radio networks, not to support pervasive computing, but because it is cheaper to reconfigure building services in software than it is to bring in an electrician. They will have dense arrays of environmental sensors not to implement smart environments, but because the sensors cost less than the energy they save by enabling fine-grained, efficient environmental control. They will eventually have simple networks of sensors capable of sensing humans and their actions, not because we in the research community think they're a good idea, but because sensing humans will be the next logical step in optimizing the productivity and efficiency of the building... in short because they will pay for themselves.

One way to understand this prediction is to say that the slow, logical, incremental progress in building technology will asymptotically approach the vision of the ambient intelligence movement: a sensor-rich, computation-rich environment that perceives the user, anticipates their needs, and acts to meet those needs like a good butler, that is without necessarily having to be told what to do [4].

The prevailing vision of Pervasive Computing is more explicitly interactive. Weiser's tabs, pads, and boards provide the ability to directly query the environment [8]. Projects like the Personal Server [7] imply the ability even to co-opt technology in the environment for personal use. While the visions of the Pervasive Computing movement is more expansive, it places a high threshold on the critical mass required for the network effect to provide any positive value. The Personal Server isn't useful unless there are displays out there that will allow you to use them. Displays are installed by businesses to convey targeted information and for branding purposes. The display owners value those messages and are unlikely to allow them to be brushed aside unless it is very clear that doing so will lead to increased profitability. It's hard to see how that will happen until there is a significant population of people walking around with Personal Servers who have demonstrated that they will, directly or indirectly, pay for the privilege.

978-1-4244-2154-1/08/$25.00 ©2008 IEEE

We think there is a potential bridge between these phases of pervasive technology deployment. The smart environments that are deployed primarily for the benefit of the building owners and operators will likely generate information that is also valuable to the people in those buildings. The sensors and analytic tools will exist. The incremental cost of enabling end-user applications within the organization becomes vanishingly small. The data simply needs to be made available on the organizations intranet.

It is likely that the information will move around the building, at least in part, over wireless links. The path to the user terminal, be it a cell phone or some other device, will therefore already be in place. The incremental cost of enabling end user services for anyone in the building will eventually be simply a software patch.

Once the smart environment becomes sophisticated enough, the value of the new applications will quickly outweigh these costs, and the environment will then engage the network effect that will accelerate development.

The challenge is that the sensor infrastructure will be tuned to the concerns of the building management, not the applications that those first end-users might actually care about. The sensors are likely to be as cheap as possible. Elevators don't care who you are, just where you're going. Air conditioners don't care what color shirt you're wearing, only where you are in the room. Economic, technological, and privacy considerations [5] will conspire to ensure that the sensors are as simple as possible.

Conner, *et al.* suggest that simple sensors, considered individually, might be enough to enable valuable applications [1]. However, it has become clear that a much richer set of applications can be enabled by considering the network of sensors as a whole, and leveraging the full spatio-temporal structure inherent to the data.

In the next section we discuss a public dataset that consists of one year of motion sensor data that we have released to facilitate work on this challenge. In Section 3 we present examples to illustrate some of the basic spatio-temporal patterns. Finally in Section 4 we sample the literature on spatio-temporal pattern recognition using this data set and consider the applications that might be enabled.

2. Public Data Set

At the Mitsubishi Electric Research Laboratory (MERL) we have been collecting motion sensor data from a network of over 200 sensors since March 2006. In 2007 we released a public data set containing well over 30 million raw motion records, spanning a calendar year and two floors of our research laboratory. We believe it presents a significant challenge for behavior analysis, search, manipulation and visualization. We have also prepared accompanying analytics such as partial tracks and behavior detections, as well as map data and anonymous calendar data marking the pattern of meetings, vacations and holidays.

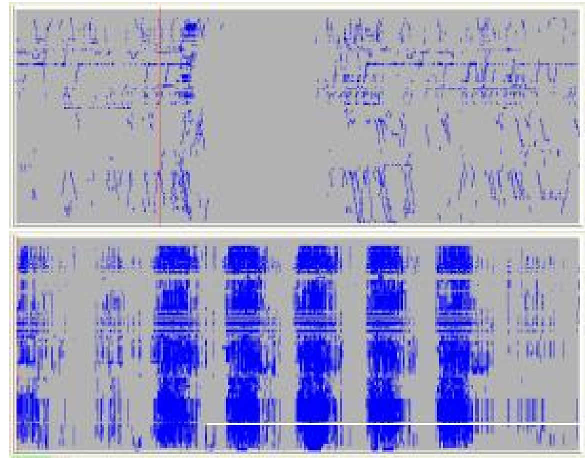

Figure 1. Time along the horizontal axis, space along the vertical axis, dots represent motion detections. **Top**: Normal activity interrupted by an evacuation. **Bottom**: A week showing diurnal and weekly patterns of activity.

This data is the residual trace from the people working in the research laboratory. It contains interesting spatio-temporal structure ranging all the way from the seconds of individuals walking down hallways, the minutes in lobbies chatting with colleagues, the hours of dozens of people attending talks and meetings, the days and weeks that drive the patterns of life, to the months and seasons with their ebb and flow of visiting employees.

2.1. Raw Motion Data

Individual wireless sensors detect the movement of people using passive infrared motion detectors. An individual or small group walking past the sensor will generate a single activation, recorded as an association between a sensor ID and a timestamp. The sensors are placed densely in the hallways and lobbies of MERL, spaced approximately two meters apart. As people move through the network, a sequence of sensor activations is recorded. It is this spatio-temporal structure, more than the individual sensor activations, that gives meaning to the data.

Even when people stand still under a sensor, their small movements will cause the sensor to trigger repeatedly, giving the data a different, distinct spatio-temporal structure.

It is interesting to consider the details of individual movements, such as in the top plot of a fire evacuation in Figure 1 (normal activity is apparent both before the evacuation and after the re-population). It is also fruitful to consider the larger structures of populations over longer spans of time, such as the bottom plot that illustrates the differences between day and nights (vertical bands) as well as the difference between work days and weekends.

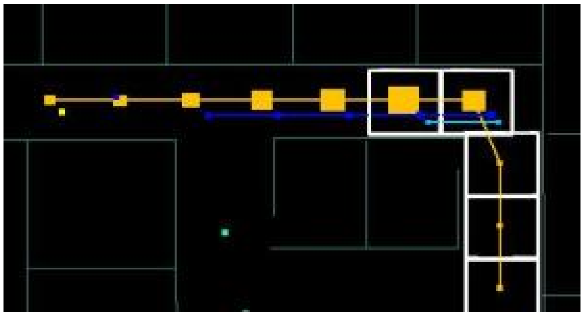

Figure 2. A sequence of motion activations (orange boxes) grouped together by their spatio-temporal proximity to each other and their isolation from other distracting.

2.2. Tracklets

When an individual moves through the space they create a structured pattern of activations that link parts of the space together in a meaningful way. Tracking is the process of recovering this structure from the raw observations. In Figure 2 we see a set of activations linked together into a *tracklet*.

2.3. Tracklet Graphs

There is an inherent ambiguity in motion sensor data. A one-bit motion sensor cannot identify individuals, or even distinguish between individuals and small groups. It is therefore impossible to track individuals through the space without some degree of ambiguity. A tracklet is a small section of a track that can be recovered unambiguously. This dataset includes a forest of graphs that represents all the known tracklets as well as the ambiguity relationships between them. All true tracks will be embedded in a graph, but each graph may allow many valid interpretations. Figure 3 is a schematic of a simple tracklet graph.

2.4. Reach

It is possible to walk the tracklet graphs to discover the possible pairings of track starts and ends, for example the simple graph in Figure 3 implies: $A \prec Y$, $B \prec Y$, $A \prec Z$, and $B \prec Z$. One possible use of this data is to estimate the probability that a trip beginning in one location will end at another location by accumulating evidence over a span of time, such as a day or a week. Figure 4 is an illustration of such an estimate.

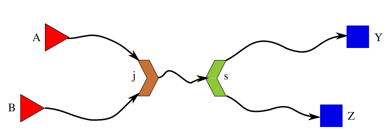

Figure 3. Tracklet Graph: people start at the triangles and end at the squares, passing through ambiguities along the way.

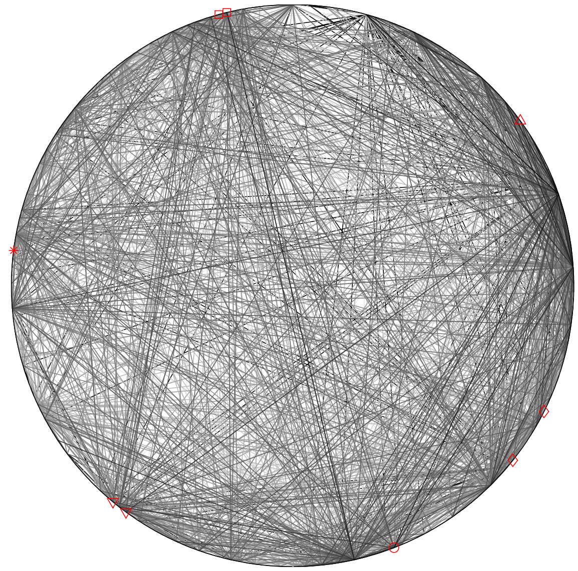

Figure 4. Dark lines indicate higher probability of a connection between localities in the space, which are represented as positions around the circle.

2.5. Symbolic Data

The dataset also includes a calibration file that associates the sensor IDs to a map of the lab. This grounds the data to the spatial context of the lab. Temporally, the data is placed in context by data from several calendars. These records indicate the times and locations of various meetings and gatherings, the dates of official holidays, and a record of the number of people who were out of the office on given days. Also included, is a daily almanac of the weather conditions in Cambridge, Massachusetts where the lab is located.

2.6. Using the Data

We invite you to download the data and apply your analytic, visualization, and interface tools. To request access, please visit `http://www.merl.com/wmd/`. More details and a video are also available at that location.

Papers that utilize this dataset must reference the technical report describing the data [11].

3. Examples of Spatio-Temporal Structure

The sensors detect fast changes in the thermal background radiation over a small region (approximately a cone 2 meters in diameter on the floor). If there is continuous motion under a sensor (such as a large group of people walking by), then the sensor will fire repeatedly, about once every 2 to 3 seconds. If people are standing or sitting under a sensor the sensor will still activate, but less often. This is due to people adjusting their stance, gesturing, or performing self-adjustment motions. Someone standing absolutely still under a sensor could theoretically avoid triggering it, but in reality it is very difficult to stand that still for very long. Larger stationary groups will tend to have higher activation rates, since the inter-arrival time of gestures and adjustments goes down.

The workday on Friday, August 4th, 2006 is a good period to find examples of many kinds of gross behaviors. That is because SIGGRAPH was in Boston that week and many people came to MERL on Friday for a demo open house, so there are large groups of people walking around the space and loitering in lobbies and hallways. This is in addition to the more typical behavior we see in the space.

3.1. Moving Around the Lab

The first example in Figure 5 shows people moving around the space. A very common activity in this data set is a single person, or possibly a couple people walking side-by-side, walking down a hallway. Several examples are shown on the left of that figure. Dots represent motion, diagonal lines represent people moving through space, in this case along corridors.

On the right side of the same figure it is possible to see a small group moving from the lecture hall to the small conference room. The trace is similar to the single-parson case, but it appears wider because each sensor fires multiple times as the group moves underneath.

3.2. A Small Meeting

There is a meeting in the conference room on and off all day on the 4th. The meeting is apparent in Figure 6 as a dotted horizontal line near the top of the figure. Even when people are sitting down and are subjectively *still*, the sensor will still occasionally register motion due to communicative gestures, shifts in posture, or self-adjustments. The inter-activation times during this session of the meeting range from a couple seconds to a couple dozen seconds.

3.3. Large Groups

The final set of examples in Figure 7 show the signature of large groups, which are fairly rare at MERL. Near the top left of that figure is another session of the same small meeting in the conference room. Compare that trace to the the large group demonstrations in the lunch room and the main reception lobby (indicated on the right top quadrant of the figure).

This group arrives at the lobby from another area while traversing the sequence of sensors highlighted by horizontal stripes on the figure, and indicated on the lower right quadrant of the figure. This group was comprised of a dozen people. Notice that the movement trace is very wide, with each sensor firing a half dozen times during the passing of the group.

4. Models

The work of Conner *et al.* at Intel calls for sensor activations to be uses in isolation, for example to determine if a particular room is occupied [1]. Here we explore the more powerful space-time models: naive spatio-temporal models, building-structured models, and trip-structured models.

4.1. Naive Spatio-Temporal Models

The naive approach attempts to build statistical models of the spatio-temporal structure in the data without the benefit of any prior knowledge about the structure of buildings or the nature of human activity.

These techniques are particularly valuable for systems that need to be self-calibrating or adaptive without human intervention. For example, it is possible for a system to autonomously learn models of short-term pattern of behavior in a space and uses those models to control a robotic camera. We have shown that this method outperforms a human operator [9].

Spatio-temporal structure can be used to predict demand for resources and enable energy- or time-saving optimizations in a variety of application domains [12].

4.2. Building-Structured Models

By inserting a bit of knowledge into the system about how buildings are put together, and how those architectural idioms are used by people, it is possible to enable more powerful inferences. For example the system can infer the location of various resources, such as meeting rooms, by detecting certain basic activities and their relationship in space and time to semantic information outside the network, such as events on the computer network, or the access control network.

This system works even when individual activity detectors perform poorly. This is important because the simple sensors we can expect are often going to be noisy and ambiguous in ways that make high-performance activity detection impossible. Fortunately high-value activities tend to be frequently repeated: e.g. visiting a shared resource such as a kitchen. A repeated activity will contribute a little information each time it is repeated. The system can look for the activity indirectly, by accumulating statistical evidence of the repeated structure over time. This works reliably even when the probability of detecting individual instances of the activity is quite low.

This type of model is good at detecting building-centered resources, such as places where people perform specific activities, independent of the individuals involved. For example, a user who wanted to know where the closest printer is could cross reference activity data with electronic logs of printing jobs to find out, even if the sensors were not

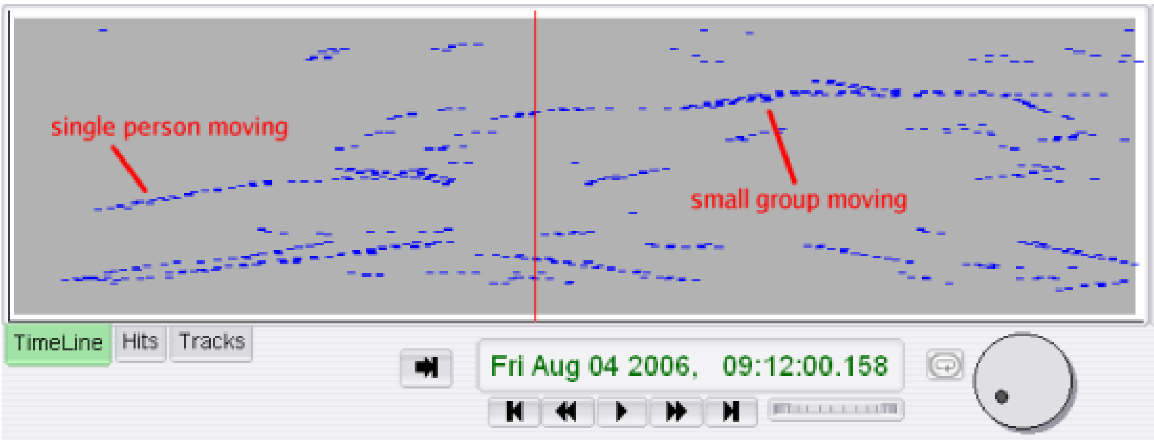

Figure 5. Dots are individual movements. Space on the vertical axis and time on the horizontal axis.

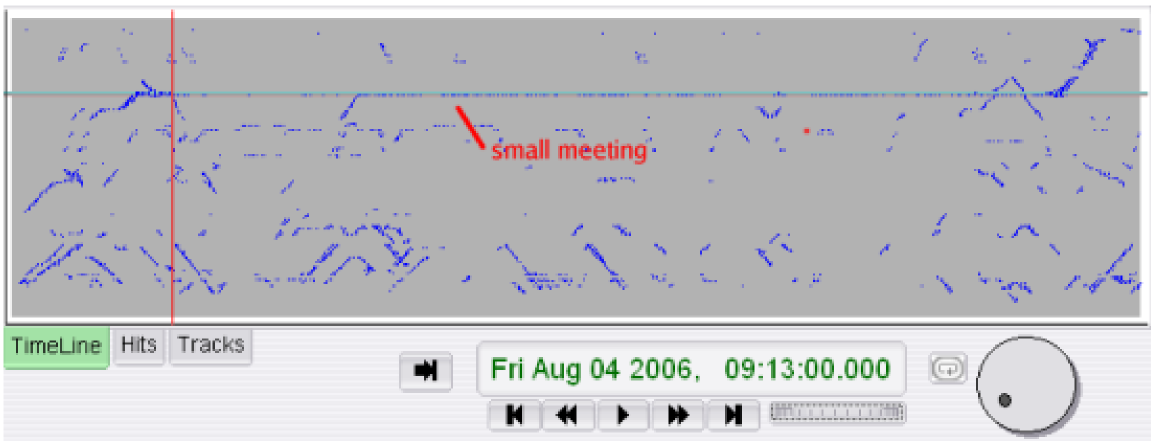

Figure 6. A horizontal line at one point in space, across time, is caused by a meeting in a conference room.

originally installed for that task and the system was never explicitly configured with that information [13].

4.3. Trip-Structured Models

For some applications it is advantageous to use models that focus on the people moving through the space, even though the sensor network may not be capable of unambiguously differentiating between individuals. The tracklet representation from Section 2.2 is an example of a data representation that supports trip-structured models. The basic unit of the representation is an unambiguous segment of a trip through the building. The tracklet graph is then a mechanism for combining the segments of an ambiguous situation into explanations such as "the person who left location A maybe have gone to location B at time T_B or location C and time T_C, but nowhere else".

These explanations are applicable to the security task since they allow pruning of the visual search that is required to find people who move between the gaps between camera views in a classical surveillance system. Knowing exactly where and when to look makes that search vastly more efficient [3].

The real power in this method, parallel to the case of the building-centered models, is when ambiguities can be statistically resolved by considering the accumulated evidence for structure in the data over many days, weeks, or even months. Sunshine-Hill, *et al.* show that it is possible to mine the tracklets graphs to synthesize plausible movements of individuals [6]. This could be used to gauge the typicality of observed movements, or to generate a set of most common paths in support of navigation applications.

Connolly, *et al.* showed that this representation is useful for extracting the social structure of a space by computing probabilities of trip beginnings and endings from the tracklet graph database and assuming that individuals have offices at particular fixed locations in the building [2]. It has also been shown that this technique can be used to detect changes in the social network structure even at the scale of a single day [10].

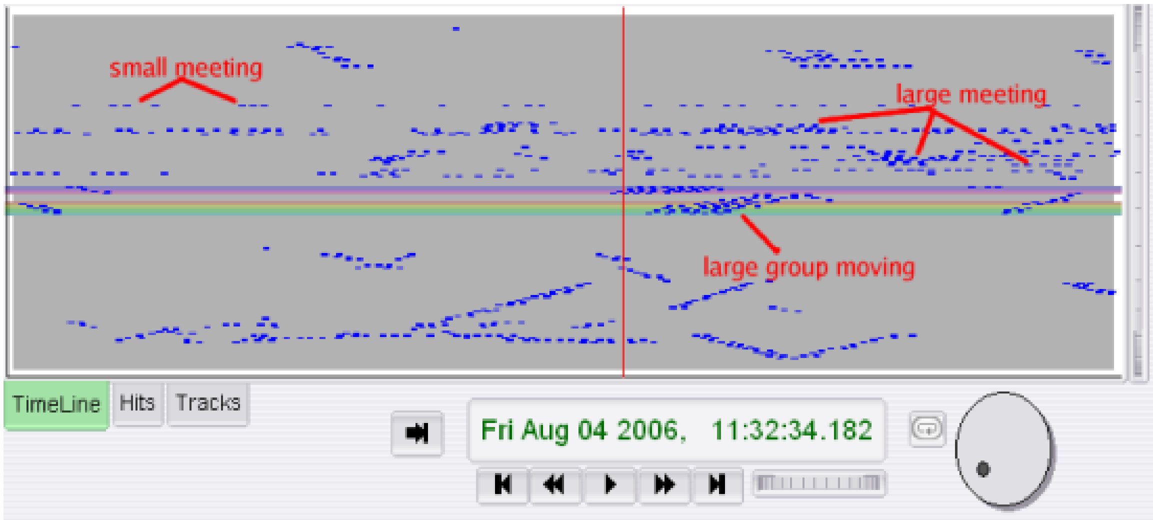

Figure 7. Large groups cause a distinctive pattern of activity.

5. Summary

People generate structures in space-time when they go about their activities in a building. These structures can be detected even by networks of the simplest sensors. With the appropriate perceptual tools these structures can be interpreted, revealing subtle and useful information about the activities in the building. It has been shown that it is possible to extract models of social interaction not only without tags, but also when it is not possible to track individuals unambiguously or even to differentiate one individual from another other than by their gross behavior. Developing these perceptual tools raises the probability that such networks will be deployed in buildings, seeding the world with powerful technology. These tools simultaneously enable high-value applications using only those initial, pragmatic sensor networks. Together these effects could help lower the cost of initiating the network effect for pervasive computing.

References

[1] W. S. Conner, L. Krishnamurthy, and R. Want. Making everyday life easier using dense sensor networks. In *Proceedings of the 3th international conference on Ubiquitous Computing*, London, UK, 2001. Springer-Verlag.

[2] C. I. Connolly, J. B. Burns, and H. H. Bui. Sampling stable properties of massive track datasets. In *ICMI Workshop on Massive Datasets*, Nagoya, Japan, November 2007. ACM.

[3] Y. A. Ivanov, C. R. Wren, A. Sorokin, and I. Kaur. Visualizing the history of living spaces. *Transactions on Visualization and Computer Graphics*, 13(6):1153–1160, 2007.

[4] A. Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):107–119, 2000.

[5] C. J. Reynolds and C. R. Wren. Worse is better for ambient sensing. In *Workshop on Privacy, Trust and Identity Issues for Ambient Intelligence*, Dublin, Ireland, May 2006. Pervasive.

[6] B. Sunshine-Hill, J. Allbeck, N. Pelechano, and N. Badler. Generating plausible individual agent movements from spatio-temporal occupancy data. In *ICMI Workshop on Massive Datasets*, Nagoya, Japan, November 2007. ACM.

[7] R. Want, T. Pering, G. Danneels, M. Kumar, M. Sundar, and J. Light. The personal server: Changing the way we think about ubiquitous computing. In *Proceedings of the 4th international conference on Ubiquitous Computing*, pages 194–209, London, UK, 2002. Springer-Verlag.

[8] M. Weiser. The computer for the 21st century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3(3):3–11, 1999.

[9] C. R. Wren, U. M. Erdem, and A. J. Azarbayejani. Functional calibration for pan-tilt-zoom cameras in hybrid sensor networks. *ACM Multimedia Systems Journal*, 12(3):255–268, December 2006.

[10] C. R. Wren, Y. A. Ivanov, I. Kaur, D. Leigh, and J. Westhues. Socialmotion: Measuring the hidden social life of a building. In *Third International Symposium on Location- and Context-Awareness*, Oberpfaffenhofen, Germany, September 2007. Springer.

[11] C. R. Wren, Y. A. Ivanov, D. Leigh, and J. Westhues. The merl motion detector dataset: 2007 workshop on massive datasets. Technical Report TR2007-069, Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, August 2007.

[12] C. R. Wren, D. Minnen, and S. G. Rao. Similarity-based analysis for large networks of ultra-low resolution sensors. *Pattern Recognition*, 39(10):1918–1931, October 2006. Special Issue on Similarity-Based Pattern Recognition.

[13] C. R. Wren and E. M. Tapia. Toward scalable activity recognition for sensor networks. In *Second International Workshop on Location- and Context-Awareness*, Dublin, Ireland, May 2006. Springer.