Large-Scale Human Activity Recognition Using Ultra-Dense Sensing

Matthai Philipose
Intel Research

Introduction

The ability to build computing systems that can observe, understand and act on day-to-day physical human activity has long been a goal of computing research. Such systems could have profound conceptual and practical implications. Since the ability to reason and act based on activity is one of the central aspects of human intelligence, from a conceptual viewpoint such a system could cast light on computational models of intelligence. More tangibly, perhaps, machines that reason about human activity could aid humans in aspects of their lives that are today considered outside the domain of machines.

Monitoring human activity is a basic aspect of reasoning about activity. In fact, such monitoring, whether of others or of ourselves, is something we all need to do: parents monitor children, adults monitor elderly parents, managers monitor teams, nurses monitor patients and trainers monitor trainees. Those following medication regimens, diets, recipes or task directions need to monitor themselves. Monitoring activity is not just ubiquitous, it also tedious and expensive. There is often no substitute for a dedicated, trained human monitor observing in detail those being monitored. Such extended observation of others results in fatigue in those observing, and resentment in those being monitored; these are classic symptoms, for instance, in caregiver-caretaker and manager-worker relationships. Of course, the necessary constant involvement of humans also makes monitoring expensive.

Tasks that are ubiquitous, tedious and expensive would usually be perfect candidates for automation. Machines don't mind doing tedious work, and expensive problems motivate the corporations that build the machines. In fact, given the aging demographics of our society, systems that notify family automatically when their elderly relatives trigger some simple alarms, such as falling, not turning off the stove, or not turning of their hot water have begun to be commercially available. However, compared to the abilities of a live-in family member, who can monitor the elders' competence in thousands of day-to-day activities, such systems only scratch the surface. In what follows, we describe in some detail a concrete application that could benefit from broad activity recognition capabilities, identify a crucial missing ingredient in existing activity recognizers, and describe how a new class of sensors, combined with emerging work in statistical reasoning, promises to significantly advance the state of the art by providing this ingredient.

An Application: The Caregiver's Assistant

Figure 1 An electronic Activities of Daily Living Form with checkmark added by the Caregiver's Assistant.

Caring for the elderly, whether in the capacity of a professional caregiver or as a family member is a common burden in most societies. Ensuring that the elder is able to perform crucial day-to-day tasks such as cooking, dressing, toileting and socializing is central to their wellbeing. Gerontologists have recognized this fact by developing a detailed list of activities, termed the Activities of Daily Living (ADL's), and metrics for scoring performances of these activities, such that an elder's score is accepted as an indicator of their cognitive health.

Professional caregivers in the US are often required to fill in ADL forms on each visit to a significant number of their patients. Unfortunately, although the data they collect is important enough that resourcing decisions such as Medicaid payments depend on it, the data collected is often inaccurate. Inaccuracies arise because much of the data is collected by interviewing elders who may have strong motives to misrepresent facts, and also because the window of data collection is narrow relative to the period being evaluated. Given the increasing pressure on caregivers' time, purely manual data collection seems unsustainable in the long run.

The Caregiver's Assistant system is intended to automatically fill out large parts of the ADL form automatically based on data collected from the elders' home on a 24/7 basis. Such an application would hopefully not just increase the quality of data collected, but (since it is a constant monitoring system) also allow for useful but currently infeasible techniques such as proactive intervention. The form on the left of Figure 1 is a prototype form. Actual forms include activities from 23 broad classes such as "housework" and "hygiene"; these classes instantiate to tens of thousands of activities such as "cleaning a bathtub" and "brushing teeth". The underlying activity recognition system thus needs to be able to track thousands of activities in non-laboratory conditions to take substantial burden off the human.

The professional caregiver could, at any given time, be provided a version of this form with potentially troublesome areas highlighted. Having this form before a visit can help the nurse prepare better for the visit. During the visit, the information can help direct interaction towards the more important issues. Formative work with roughly one hundred caregivers from around the country indicates that such a system could be quite useful, at least for caregivers.

Features for Discriminating Between Many Activities

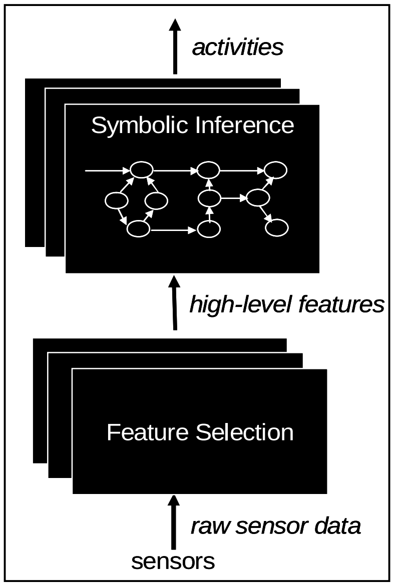

Figure 2 A typical activity recognition system.

The process of recognizing mundane physical activities can be understood as mapping from raw data from sensors monitoring the world to a label denoting an activity. Figure 2 shows how systems performing the mapping are traditionally structured. *Feature selection* modules typically work on high-dimensional, high-frequency data coming directly from sensors (such as cameras, microphones and accelerometers) to identify relatively small numbers of semantically higher-level features such as objects in images, phonemes in audio streams and motions in accelerometer data. *Symbolic inference* modules reason about the relationship between these features and activities in a variety of ways. The reasoning may include identifying ongoing activities, detecting anomalies in the execution of activities, and performing actions to help achieve the goal of the current activities.

Both feature selection and inference techniques have been investigated extensively. Depending on the feature they seek to use, researchers can draw from extensive bodies of work: objects, faces, automobiles, gestures, edges and motion flows (to take examples from the computer vision community alone) each have dedicated sub-communities of researchers. On the other hand, once features are selected, an activity recognition system could choose from a very large space of model representations and inference techniques. These techniques differ, for instance, in whether they support statistical, higher-order, or temporal reasoning, the degree to which they learn and the amount of human intervention they need in doing so, and the efficiency which they process various kinds of features, especially higher-dimensional ones. In Figure 2, we indicate the variety of selection and inference algorithms by stacks of boxes.

However, this profusion of options has not translated into an activity inferencing system capable of recognizing large numbers of day-to-day activities in natural environments. A key underlying problem is that no existing combination of sensors and feature selector has been shown to detect robustly the features necessary to distinguish between thousands of activities. For instance, objects used during activities have long been thought to be a crucial discriminator. However, existing object recognition and tracking systems [6] tend not to work very well when applied to a large variety of objects in unstructured environments. Activity recognition systems based on tracking objects therefore tend to be customized for particular environments and objects, and restricted in their utility as general purpose day-to-day activity recognizers. Given that producing each customized detector is itself a research task, the goal of general-purpose recognition has not surprisingly remained unattained.

A new class of small, wireless sensors to individual objects seems likely to provide a practical means of detecting objects used in many day-to-day activities [3, 4]. Given this stream of objects, recent work has shown that even simple symbolic inference techniques are sufficient for tracking the progress of these activities.

Detecting Object Use with RFID-Based Sensing

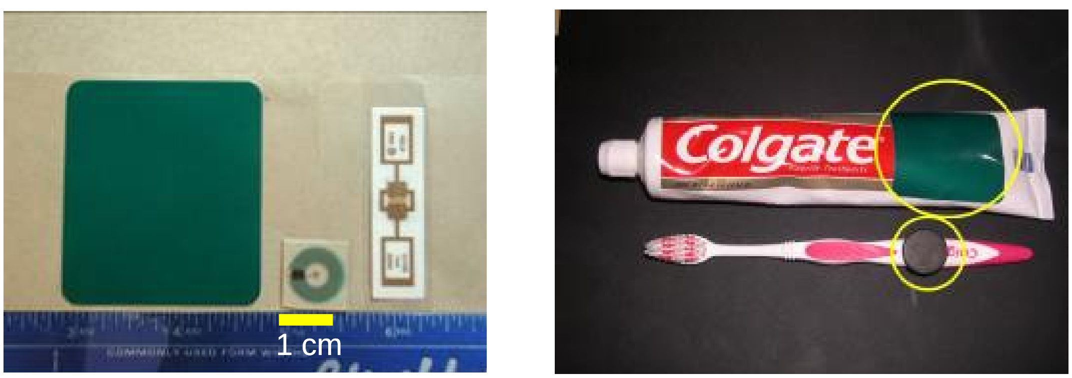

Figure 3 Radio Frequency Identification tags (left). A tagged toothbrush and toothpaste (right).

A passive Radio Frequency Identification (RFID [1]) tag is a postage-stamp-sized wireless, battery free transponder that, when interrogated (via radio) by an ambient reader, returns a unique identifier (see Figure 2). Each tag consists of an antenna, some protocol logic and optional non-volatile memory. Tags use the energy of the interrogating signal to return a 64 to 128-bit identifier unique to each tag, and when applicable, data stored in on-tag memory. Short-range tags, which are inductively coupled, have a 2 to 30 cm range, whereas long-range backscatter based tags have a 1 to 10 m range. Tags are available off the shelf for less than 50 cents a tag. Short range readers are priced in the low hundreds of dollars, whereas long-range ones are in the low thousands. Current trends point to a step drop-off in price of both tags and readers in the next few years.

If an RFID tag is attached to an object and the tag is detected in the vicinity of a reader, we can infer that the attached object is also present. Given their object-tracking abilities, RFID-based systems are currently under serious consideration for such commercial applications as supply chain management and asset tracking; existing uses include livestock tracking, theft protection in the retail sector and facilities management. The promise of RFID as viable system for tracking presence of large numbers of objects brings up the question of whether it can be used as the basis of a system for tracking the objects used by people whose activities we wish to track. Since a sensor can be attached to each object, we have, in principle an "ultra-dense" deployment of sensors that could allow each tagged object to report when it is in use.

Neither short-range nor long-range RFID systems, as conventionally designed are quite up to the task of detecting object use in a manner useful for activity tracking. Short-range RFID readers are typically bulky handheld units (similar to barcode readers) that are intentionally "swiped" on tags that are to be read. It is clearly impractical to expect those whose activities are to be tracked (whether they are elders or medical student) to carry a scanner and swipe tagged objects in the middle of their day-to-day tasks. Long-range tags, on the other hand, do not require any explicit cooperation from those being tagged: readers in the corner of a room can detect tags within it. Unfortunately, since a conventional RFID tag simply reports the *presence* of a tagged object is in the field of a reader, and not its *use*, these readers cannot tell us when objects are being used either. They simply report the list of all objects in the room they are monitoring. As we describe below, however, each of these modalities can be re-engineered to usefully and unobtrusively detect object use, in addition to object presence.

Figure 4 The iBracelet: Close-up (left) with quarter for comparison. In use (right).

Figure 4 shows how the short-range RFID reader can be adapted to become an unobtrusive object-use sensor. Essentially, the RFID reader, a radio with built-in processor, non-volatile memory and a power supply are integrated into a single bracelet called the iBracelet (on the left of the figure) [2]. The antenna of the RFID reader is built into the rim of the bracelet. When turned on, the bracelet scans for tags at 1 Hz at a range of 20-30 cm. Any object, such as the water jar on the right of the figure, that has a tag within a few 10-15 cm of its grasping surface, can therefore be identified as having been touched. The data can either be stored on board (for later offloading through a data port) or immediately radioed off board. The bracelet can currently read for 30 hours between charges when storing data locally, and for roughly 10 hours when transmitting the data. Careful placement of tags on objects can reduce false negative rates, i.e., tags being missed. However, given the range of the bracelet, “accidental” swipes of objects are unavoidable. The statistical framework that processes the data should therefore be able cope with these. Early studies indicate that the iBracelet with 40-cent inductively coupled tags are a quite practical means of detecting object touch, and therefore object use.

Some people or applications may deem wearing the bracelet an unacceptable requirement. WISP’s (Wireless Identification and Sensing Platforms [5]) may be a useful way of detecting object-use in these cases. WISP’s are essentially long-range RFID tags that have sensors integrated into them. These tags use incident energy from distant readers not only to return a unique identifier, but also to power their onboard sensor and communicate the current value of the sensor to the reader. For activity inferencing applications, so-called α-WISP’s (which are about the size of a large band aid; see Figure 5), which have integrated accelerometers are attached to the objects whose movements are to be tracked. When a tagged object is used, more often than not the accelerometer is triggered, and the ambient reader notified. A single room, which may contain hundreds of tagged objects (most of them inactive at any given time), may be monitored by a single RFID reader. A complication with WISP’s is that the explicit correspondence between the person using the object and the object being used is now lost, so that higher-level inference software may need to track this correspondence implicitly if necessary.

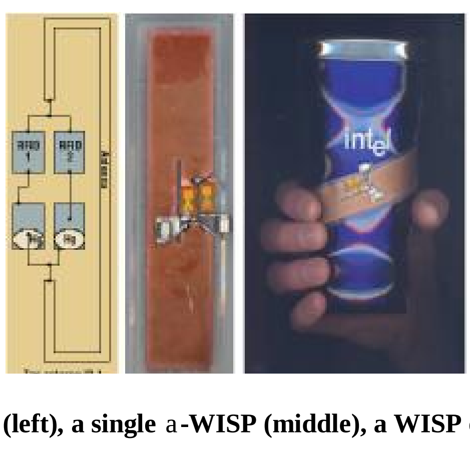

Figure 5 WISP's: Schematic (left), a single a-WISP (middle), a WISP on a coffee mug (right).

Inference on Object-Use Data

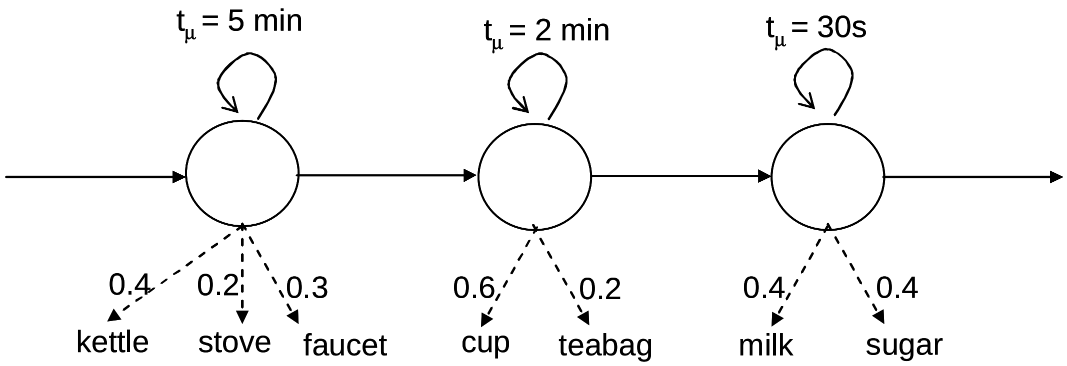

Figure 6 A simple probabilistic model for making tea.

Given the sequence of objects detected by RFID-based sensors, it is the job of the inference system to infer the activities happening. The inference system relies on a model translating from observations (in this case, the objects seen) to the activity label. Recent work [4] has shown that even very simple statistical models of activities are sufficient to distinguish between dozens of activities performed in a real home. Figure 6 shows the model for making tea as an example. Each activity is represented as a linear sequence of steps. Each step has a specified average duration, a set of objects likelihood to be seen in that step, and the probability that one of these objects will be seen in an observation window. In the figure, the first step (corresponding to boiling tea) takes five minutes on average; in each one-second window, there is a 40-, 20- and 30% chance respectively of a kettle, stove or faucet being used. Experiments in a real home with 14 subjects, each performing a randomly selected subset of 66 different activities selected from ADL forms, and using activity models constructed by hand to automatically classify the resulting data, have yielded above 70% (and often close to 90%) accuracy in activity detection.

Although the models are simple, it is impractical to model tens of thousands of activities by hand. The fact that the features to be recognized are English words representing objects, and that the label to be mapped to is an English phrase (such as “making tea”) lead to an interesting observation: the processing of building a model is essentially a process of translating probabilistically from English phrases to words. Recent work [7] has used this observation to completely automatically extract these translations using word co-occurrence statistics from the text corpora such as the web. Intuitively, if of 1 million web pages that mention “making tea”, 600,000 mention “faucet”, these systems accept 60% as the rough probability of using a faucet when making tea. These crude “commonsense” models can then be used as a basis for building customized models for each person by applying machine learning techniques to data generated by that person. Experiments on the above dataset have shown that these completely automatically learned models can recognize activities correctly roughly 50% of the time. Analysis of these corpus-based techniques has also provided indirect evidence that object-based models should be sufficient to discriminate between thousands of activities.

Conclusions

Monitoring day-to-day physical activity is a tedious and expensive task performed by most humans. Automating the monitoring therefore has the potential of tangibly easing the lives of many people. Traditional approaches to activity recognition have not been successful at monitoring large numbers of day-to-day activities in unstructured environments, partly because of their inability to identify sufficiently discriminative high-level features robustly. A new family of sensors, based on Radio Frequency Identification, is able to simply and accurately identify most of the objects used in activities. Given this rich stream of features, even simple statistical models can classify large numbers of activities with good accuracy. Further, these models are simple enough that they can extracted automatically from massive text corpora such as the web and customized on observed data to good effect.

Acknowledgments

This paper describes work done by the author jointly with the SHARP group at Intel Research Seattle and with researchers at the University of Washington. Specifically, the iBracelet was joint work with Adam Rea and Ken Fishkin. The work on WISP’s was done jointly with Joshua Smith and the WISP team. The work on mining models was done jointly with Mike Perkowitz, Danny Wyatt and Tanzeem Choudhury. Inference techniques used were developed jointly with Dieter Fox, Henry Kautz and Don Patterson.

References

1. K. Finkenzeller, *RFID Handbook*, 2nd ed., John Wiley & Sons, 2003.
2. K. Fishkin, M. Philipose, and A. Rea, “Hands-On RFID: Wireless Wearables for Detecting Use of Objects”, to appear, International Symposium on Wearable Computing, 2005.
3. E. Munguia Tapia, S. S. Intille, and K. Larson, "Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors," in Proceedings of PERVASIVE 2004, vol. LNCS 3001, pp. 158-175.
4. M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, H. Kautz and D. Hähnel . "Inferring Activities from Interactions with Objects". IEEE Pervasive Computing, October 2004, pp. 50-57.
5. M. Philipose, J. Smith, B. Jiang, A. Mamishev, S. Roy, and K. Sundara-Rajan, "Battery-Free Wireless Identification and Sensing," IEEE Pervasive Computing, vol. 4, no. 1, pp. 37-45, 2005.
6. B. C. S. Sanders, R. C. Nelson, and R Sukthankar, “The OD Theory of TOD: The use and limits of temporal information for Object Discovery,” In Proc. AAAI, 2002.
7. D. Wyatt, M. Philipose, T. Choudhury, “Unsupervised Activity Recognition Using Automatically Mined Common Sense”, In Proc. AAAI, 2005.