Proceedings of 2005 CACS Automatic Control Conference
Tainan, Taiwan, Nov 18-19, 2005

# Event-driven Activity Recognition from Heterogeneous Sensor Data

C. Y. Lin and Y. J. Hsu

Department of Computer Science and Information Engineering
National Taiwan University, Taipei, TAIWAN
{r93922129, yjhsu}@ntu.edu.tw

## Abstract

A smart environment should monitor and recognize the activities of people living within the space in order to provide timely support for safety, comfort, and convenience. This research proposes an approach to activity recognition based on data collected from sensors embedded in the environment. A variety of simple sensors are used to detect different aspects of the human behaviors. This paper presents *event-driven activity recognition* using an *event-action-activity model*. A multi-agent architecture is proposed to enable context-aware services based on user activities gleaned from aggregating and interpreting sensor data. We conducted experiments to evaluate the system's performance in recognizing a number of routine everyday activities. Three applications in personal assistance are described in this paper.

**Key Word: RFID, activity recognition, location tracking.**

## 1. Introduction

Activity recognition in real-time could also allow the development of just-in-time learning environments that educate and inform people by presenting information at the right time as they move through the environment. To create algorithms that detect activities, computational models that capture the structure of activities must be developed. The behavior of an individual can be characterized by the temporal distribution of his activities such as patterns in timing, duration, frequency, sequential order, and other factors such as location, cultural habits, and age. There are at least four ways for a computer to automatically acquire data about people's activities using sensor systems: (1) ask the individual, as in experience sampling, (2) remotely observe the scene using audio, visual, electromagnetic field, or other sensors and interpret the signal readings, (3) attach sensors to the body and interpret the signal readings, and (4) attach sensors to objects and devices in the environment and interpret the sensor readings.

In this work, the performance of activities recognition algorithms under the techniques: sensor data must be efficiently modeled, structured to allow for aggregation is assessed. The results for activity recognition are based on object tagged data collected from worn RFID reader, location data computed by wireless locating system, and environmental sounds aggregated from wearable motes. In this paper, we focus on accessing and manipulating an archive of sensed data to support applications in everyday computing, using concept hierarchies to relate sensed data to real-world definitions.

## 2. Hardware Devices

In the proposed system, we attach short-range RFID tags to everyday objects for identification. Meanwhile, RFID readers (shown in Figure 1) are embedded in wearable personal items, such as rings or wrist watches, to identify tagged objects that are being handled by the user. For location tracking, a user carries a PDA equipped with the Ekahau Positioning Engine 2.1 (EPE). It utilizes existing Wi-Fi network infrastructure to facilitate user mobility and asset visibility. Berkeley Motes (shown in Figure 2) will be used for speaking detection.

**Figure 1. Wearable RFID Reader and everyday object Tagged**

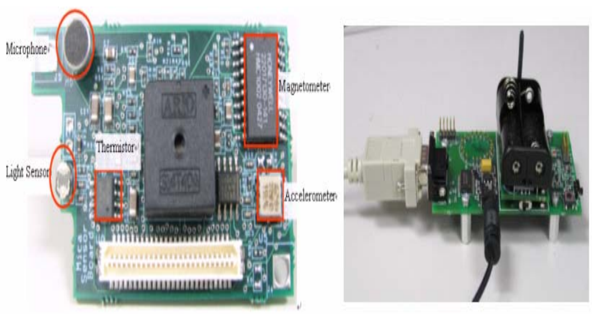


**Figure 2. A multi-sensor board MTS310 (left) and a**

**serial gateway MIB510**

## 3. Multiagent System Architecture

Based on the three sorts of sensors are put to use. We implemented the architecture based on multiagent system. The reason is though sensors aggregate lots of data to system, in fact, that data can't be used efficiently. We always want to deploy more and more sensors and have sensor-rich environment, but never use that data intelligently, we'll go to waste that data. If we let intelligent agents to build meaningful structure and common sense to that data, this helps a lot to infer and earn more information than that data just be used purely. In Figure 3, we set up six intelligent agents (Object Tracking Agent, Location Tracking Agent, Speaking Tracking Agent, Time Tracking System, Reasoning Agent, and Service Agent) to deal with sensor data, and give it more semantic meaning for reasoning.

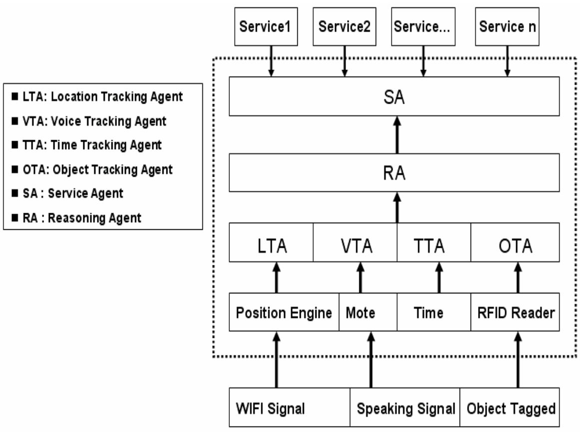


**Figure 3. System architecture based on multiagent system**

### 3.1 Object Tracking Agent

From our wearable RFID-Tag reader, Object Tracking Agent (OAT) can read the tag id embedded on the surface of objects. Furthermore, it can identify the object from its tag id, and other properties such as category, owner and RFID information that the object has from its object domain mapping. Details are described in Figure 4(a).

### 3.2 Location Tracking Agent

The Location Tracking Agent (LTA) runs on a notebook computer with the EPE 2.1 server installed, while the Ekahau Clients are installed in a number of mobile devices, such as a PDA. The EPE 2.1 server will monitor the locations of all Ekahau Clients. Whenever, the location of any client changes, the LTA captures the current time with the new location in its repository, and translates location information from the coordinates returned by EPE 2.1 into labels denoting physical areas, such as lab rooms, toilet, hallway, and so on. Figure 4(c) shows the technique for location data translation

### 3.3 Voice Tracking Agent

Motes data are received as Berkeley Motes data format through a gateway linked to PC, which is a combination of programming board and a communication board. Received data are collected by Voice Tracking Agent (VTA) in Figure 4(d). VTA calculates received Motes data to judge if the person is speaking. Besides, VTA uses two steps to detect speaking events. Berkeley Motes are programmed to send one data packet every 1.22 seconds. One packet contains 10 values of sensed sound volume. First, among 10 values in a packet, maximum and minimum are found out. If the difference value between the maximum and minimum is greater than 4, the corresponding packet duration is judged as speaking, otherwise, not speaking. Second, to filter noise single speaking packet is considered as a false positive because a human speaking usually goes on longer than 1.22 seconds. Hence, second level Speaking Detection clusters first level detected result by majority vote of 5 nearest neighborhood Speaking Buffer.

### 3.4 Time Tracking Agent

Time is also an important context. We sometime miss the cue of duration in activity recognition. Duration can display the degree and how much time spent in doing activity. Time Tracking Agent (TTA) in Figure 4(b) not only records the current time, but also analysis the time mapping to "Morning", "Noon", "Afternoon", "Evening" and other physical definitions. Besides, TTA supports past tracking, current prediction, and future scheduling. Time with a semantic description helps lots for reasoning.

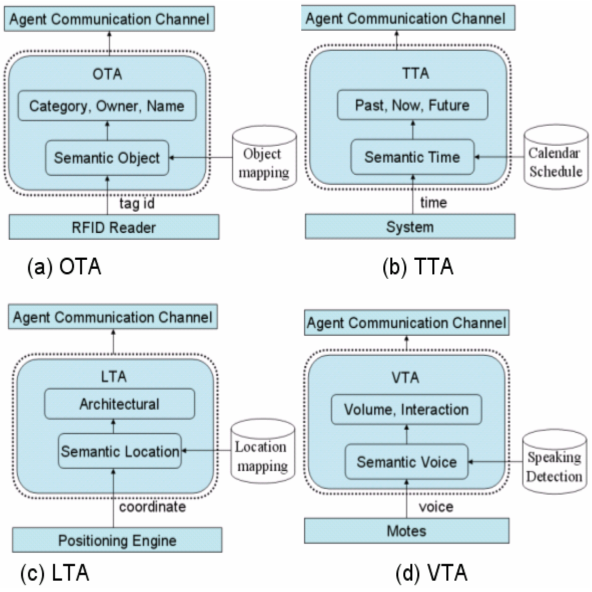


**Figure 4. Multi-agent working domain and interface.**

### 3.5 Reasoning Agent

Reasoning Agent (RA) will receive the sensor data from OTA, LTA, TTA, and VTA (shown in Figure 7(e)).The main work for RA is to infer the physical activities by aggregating the context data. RA uses Activity Model to infer the activities. Context data is modeled as a structure with common sense. Four context data in our Activity Model are object touched, location transmission, time slice, and speaking events.

The Bayesian network structure for activity recognition we built learns from the experience data of inhabitant's daily life. Each activity done by human can be traced from the object sequence, location, duration, sounds, and so on. The object sequence is the main feature for activity recognition, and location, duration, sounds minister to support the evidence.

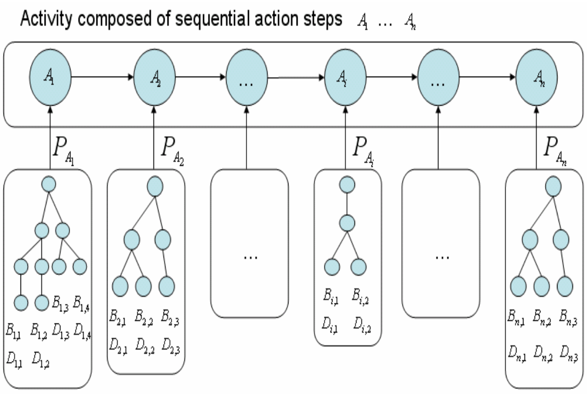


**Figure 5. Activity model for activity recognition**

In Figure 5, activity is divided into n sequential action steps $A_1..A_n$. Each action step is proved to be true while $P_{Action} \geq 0.5$. OTA aggregates object data and constructs that data in structured tree model. Each object three has its branch ($B_{ij}$) and depth ($D_{ij}$), $i$ means the number of action number, and $j$ means the order of branch. In this way, we can compute the action probability $P_{A_i} = \sum_{j=1}^{m} D_{Max} / D_{ij}$ with observed object data collected from OTA. $D_{Max}$ is the observed object sequence that matches the maximum depth in object tree model, and $m$ is the number of branch in object tree model. So the activity probability will be obtained by $P_A = \prod_{i=1}^{n} P_{A_i}$, and we set threshold $H = \prod_{i=1}^{n} 0.5 = (0.5)^n$. If $P_A > H$, then the activity A will be taken happened.

We take Make Tea as an example in Figure 6. Make Tea can be divided into Boil Tea, Steep Tea, and Flavor Tea three sequential action steps. Each action is decided by the corresponding probability $P_{Boil}$, $P_{Steep}$ and $P_{Flavor}$. If OTA got the object sequential data like {kettle, stove, cup, milk}, RA will compute the probability in following way.

$$P_{A_i} = \sum_{j=1}^{m} D_{Max} / D_{ij}$$

$$P_{Boil} = \sum_{j=1}^{2} D_{Max} / D_{1,j} = \frac{2}{3} + \frac{0}{3} = \frac{2}{3},$$

$$P_{Steep} = \sum_{j=1}^{1} D_{Max} / D_{2,j} = \frac{1}{2},$$

$$P_{Flavor} = \sum_{j=1}^{1} D_{Max} / D_{3,j} = \frac{1}{2},$$

$$P_{MakeTea} = P_{Boil} \times P_{Steep} \times P_{Flavor} = \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{6}$$

$$H_{MakeTea} = \prod_{i=1}^{3} 0.5 = (0.5)^3 = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8},$$

$$P_{MakeTea} > H_{MakeTea},$$

So the activity of Make Tea is recognized as really happened.

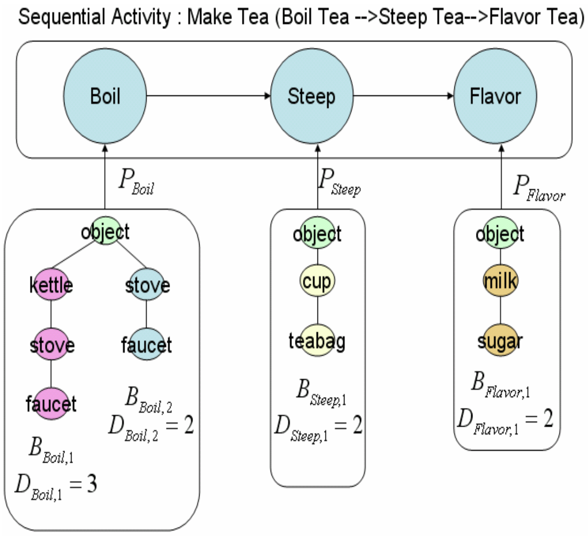


**Figure 6. Activity model for activity recognition**

In RA, sensor model and activity model are taken to infer human behavior.

In sensor model, each agent controls one sensor data input (e.g. LTA controls the location data, and OTA dominates the object data), and deals with that data automatically. RA collects that data translated by agents, and takes a look on the conditional probability table (CPT) values to determine which action is done by the human now.

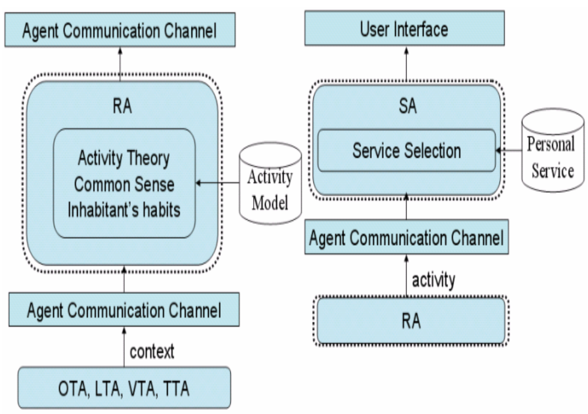


**Figure 7. Multi-agent working domain and interface.**

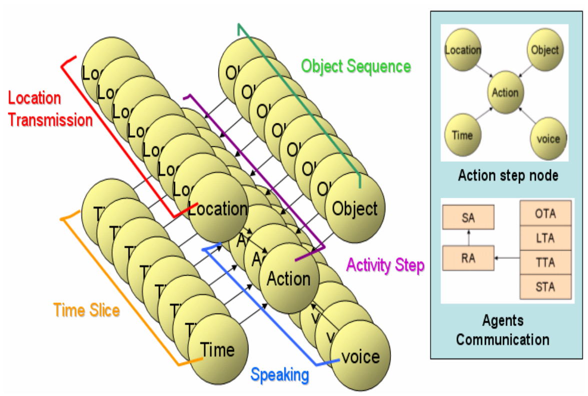


**Figure 8. Action step decided by context data in RA.**

### 3.6 Service Agent

Service Agent (SA) provides user service according to the activity inferred by RA. Personal Service Database (PSB) is built in SA. Our personal service focuses on helping human in every time. We setup three applications in SA to assist people in daily life. Figure 7(f) shows the working structure of SA.

First is Object Locate Service (OLS). Searching a certain thing is really tedious for people, especially when we need the item in emergency. So how to find the item as soon as possible is a really practical problem. We solved the problem with simple mapping past object to location information.

Second is Memory Assistant Service (MAS). With the advances in modern medicine and technology, human life expectancy has increased significantly during the last century. Issues associated with the aging population, e.g. high medical costs and lack of companionship, have become critical in nearly every country around the world. Exploring advances in information, sensor, and control technology to create solutions for "successful aging in place" has attracted much attention. Declines in memory capabilities of older adults lead to difficulties in remembering tasks (e.g. taking medication) and handling interruption. We are working on MAS to recognize forgotten or interrupted tasks and to assist the occupant in resuming these tasks. While SA gets the activity inferred from RA, SA will choose a service mapping to the recognized activity from the PSB.

Third is Future Remind Service (FRS). We have noticed that reminding is one of the most frequent issues users deal with in their daily life. Often I forget to take my house key or cellular phone when I leave the home in the morning. We set rules in our PSB, while we want to go out, SA will remind all objects we need on time. So if RA recognized that the activity was go to work or go out, it will check the PSB and do correspondent reminding.

Through the activity recognition technology, we can use it to create more and more useful application, and life will be much more easy-going.

### 3.7 Agent Communication

The communication of our agents is a FIPA like contract interaction protocol specification. Our communication message format contains the identification of which agent sends a request, and which action that the agent requests another agent to do, and the time when an agent makes a request, and other information needed to pass to another agent. When the agent received a request, it can understand the meaning of the request, and response to the request, therefore the agents can negotiate with each other. Table 1 is the message format that we defined for agents.

**Table 1. OSL Message format for communication.**

| Messages for Object Locating | | | | |
|---|---|---|---|---|
| *Sender* | *Receiver* | *Request* | *Time Stamp* | *Content* |
| SA | OTA | request object | system time | object name |
| OTA | SA | null | system time | time or location |
| SA | OTA | get location | system time | object name |
| LTA | SA | null | system time | object location |

We take OLS as an example. While a person makes a request for the location of certain object, SA will send a message called "request object" to OTA. OTA will receive the request, and know which action it needs to do. So it will check its object domain knowledge, to find if the object exists, if the object has location attribute, it means that the object always stays in a certain place (like refrigerator in kitchen, sofa in living room, and television in room or in living room), then it will directly return the location of the data, otherwise it will return the time when the object was touched. SA will get the response with the information of location or touch time, if the information is location, then it will response to the user, if the information is touch time, then SA will make a request called "get location" with the information of touch time to LTA. LTA will analyze the request, and find the location, where the people touch the object at that time. So LTA will send the location to SA, then SA will response to the user.

## 4. Experiments

We put everyday objects with RFID tag in our laboratory. Besides, we are equipped with wearable item with RFID reader, and use RS232 transmission line to connect our personal server.

In the experiment, we assumed that each activity is just a single action step. Besides, only object and location information are used to test and verify our model. Table 2 shows the activities defined in our experiment. We use our multi-agent system to infer the activities done by people. The result is show in Table 3. It's not so satisfied, but if we add the location information from LTA, we can easily distinguish No.1 and No.4 activities from their different working space compared with Table2.

**Table 2. Description of four defined objective activities**

| No | Activity | Related Objects | Related Location |
|---|---|---|---|
| 1 | Make Tea | Kettle, stove, faucet, teapot, teabag, milk, sugar, cup | Kitchen |
| 2 | Wash Dishes | Dish soap, sponge, rag, dishwasher | Kitchen |
| 3 | Brew Coffee | Coffee reservoir, tap, coffeemaker, coffee box, cup | Living Room |
| 4 | Make Tea | Thermos, teapot, teabag, milk, sugar, cup | Bed Room |

**Table 3. Experimental results for activity recognition without location information**

| No | True Positives | False Negatives | Recall (%) |
|---|---|---|---|
| 1 | 18 | 9 | 66 |
| 2 | 8 | 2 | 80 |
| 3 | 8 | 2 | 80 |
| 4 | 10 | 0 | 100 |

## 5. Related Work

More complex sensors such as cameras in computer vision have also been used for recognizing activities. Computer vision sensing for tracking [1, 2, 3] and action identification [4, 5, 6]) often works in the laboratory but fails in real home settings due to clutter, variable lighting, and highly varied activities that take place in natural environments. Little of this work has been extensively tested in the field due to the complexity of dealing with changes in the scene, such as lighting, multiple people, clutter, etc. Finally, because sensors such as microphones and cameras are so general and most commonly used as recording devices, they can also be perceived as invasive by some people.

There has been some previous research in algorithms for recognizing activities and patterns of activities from sensor data on data collected from living environments. Dynamic Bayesian networks (DBNs) [7] are a specific type of Bayesian network that graphically encode dependencies among sets of random variables which evolve in time. Hierarchical hidden semi-Markov models (HHSMMs), specific types of DBNs, have been used to track the daily activities of residents in an assisted living community [8]. Even though DBNs have proven to be one of the most powerful representations for temporal events and efficiently fusion information from multiple sensors [9], the complexity of the networks and learning algorithms make it difficult to apply them in problems involving hundreds of low-level sensors.

CybreMinder [10] is an early research of using context information for delivering messages via different ways in right situation. Messages can be voice message, e-mail or displaying on nearby displays. The user context here can be a person, location, time, activity, etc. In the similar way, Gate Reminder [11] also tries different reminding ways through different kind of interfaces deployed in a future smart home, as its name. This project sets up a home appliance located in the front door to achieve transparent interaction using RFID technology. The

RFID reader will real time capture people and objects identification by passive tags, match home member's item list with their predefined schedule, then provide the effective reminders according to the right contexts. In their work, a series of design principle and user studies were given.

Another kind of application aims to use reminder as the assisted cognition tool for caring the elders and people with memory impairment. Autominder [12] is a project to support activity reminder and help clients remain their daily living activities, such as eating, performing hygiene, taking medicine and toileting. It keeps the client's activity model for further planning and has scalability for addition or modification of an activity. This system can be further implemented on a nurse robot for home care.

## 6. Conclusion

The work described here is preliminary but demonstrate that ubiquitous, simple sensor devices can be used to recognized activities of daily living from real homes. Moreover, the proposed multiagent system presents an alternative meaning to sensor data. Therefore, a pre-trained model should be learned from the inhabitant's history in daily life. Our future work is focused on how to combine these raw data in reasonable way from common sense and define the context data in the view from physical world.

## 7. References

[1] S.S. Intille, J. Davis, and A. Bobick. Real-time closed-world tracking. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 697–703. IEEE Press, 1997.

[2] S. Stillman, R. Tanawongsuwan, and I. Essa. A system for tracking andrecognizing multiple people with multiple cameras. In Proceedings of the Second International Conference on Audio-Vision-based Person Authentication. 1999.

[3] I. Haritaoglu, D. Harwood, and L. Davis. W4: Who, When, Where, What: A real time system for detecting and tracking people. In Third International Conference on Automatic Face and Gesture. Nara, 1998.

[4] J.W. Davis and A.F. Bobick. The representation and recognition of action using temporal templates. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 928–934. IEEE Press, 1997.

[5] D.J. Moore, I.A. Essa, and M.H. Hayes III. Exploiting human actions and object context for recognition tasks. In Proceedings of the 7th IEEE International Conference on Computer Vision. IEEE Press, 1999.

[6] X. Sun and C.W. Chen. Probabilistic motion parameter models for human activity recognition. IEEE International Conference on Pattern Recognition (ICPR02), 1, 2002.

[7] K. P. Murphy. Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, University of California, Berkeley, 2002.

[8] H. Kautz, O. Etziono, D. Fox, and D. Weld. Foundations of assisted cognition systems. technical report cse-02-ac-01, University of Washington, Department of Computer Science and Engineering, 2003.

[9] A. Garg, V. Pavlovic, and J.M. Rehg. Audio-visual speaker detection using dynamic bayesian networks. IEEE International Conference in Automatic Face and Gesture Recognition, 2000.

[10] A.K. Dey and G.D. Abowd, "Cybreminder: A context-aware system for supporting reminders," Proc. of the 2nd international symposium on Handheld and Ubiquitous Computing (HUC '00), London, UK, Springer-Verlag , 2000, pp 172–186.

[11] S.W. Kim, M.C. Kim, S.H. Park, Y.K. Jin and W.S. Choi, "Gate reminder: A design case of a smart reminder," Proc. of the conference on Designing interactive systems (DIS '04), New York, NY, USA, ACM Press, 2004, pp 81–90.

[12] M.E. Pollack, L. Brown, D. Colbry, C.E. McCarthy, C. Orosz, B. Peintner, S.Ramakrishnan and I. Tsamardinos, "Autominder: An intelligent cognitive orthotic system for people with memory impairment," Robotics and Autonomous Systems, Vol. 44, 2003, pp 273–282.

## Acknowledgements

This research is partially supported by the National Science Council of Taiwan, R.O.C, under grant NSC-93-2218-E-002-148.