Human Intention Detection and Activity Support System
for Ubiquitous Sensor Room

Yasushi Nakauchi¹ Katsunori Noguchi² Pongsak Somwong²
Takashi Matsubara²

¹Inst. of Engineering Mechanics and Systems
University of Tsukuba
Tsukuba, Ibaraki 305-8573, Japan
nakauchi@kz.tsukuba.ac.jp

²Department of Computer Science
National Defense Academy
1-10-20 Hashirimizu, Yokosuka 239-8686, Japan
{g41039, s6904, matubara}@nda.ac.jp

Abstract

In this paper, we propose the human behavior detection and activity support environment Vivid Room. Human behaviors in Vivid Room are detected by many kinds of sensors embedded in the room (i.e. magnet sensors for doors/drawers, micro-switches for chairs, ID-tags for humans.) and that information is collected by sensor server via RF-tag system and LAN. In order to recognize meaningful human behaviors (i.e. studying, eating, resting, etc.), we’ve employed ID4 based learning system. Also we’ve developed human activity support system by using sound and voice by taking account of human behaviors in the room. The experimental results, which denote the accuracy of human behavior recognition and the quality of human support, are also shown.

1 Introduction

Recent development of Information Technology is making electric household appliances computerized and networked. If the environments surrounding us could recognize our activities indirectly by sensors, the novel services, which respect our activities, can be possible. This idea has initially proposed by Weiser as ubiquitous computing [16] and are emerged as Aware Home [3], Intelligent Space [5], Robotic Room I, II [12, 8], Easy Living [2, 4], Smart Rooms [9, 10], etc.

One of the most important factors for such systems is the recognition of human intentions ¹ by using ubiquitous sensors. Intelligent Space detects the position of human by using multiple cameras on the ceiling and makes a mobile robot to follow the human [5]. Easy Living also detects position of human and turns the light close to the human on [2, 4]. These systems are considered as providing services by taking account of human intention on where the human intends to move. On the other hand, one of the applications Robotic Room I provides, uses the human intention expressed more explicitly. As the finger pointing by a patient lying on a bed is recognized by vision, Robotic Room I makes the long arm robotic manipulator to hand the pointed object to the patient [12].

Though, the above mentioned applications are using human intention, in general, there are various kinds of human activities in a space. If the system could recognize the various kinds of human intention, it can provide more sophisticated services. Asaki et al. have proposed the human intention (i.e. changing clothes, preparing meals, etc.) recognition system by using state transition model [1]. But since the state transition model is pre-programmed one, it is difficult to extend the recognizer to adapt various kinds of sensor settings nor various kinds of human intention. Moore et al. have proposed Bayesian classification method, which enables to recognize the various kinds of human intention by using learning mechanism [6, 7]. But since their system uses only the knowledge on the temporal order of actions human caused, the kinds of intention that can recognize are rather limited.

For example, if 1) some books are stacked on the desk, 2) the door of bookshelf are opened and closed often, and 3) a human stands up from a chair and then opened the bookshelf, we can imagine that the person is arranging the books into the bookshelf. For this, we need to use various kinds of information observed, not

¹ The meanings of the human intention have wide spectrums. In this paper, we use the word intention as the behaviors, which are recognizable from the external observation.
only the current status of objects and the temporal order of events observed, but also the frequency on the usage of objects at the same time. Also, the system should be adjustable to the arbitrary configurations of room and kinds, numbers, and arrangements of sensors used.

In this paper, we propose the human intention detection and activity support environment Vivid Room. Human behaviors in Vivid Room are detected by many kinds of sensors embedded in the room (i.e. magnet sensors at doors/drawers, micro-switches at chairs, ID-tags at humans.) and that information is collected by sensor server via RF-tag system and LAN. In order to recognize human intentions (i.e. studying, eating, resting, etc.), we’ve employed ID4 based learning system. We make the learning system to use the information not only the current status of objects and the temporal order of events observed, but also the frequency of activities human does as well, to recognize the various kinds of human intention. Also, by equipping the learning functions, the system becomes adjustable to the variety of intentions, which need to be recognized for supporting human. The usage of RF-tag system with the learning mechanism, we believe the system becomes adjustable to the arbitrary configurations of room and kinds, numbers, and arrangements of sensors used.

We develop human activity support system in Vivid Room by using sound and voice by taking account of human intentions. We also conduct experiments to verify the accuracy of human intention recognition and the quality of human activity support.

2 System Design

2.1 Vivid Room

The outlook of Vivid Room is as shown in figure 1. All the room equipments (i.e. doors, drawers, bookshelves, chairs, workstations, etc.) are sensible as shown in table 1.

For example, the opening and closing of doors/drawers are sensed by magnet sensors attached on the them (see figure 2 and 3), whether the human is sitting on the chair or not is sensed by micro-switch attached on the seat (see figure 4). These information obtained by switches are transmitted via Spider III RFID (Radio Frequency IDentification) System developed by RF Code Inc. (see figure 5) [15]. With the usage of RF-tag system for the sensor networking, the system itself becomes adjustable for the arbitrary configuration of the room.

The existence of human is directory sensed by RF-tag possessed by the human. The login and logout information on the workstations are obtained by analyzing “utmp” login information on UNIX file system.

On the other hand, Vivid Room is also equipped with expressible information, which will be used for human activity support. So far, the functions such as the musical performance of arbitrary MIDI sounds, the speech by synthesized voice, the vibration of chair seats, and the light control of desk lamp are available (see table 2).

2.2 Human Intention Recognition System

Depending on what kind of services the system is willing to provide, the varieties of human intentions need to be recognized may vary. For example, when a human is sitting on a chair and if a robot (one of the system components) recognizes he/she is intending to study, it will be nice the robot to offer him/her a cup
of coffee by voice. On the other hand, if the robot recognizes he/she is intending to make a phone call, it will be nice the robot to hand him/her a memo pad silently.

To make the system adjustable to the variety of intentions, we've employed C5.0 learning system [11], which is based on ID4 learning algorithm [13]. A learning instance of C5.0 consists of attribute-value pairs and the class to be classified. It produces a decision tree (classification rules) by using information theoretic approach aimed at minimizing the expected number of tests to classify the objects.

The most direct and easiest manner is to consider the current status of sensors shown in table 1 as attributes. But these attributes are insufficient. For example, the frequent opening and closing of the bookshelf may denote that the person is arranging books. Also, if a person sitting on the chair has accessed to the refrigerator before he/she has seated, it may denote he/she is eating something.

Therefore, add to the current status, we’ve employed the frequency of sensor status changes and the previous event as the attributes of each sensor. The frequency is the discrete value such as frequent \((n \geq 10)\), often \((n \geq 5)\), some \((n \geq 1)\) and never \((n = 0)\) where \(n\) is number of status changes (i.e. opened and closed) of the sensor within the past 10 minutes. The value of previous event is consists of the combination of sensor name and its status (i.e. door1_closed, chair1_seated, etc.), which happened within the past 5 minutes\(^2\).

An example of leaning instances is as shown in table 3. The previous event possesses the only one previous event for each sensor. But as shown in the example, the causality information such as the refrigerator has opened, then closed, then the user has seated on the chair1 are expressed within a learning instance. The overall procedures the system performs are as shown in table 4.

---

**Table 1: Sensible information in Vivid Room.**

<table>
<thead>
<tr>
<th>Sensors</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entrance door (left/right)</td>
<td>closed, half-opened, full-opened</td>
</tr>
<tr>
<td>Chairs (two chairs)</td>
<td>seated, not-seated</td>
</tr>
<tr>
<td>Drawers of desk (7 drawers)</td>
<td>closed, opened</td>
</tr>
<tr>
<td>Shelf on desk (left/right)</td>
<td>closed, opened</td>
</tr>
<tr>
<td>Drawers on desk (5 drawers)</td>
<td>closed, opened</td>
</tr>
<tr>
<td>Refrigerator door</td>
<td>closed, opened</td>
</tr>
<tr>
<td>Bookshelf (left/right)</td>
<td>closed, half-opened, full-opened</td>
</tr>
<tr>
<td>Electric pot</td>
<td>placed, lifted</td>
</tr>
<tr>
<td>Person</td>
<td>existent, non-existent</td>
</tr>
<tr>
<td>Workstations (two WSs)</td>
<td>logout, login (without idle),</td>
</tr>
<tr>
<td></td>
<td>login (with idle)</td>
</tr>
</tbody>
</table>

**Table 2: Expressible Information in Vivid Room.**

<table>
<thead>
<tr>
<th>Peripherals</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speakers</td>
<td>MIDI sound, synthesized voice</td>
</tr>
<tr>
<td>Vibrators</td>
<td>on/off</td>
</tr>
<tr>
<td>Desk Lamps</td>
<td>on/off</td>
</tr>
</tbody>
</table>

---

\(^2\) When no events were observed within the time limits, the values of the frequency and the previous event are **NULL**.
3 Implementation

3.1 System Configuration

The system configuration of sensor server is as shown in figure 6. The sensor information transmitted by RF-tags is received by Spider III Reader and is collected via RS-232c. Other information such as the login status of workstations is collected via LAN.

The server also controls the on and off of the desk lamps and the vibrators via relay circuits. The MIDI sounds and the synthesized voice are also played via speakers. All these server software were developed by Java.

To confirm the current sensor status, we have developed the graphic interface as shown in figure 7. In the interface, the sensor statuses with the room layouts of Vivid Room are depicted as Java animations.

3.2 Human Intention Recognition System

In this research, we assumed Vivid Room as an office. We also have assumed the intentions occur in the office were studying, arranging, eating, and resting.

In order to recognize these human intentions from the sensed data, we apply ID4 based learning system. To obtain the rules (or decision tree) for classifying these four statuses, we need to have plenty of learning instances. So we have asked 10 subjects to perform as if he/she is doing one of the four intentions. The examples of instructions for the subjects were as follows.

**Studying** “Translate the Japanese sentences to English.” (The subjects were given Japanese one page documents. They also indicated that the dictionaries are in the bookshelf on the desk and the stationeries are in the right up most drawer of the desk.)

**Arranging** “Sort and store the journals in the bookshelf.” (100 journals of the three societies were...
stacked in random order on the desk. On the bookshelf, the kinds of journals to be stored were indicated.)

**Eating** “Make a cup of coffee and drink.” (The hot water is in the pot. The cup and the cans of instant coffee, sugar, and creams were placed on the refrigerator.)

**Resting** “Relax and read the magazines you like.” (Several kinds of magazines were stacked on the desk.)

We also have developed the software as it automatically generates the learning instances as shown in table 3 from the periodically updated sensor data. We have collected about 400 learning instances while the subject were performing the indicated tasks. Then, we made C5.0 to produce the classification rules from the learning instances. The typical classification rules for each intention were as follows.

**Studying** If the opening and closing of the bookshelves or the refrigerator have not observed within the past 5 minutes and the user is logging in the workstation without idling, then the user is studying.

**Arranging** If the opening and closing of bookshelves have observed frequently and the user is not sitting on a chair, then the user is arranging.

**Eating** If the refrigerator has opened and closed, or the pot has lifted before the user sits on a chair, then the user is eating.

**Resting** If the opening and closing of the bookshelves have not observed within the past 5 minutes and the user has logged out from the workstation, then the user is resting.

We also have developed the system to recognize the human intentions from the current sensor status by using the above mentioned classification rules. An example of recognition results were as shown in figure 8.

The horizontal axis denotes the time for 60 minutes. As shown in the figure, the human intentions that vary minutely have observed.

3.3 Human Action Support System

In order to demonstrate the possibilities of human action supports that take account of human intentions, we have developed two kinds of applications.

3.3.1 Ambient Sounds

One of the characteristics of ubiquitous systems is the less involvement between human and the system. The system monitors the human activities without disturbing human’s free activities. Therefore, we believe one of the promising applications is to provide the atmospheric services that change humans’ mood.

So we have developed the application, which plays ambient sounds by taking account of the human intentions. The varieties of the sounds played according to each intention are as follows.

**Studying** Playing classic music for not disturbing and for easing the human who is studying.

**Arranging** Playing march music for encouraging the human who is arranging objects.

**Eating** Playing popular music for refreshing the mood.

**Resting** Playing twitters of birds and sounds of stream for relaxation.

---

3 The classification rules were generated by off-line. The fixed rule sets were used for the on-line recognition.
3.3.2 Situated Voice Assistants

The other application is more task oriented one. Since the system could know the intentions of human, we have developed the voice assistant application by taking account of what the human is intended to do.

We’ve assumed that each drawers of desk and each shelf of bookshelves in Vivid Room have specific purposes (i.e. storing stationeries, cutleries, snacks, journal papers, magazines, etc.). The examples of reactions the system will take when it recognized each intention are as follows.

Studying When the human opened the drawers that are not for stationeries, the system speaks as “the stationery is in the right up most drawer of the desk.” When the human opened the doors of bookshelves that are not for dictionaries, the system speaks as “the dictionaries are in the bookshelf to your back.”

Arranging When the human opened arbitrary drawers of the desk, the system speaks as “the stationeries are in the right up most, the cutleries are in the right middle, and the snacks are in the right down most drawers.” When the human opened any doors of bookshelf, the system speaks as “the magazines are in the upper and the journals are in the lower shelves.”

Eating When the human opened the drawers that are not for cutleries, the system speaks as “the cutleries are in the right middle drawer of the desk, and the cold drinks are in the refrigerator.”

Resting When the human opened the drawer that is not for snacks, the system speaks as “the snack is in the right down most drawer of the desk.”

4 Experiments and Discussions

4.1 Human Intention Recognition

In order to measure the human intentions recognition qualities, we have conducted the experiments with 10 subjects who are different from the ones used for generating the classification rules. We also have instructed the new subjects to do the tasks as shown in section 3.2.

We’ve obtained about 300 sensed instances with their intentions. Then, we made the system to classify the instances into the four kinds of classes (intentions). As the results, we’ve confirmed the system could recognize the new subjects’ intentions at the accuracy of 93.7%.

As mentioned in section 3.2, the learning instances for producing the decision tree were automatically generated from the sensor data. Though the information on the configuration of the room and the kinds, the numbers, and the arrangements of sensors were not explicitly input to the system, we have confirmed that the system could produce the classification rules, which classify the human intentions at the very high accuracy. Also, from the typical classification rules shown in section 3.2, we can see that some of the rules are using the frequency of the sensor activities. This fact denotes that the employment of the frequency, which was not took into the consideration in the previous researches, contributes to the classification of human intentions at high accuracy.

4.2 Support of Human Action

To evaluate the applications described in section 3.3, we have conducted the questionnaire to the subjects. In the questionnaire form, we have asked 10 subjects to score (from −2 to 2) how much the services improved their task (work efficiency) and how comfortable with the services (comfortableness) with comments for each of the applications. The evaluation results were as shown in figure 9 and 10.

As for Ambient Sounds, the subjects felt the improvement of the work efficiencies except for the studying and also felt the comfortableness at each of categories (see figure 9). Some of the subjects commented that they felt even the classic music are noisy when they are concentrating on studying, though most of them felt comfortable. So we are planning to extend the system so that it recognizes human intentions more in detail.

As for Situated Voice Assistants, the subjects felt the improvement of the work efficiencies at each of categories (see figure 10). Especially for the subjects who were arranging objects, the voice assistance were scored high. This phenomena is explained by the one of the comments from the subjects: “the consist and timely explanations indicating where should I store the objects were very useful.”

On the other hand, as for the comfortableness, the averaged scores of studying, eating and resting were about 0. Only when the subjects were arranging ob-

4 For example, opening/closing of the bookshelves have observed frequently in the case of arranging, opening/closing of the refrigerator door has observed in the case of eating, and etc.
objects, they felt comfortable. Since this application is aimed for improving the work efficiencies, from the results that the subjects were not feeling uncomfortableness (except for the small uncomfortableness in eating), we could confirm the system has not disturbed the subjects.

5 Conclusion

In this paper, we proposed the human intention detection and activity support environment Vivid Room. For the human intention detection, our system employed the information, not only the current status of objects and the temporal order of events observed, but also the frequency on the usage of object at the same time. Also, our system is adjustable to the arbitrary configuration of room and kinds, numbers, and arrangement of sensors by employing ID4 based learning algorithm and ID-tag based sensor network. We have developed human activity support applications in Vivid Room by taking account of human intentions. From the experimental results with subjects, we’ve confirmed the accuracy of the human intention recognition and the efficiency of human activity support applications.

As the future works, we are planning to extend the sensibility abilities to obtain human’s standing positions in Vivid Room. Also we are planning to provide the physical assistance by using autonomous mobile robots.

References


