Proceedings of the 2004 IEEE International Workshop on
Robot and Human Interactive Communication
Kurashiki, Okayama Japan September 20-22, 2004

Human Behavior Recognition for Cooking Support Robot

T. Fukuda[1] Y. Nakauchi[2] K. Noguchi[3] T. Matsubara[3]
[1]College of Eng. Syst. [2]Grad. School of Syst. and Info. Eng. [3]Dept. of Computer Science
Univ. of Tsukuba Univ. of Tsukuba [3]National Defense Academy
Tsukuba, Ibaraki 305-8573, Japan Yokosuka 239-8686 Japan
E-mail: fukuda@hri.iit.tsukuba.ac.jp, nakauchi@iit.tsukuba.ac.jp
noguopanda@yahoo.co.jp, matubara@nda.ac.jp

Abstract

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. In this paper, we propose human activity recognition system, which infers the next human action by taking account of the past human behaviors observed so far. We also developed cooking support robot, which suggests what the human should do next by voice and gesture.

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** [17] and been emerged as Aware Home [5], Intelligent Space [8], Robotic Room I, II [16, 11], Easy Living [3, 7], Smart Rooms [14, 15], etc.

One of the most important factors for such systems is the recognition of human behavior 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 [8]. Easy Living also detects position of human and turns the light close to the human on [3, 7]. These systems are considered as providing services by taking account human intentions on where to move. On the other hand, one of the applications Robotic Room I provides, uses the human intention expressed more explicitly. When 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 [16].

In order to recognize implicit human intentions, Asaki et al. have proposed the human behavior (i.e. changing clothes, preparing meals, etc.) recognition system by using state transition model [1]. Moore et al. have proposed Bayesian classification method, which enables to recognize the various kinds of human behavior by using learning mechanism [9, 10]. We also have proposed human intention recognition method by using ID4 based learning algorithm and succeeded to recognize what a human intend to do such as study, eat, rest, etc [12].

With these researches, the certain kinds of human statuses are recognized. But even if we know that a human is cooking something, the varieties of human supports are rather limited. Suppose that a human is making a cup of coffee, it will be nice to suggest where the cream is, when a human took a cup and an instant coffee. In order to realize such suggestions, the system should know the time series of procedures and infer the next action human should execute by taking account of observed human actions obtained so far.

In this paper, we propose human activity recognition system, which infers the next human action by taking account of the past human behaviors observed so far. All the merchandise in supermarkets has one dimensional bar code. But in near future, they will be replaced to IC tags, which maintain maker name, kinds of merchandise, place of production, expiration date, etc. This means that all the properties in home will be labeled by IC tags. So in this paper, we presume foods, cooking tools, tableware, and cutlery in kitchen are labeled by IC tags and the movements of these items can be observed by antenna placed on shelves and kitchen counters. We also develop human activity support system in the kitchen by using mobile robot.

0-7803-8570-5/04/$20.00 ©2004 IEEE

2 System Design

2.1 Inference from Series of Human Actions

We define an observed action by sensors as **action** a_i and a set of actions as $A = \{a_1, a_2, ..., a_n\}$. We also define a set of time series actions in arbitrary length as **action pattern** p_i and a set of action patterns as $P = \{p_1, p_2, ..., p_m\}$. Suppose that we have observed the action pattern $p_o = \{\mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}\}$ by watching a human and there exists an action pattern $p_i = \{..., \mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}, a_6, ...\}$ in database P, which is the collection of action patterns observed so far. We could find the same time series action pattern in p_i and can infer that the next action the human should execute is action a_6.

Human sometimes behaves redundantly or concurrently. For example, p_o may contains a_n as $\{\mathbf{a_3}, \mathbf{a_2}, a_n, \mathbf{a_4}\}$ or p_i may contains a_n as $\{..., \mathbf{a_3}, a_n, \mathbf{a_2}, \mathbf{a_4}, a_6, ...\}$. These actions are considered as noise when the original time series actions have meanings such as the procedures for making coffee, cooking hamburgers, etc. So we must develop the inference system, which could infer the next human action even if such noises are contained.

Add to that, human procedures (time series actions) may have branches. For example, one may add sugar and the other may add cream after he/she made black coffee. This phenomena mean that there may exist $p_j = \{..., \mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}, a_7, ...\}$ add to p_i mentioned above. If the inference system uses not only time serial information but also frequency of action patterns observed so far, it can predict more preferable next action human should perform. For example, you can easily find that something strange, when you see your familiar person who always drink black coffee added cream in it.

2.2 Time Sequence Data Mining

In this paper, we propose the inference system, which takes account of frequencies in the action patterns observed so far. At first, we will explain briefly the typical algorithms that extract temporal orders in time series patterns.

Apriori algorithm proposed by Agrawal is one of the famous data mining method for temporal sequential data [2]. We will explain Apriori algorithm by using examples. Supposed that there are four time series data sets $p_1 = \{\mathbf{a_3}, a_2, \mathbf{a_4}, a_6\}, p_2 = \{\mathbf{a_3}, a_2, \mathbf{a_4}, a_6\}, p_3 = \{\mathbf{a_3}, a_2, a_5, a_6\}, p_4 = \{\mathbf{a_3}, a_1, \mathbf{a_4}, a_6\}$ in database. Apriori algorithm extracts the partial sequences from the data sets by taking account the number of occurence and certainty given by a user. For example, it finds the partial sequences $\{\mathbf{a_3}, \mathbf{a_4}\}$, which means "$\mathbf{a_4}$ occurs after $\mathbf{a_3}$". The certainty is the occurence ratio (i.e. $\mathbf{a_4}$ happens after $\mathbf{a_3}$ at the ratio of 75%). It is known experimentally that the calculation costs increase exponentially as the number of data sets increases with the Apriori algorithm.

On the other hand, Pei proposed PrefixSpan algorithm, which extracts a multiple frequency patterns efficiently in terms of computational costs [13]. Supposed that there are four time series data sets $p_1 = \{\mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}, \mathbf{a_6}\}, p_2 = \{\mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}, \mathbf{a_6}\}, p_3 = \{\mathbf{a_3}, \mathbf{a_2}, a_5, \mathbf{a_6}\}, p_4 = \{\mathbf{a_3}, a_1, \mathbf{a_4}, \mathbf{a_6}\}$ as same as the above example. PrefixSpan extracts the partial sequences with the number of occurence as shown in figure 1. $\{a_3/4, a_2/3, a_4/2, a_6/2\}$ in the figure denotes that there is a time series data $\{\mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}, \mathbf{a_6}\}$ with the frequency shown as suffix (i.e. the occurrence of $\mathbf{a_3}$ alone is 4 and the occurrence of $\{\mathbf{a_3}, \mathbf{a_2}, \mathbf{a_4}, \mathbf{a_6}\}$ as the time series data is 2).

$a_1/1$	$a_4/1$	$a_6/1$	
$a_1/1$	$a_6/1$		
$a_2/3$	$a_4/2$	$a_6/2$	
$a_2/3$	$a_5/1$	$a_6/1$	
$a_2/3$	$a_6/3$		
$a_3/4$	$a_1/1$	$a_4/1$	$a_6/1$
$a_3/4$	$a_1/1$	$a_6/1$	
$a_3/4$	$a_2/3$	$a_4/2$	$a_6/2$
$a_3/4$	$a_2/3$	$a_5/1$	$a_6/1$
$a_3/4$	$a_2/3$	$a_6/3$	
$a_3/4$	$a_4/3$	$a_6/3$	
$a_3/4$	$a_5/1$	$a_6/1$	
$a_3/4$	$a_6/4$		
$a_4/3$	$a_6/3$		
$a_5/1$	$a_6/1$		
$a_6/4$			

Figure 1: Time series data generated by PrefixSpan.

Since PrefixSpan is better in computational costs and easy to utilize, we employ PrefixSpan as the basic engine for extracting the time sequence data from the observation of human behaviors.

2.3 Human Behavior Inference Algorithm

In this section, we'll explain the human behavior inference algorithm that takes noises both in the time sequence data within the database and in the human observation data into consideration.

At first, we'll define the terms used in the algorithm. Every behavior observed by sensors is defined as **input data** w_i. Suppose that the latest input data is w_i and the number of W input data as $\{w_{i-W+1}, \cdots, w_i\}$ observed recently, these W time series data is defined as **window data** of width W.

The inference to predict human's next behavior is to find the same time sequence data as the window data from the behavior database, which consists of enormous amount of time sequence data observed from the human behaviors in the past. We employ PrefixSpan for generating partial time sequence data from the behavior database.

The matching between the input data and the behavior database is done by window size. For example, if we start to find the time series input data of window size 5 and could not find the exactly same sequence, we'll reduce the window width to $4, 3, 2$. In this way, even if the input data contains some noises, we will be able to find the exactly same data form the database. The maximum window size used at the beginning of search is defined as W_{max} and the minimum one is as W_{min}. In order to infer the human's next action, we need to know several time sequence data as a hint, so W_{min} is used for terminating the search.

The **inferred event** is the action that occurs succeeding to the matched time sequence with the window size W at the highest certainty. The **certainty** is calculated as follows.

$$certainty = \frac{O_{ia}}{O_{pa}}, \quad (0 \leq certainty \leq 1) \qquad (1)$$

where O_{ia} is the Occurence of Inferred Action and O_{pa} is the Occurence of Previous Action to the Inferred Action calculated by PrefixSpan (see figure 2).

The outputs of the inference engine are one of the following three kinds; EOS (End Of Sequence), inferred event with certainty, or NULL. EOS is the output when the observed time series data matched in the database but the observed most recent event is the end of sequence in the matched database. So the inference engine could not infer the next event in this case. When the inference engine found the matched time series data and there was a succeeding event in the database, it outputs the succeeding event as the inferred event with the certainty calculated by formula 1. NULL is the output when the inference engine could not find the matched data in the database even though it reduced the time series actions till window width becomes W_{min}.

The overall algorithm of our proposed inference engine is as follows. 1) At first, it creates the window with the size $W = 5$ and initializes all contents as NULL. 2) Then, it inputs the observed events at most $W = 5$ to the window so that the most recent event becomes w_W. 3) It finds the exactly same time series data as the window in the database.

4) If the number of matched data was singular, it outputs the inferred event ($(W + 1)th$ event in the database) with its certainty calculated by formula 1. If $(W + 1)th$ event does not exist, it outputs EOS.

5) If the number of matched data was plural, it selects the most high certainty ones, then selects the most long sequenced ones [1]. If the multiple candidates still remain even if it applied the choosing strategy, it selects the arbitrary candidates as the matched data. Then, it outputs the inferred event or EOS as same as the procedure 4.

6) If there were no matched data from the procedure 3, it reduces the window size to $W = W - 1$ as shown in figure 3. Then, it finds the matched data with these multiple windows as same as the procedure 3. When the window size becomes $W < W_{min}$, it outputs NULL since it could not find any matched data in the database. The above mentioned procedures are summarized in table 1.

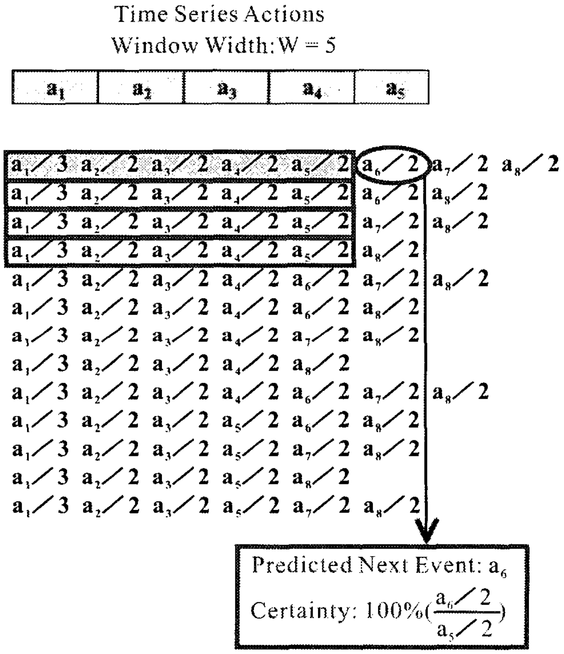

Time Series Actions
Window Width: W = 5
a_1 a_2 a_3 a_4 a_5
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_5/2$ $a_6/2$ $a_7/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_5/2$ $a_6/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_5/2$ $a_7/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_5/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_6/2$ $a_7/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_6/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_7/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_4/2$ $a_6/2$ $a_7/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_5/2$ $a_6/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_5/2$ $a_7/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_5/2$ $a_8/2$
$a_1/3$ $a_2/2$ $a_3/2$ $a_5/2$ $a_7/2$ $a_8/2$
Predicted Next Event: a_6
Certainty: $100\%(\frac{a_6/2}{a_5/2})$

Figure 2: Matching algorithm.

2.4 Activity Support of Human

If the system suggests a human for the next action he/she should do, it will be very helpful. At that time, it is important not to disturb human's free activities

[1] This is because that the longer the time series data in the database becomes, the more it describes the detailed procedures.

Table 1: The procedure of matching algorithm.

No.	Procedures
1	Set $W = 5$ and creates the window $\{w_1, w_2, w_3, w_4, w_5\}$ with all values as NULL.
2	Inputs the observed at most $W = 5$ events to the window so that the most recent event becomes w_W.
3	Finds the exactly same time series events as of the window in the database. No. of matched data is singular. $\Rightarrow$ go to 4. No. of matched data is plural. $\Rightarrow$ go to 5. No. of matched data is none. $\Rightarrow$ go to 6.
4	Outputs the inferred event ($(W + 1)th$ event in the database) with its certainty calculated by formula 1. Outputs EOS if $(W + 1)th$ event does not exist.
5	Selects the most high certainty ones, then selects the most long sequence ones.
6	Reduces the window size to $W = W - 1$. $\Rightarrow$ go to 3. If the window size becomes $W < W_{min}$, it outputs NULL.

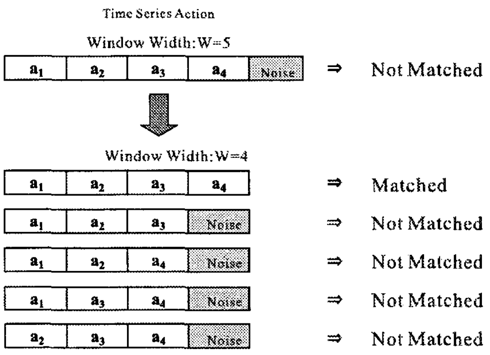

Time Series Action
Window Width:W=5
a1 a2 a3 a4 Noise ⇒ Not Matched
Window Width:W=4
a1 a2 a3 a4 ⇒ Matched
a1 a2 a3 Noise ⇒ Not Matched
a1 a2 a4 Noise ⇒ Not Matched
a1 a3 a4 Noise ⇒ Not Matched
a2 a3 a4 Noise ⇒ Not Matched

Figure 3: Reduction of window size.

derived from his/her own preferences. So in this research, we develop the mobile robot that recommend the human's next action by voice and gesture.

In order to reduce the uncomfortable or unsuitable recommends, we employed the threshold to the certainty (see formula 1). We made the robot to suggest the human's next action only when the certainty of inferred event is above the threshold.

We conducted the experiments with ten subjects and collected their comments if the recommendations (inferred events) are suitable or not. As the results, we confirmed that the most of certainties, which the subjects felt unsuitable were below 0.55. So we've set the threshold to 0.55 and made the robot not to issue the recommendation whose certainty is below it.

3 Implementation

3.1 IC Tag System

We presume that all the merchandise in supermarkets and department stores will have IC tags in near future as the replacement of bar codes. As the results, most of the items in house and office will have IC tags, which enables the trace of their locations and movements by antenna.

In this research, we employed smart tag system developed by Feig Electronics Co. Ltd. The size of IC tag (sticker label) is about $2cm \times 4cm$. The size of the antenna is about $30cm \times 40cm$ and it can read/write the information from/on IC tags, which are closer than about $15cm$.

We've attached IC tags to the items (cup, glass, pot, instant coffee, tea bag, cream, sugar, potato, carrot, spoon, folk, knife, medicine box, disinfectant, cotton, adhesive plaster, etc.), which are usually available in kitchen or home (see figure 4).

These tag information are obtained by PC via RS-232C serial link. The inference engine is implemented on the PC and the inferred event is transferred to the mobile robot via wireless LAN (see figure 5).

3.2 Inference of Human's Next Action

In order to obtain the learning instances for PrefixSpan, we asked ten subjects to perform four kinds of tasks, which are 1) make a cup of coffee, 2) make a cup of tea, 3) treat a cut on finger, and 4) take a medicine for cold. The examples of the learning instances are as shown in figure 6.

In order to predict precise human behaviors, we've employed not only the IC tag information that de-

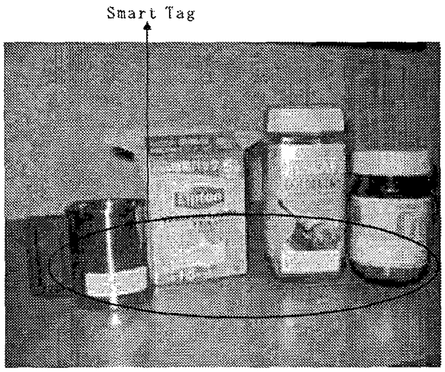


Figure 4: The items labeled by IC tags.

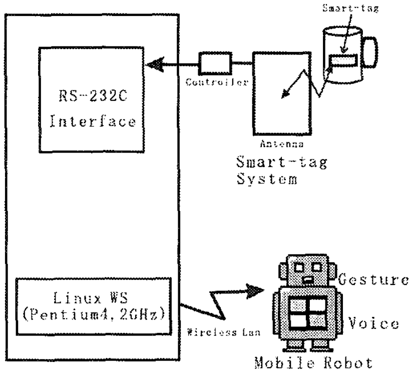
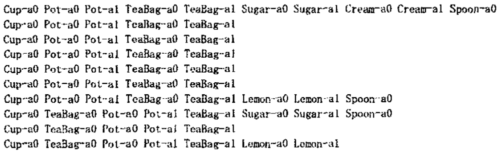


Figure 5: The system configuration of cooking support robot.

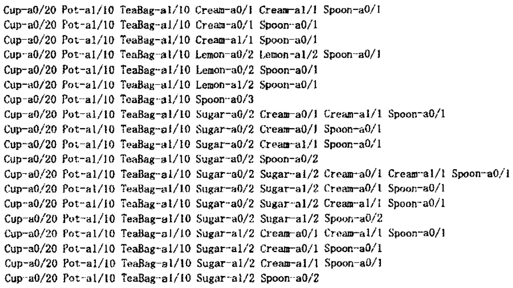


```
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1 Sugar-a0 Sugar-a1 Cream-a0 Cream-a1 Spoon-a0
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1
Cup-a0 Pot-a0 Pot-a1 TeaBag-a0 TeaBag-a1 Lemon-a0 Lemon-a1 Spoon-a0
Cup-a0 TeaBag-a0 Pot-a0 Pot-a1 TeaBag-a1 Sugar-a0 Sugar-a1 Spoon-a0
Cup-a0 TeaBag-a0 Pot-a0 Pot-a1 TeaBag-a1
Cup-a0 TeaBag-a0 Pot-a0 Pot-a1 TeaBag-a1 Lemon-a0 Lemon-a1
```

Figure 6: Example of learning data.

notes the name of item, but also the information of place where it has sensed (a: cupboard, b: cabinet, c: medicine box) and the human action (0: taken out, 1: stored). For example, the event Spoon-a0 denotes that the spoon has taken out from the cupboard.

The time sequence database generated by PrefixSpan from the learning data shown in fugure 6 is as shown in figure 7. For example, the data {Cup-a0/20, Pot-a1/10, TeaBag-a1/10, Spoon-a0/3} denotes that the event Cup-a0 alone has observed 20 times in the learning data. But whole the sequence of {Cup-a0, Pot-a1, TeaBag-a1, Spoon-a0} (the cup has taken out from the cupboard, the pot has stored to the cupboard, the tea bag has stored to the cupboard, and the spoon has taken out from the cupboard) has observed 3 times.



```
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Cream-a0/1 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Cream-a0/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Lemon-a0/2 Lemon-a1/2 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Lemon-a0/2 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Lemon-a1/2 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Spoon-a0/3
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Cream-a0/1 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Cream-a0/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Spoon-a0/2
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Sugar-a1/2 Cream-a0/1 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Sugar-a1/2 Cream-a0/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Sugar-a1/2 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a0/2 Sugar-a1/2 Spoon-a0/2
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a1/2 Cream-a0/1 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a1/2 Cream-a0/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a1/2 Cream-a1/1 Spoon-a0/1
Cup-a0/20 Pot-a1/10 TeaBag-a1/10 Sugar-a1/2 Spoon-a0/2
```

Figure 7: Example of time series data generated by PrefixSpan.

3.3 Cooking Support Robot

We employed mobile robot Robovie developed at ATR [6] for the cooking support robot (see figure 8). We made the robot to recommend the inferred next human action by using synthesized voice and gestures.

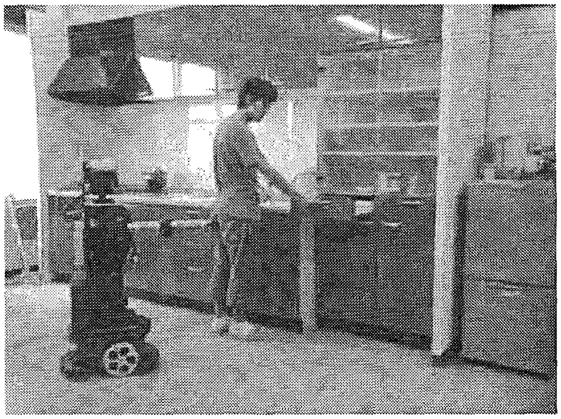


Figure 8: Cooking support robot recommending a presumed next action to human by voice and gesture.

As the whole system, we confirmed that the following supports have realized. When a user took out a cup and an instant coffee from the cupboard, the robot recommends the next action by saying "sugar is in the cupboard" and by turning towards the cupboard and pointing the shelf where the sugar located by its hand. Also, when a user took a medicine for cold and stored it in the medicine box, the robot recommends the next action by saying "the medicine box should be stored in the shelf" by pointing the shelf. These recommendations are automatically generated from the inferred events such as Suger-a0, MedicineBox-b1, etc.

4 Conclusions

In this paper, we proposed human behavior recognition system, which infers the typical next human action by taking account of the accumulated human behaviors observed in the past. We also developed the cooking support robot, which recommends the presumed next human action by voice and gestures.

We are currently conducting the quantitative evaluation of the inferred recommendations by subjects. As the future work, we are planning to extend the system so that it detects more precise and detailed human activities by using heterogeneous sensors such as vision, laser, etc.

References

[1] K. Asaki, Y. Kishimoto, T. Sato and T. Mori, "One-Room-Type Sensing System for Recognition and Accumulation of Human Behavior –Proposal of Behavior Recognition Techniques–," *Proc. of JSME ROBOMEC'00*, 2P1-76-119, 2000.

[2] R. Agrawal and R. Strikant, "Fast algorithms for mining association rules", *Proc. of the 20th International Conference on Very Large Databases*, pp.487-499, 1994.

[3] B. Brumitt et al., "Easy Living: Technologies for Intelligent Environments," *Proc. of International Symposium on Handheld and Ubiquitous Computing*, 2000.

[4] B. Brumitt, B. Meyers, J. Krumm, A. Kern, and S. Shafer, "EasyLiving: Technologies for Intelligent Environments", *Proc. of International Symposium on Handheld and Ubiquitous Computing*, pp.12-29, 2000.

[5] I.A. Essa, "Ubiquitous sensing for smart and aware environments: technologies towards the building on an aware home," *Position Paper for the DARPA/NFS/NIST workshop on Smart Environment*, 1999.

[6] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu, "Robovie: A robot generates episode chains in our daily life", *Proc. of Int. Symposium on Robotics*, pp.1356-1361, 2001.

[7] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale and S. Shafer, "Multi-Camera Multi-Person Tracking for Easy Living," *Proc. of 3rd IEEE International Workshop on Visual Surveillance*, pp.3-10, 2000.

[8] J. Lee, N. Ando, and H. Hashimoto, "Design Policy for Intelligent Space", *Proc. of IEEE International Conference on System, Man and Cybernetics (SMC'99)*, pp.12-15, 1999.

[9] D.J. Moore, I.A. Essa, and M.H. Hayes III, "ObjectSpaces: Context Management for Human Activity Recognition", *Georgia Institute of Technology, Graphics, Visualization and Usability Center, Technical Report*, #GIT-GVU-98-26, 1998.

[10] D.J. Moore, I.A. Essa, and M.H. Hayes III, "Exploiting Human Actions and Object Context for Recognition Tasks", *Proc. of The 7th IEEE International Conference on Computer Vision*, pp.80-86, 1999.

[11] T. Mori, T. Sato et al., "One-Room-Type Sensing System for Recognition and Accumulation of Human Behavior," *Proc. of IROS2000*, pp.345-350, 2000.

[12] Y. Nakauchi et al., "Vivid Room: Human Intention Detection and Activity Support Environment for Ubiquitous Autonomy", *Proc. of IROS2003*, pp.773-778, 2003.

[13] J. Pei et al., "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth", *Proc. of International Conference of Data Engineering*, pp.215-224, 2001.

[14] A. Pentland, "Smart Rooms," *Scientific American*, pp.54-62, 1996.

[15] A. Pentland, R. Picard and P. Maes "Smart Rooms, Desks, and Clothes: Toward Seamlessly Networked Living," *British Telecommunications Engineering*, Vol.15, pp.168-172, July, 1996.

[16] T. Sato, Y. Nishida, and H. Mizoguchi, "Robotic Room: Symbiosis with human through behavior media", *Robotics and Autonomous Systems 18 International Workshop on Biorobotics: Human-Robot Symbiosis*, Elsevier, pp.185–194, 1996.

[17] M. Weiser, "The Computing for the Twenty-First Century", *Scientific American*, pp.94-104, September, 1991.