Applying Ontology and Probabilistic Model to Human Activity Recognition from Surrounding Things

Naoharu Yamada,[†] Kenji Sakamoto,[†] Goro Kunito,[†]
Yoshinori Isoda,[†] Kenichi Yamazaki[†] and Satoshi Tanaka[††]

This paper proposes human activity recognition based on the actual semantics of the human's current location. Since no predefined semantics of location can adequately identify human activity, we automatically identify the semantics from things by focusing on the association between things and human activities with the things. Ontology is used to deal with the various possible representations (terms) of each thing, identified by a RFID tag, and a multi-class Naive Bayesian approach is applied to detect multiple actual semantics from the terms. Our approach is suitable for automatically detecting possible activities even given a variety of object characteristics including multiple representations and variability. Simulations with actual thing datasets and experiments in an actual environment demonstrate its noise tolerance and ability to rapidly detect multiple actual semantics from existing things.

1. Introduction

Owing to the downsizing and increasing sophistication of computing appliances, the ubiquitous computing environment proposed by Weiser [27] is becoming reality. In the ubiquitous computing environment, people will enjoy novel services called "ubiquitous services". Which ubiquitous services are appropriate depends on the user's activities. While traditional services are reactive and uniform for every user, ubiquitous services are proactive and adaptive to each user. For example, when a user is shopping in a food court, the system can tell him what is in his refrigerator and what is missing. When the user is cooking, the system can talk the user through the recipe of an additional dish that uses the left-over food. When the user drops his umbrella in a shop or a train, the system can warn the user. One essential issue in achieving ubiquitous services is how to recognize human activities since the services provided should depend on the user's activities, not his explicit requests. If the system judges the activity wrongly, the user will be annoyed by the useless services.

Some context-aware systems have focused on using location to recognize human activity [16],[26]. Location-based human activity recognition works well to some extent if it is restricted to locations whose capability is strictly limited such as the lavatory or bathroom. However, the location-based approach is not sufficient since it assumes that the semantics of location are static. In fact, most locations allow more than one activity and the activities conducted at a location vary over time. For example, in a living room we take meals at certain times, work at certain times, and chat with friends at certain times.

Some recent papers have focused on recognizing human activities from things [5],[22]. Since most things have a specific function, the system can identify activities from the series of things that the user is using or interacting with. Though this approach can handle dynamically changing activities, it is insufficient since the semantics of location are ignored. A good example is the user who drops a possession. In this case, the location approach has difficulty in distinguishing between unintentionally dropping the possession and intentionally putting it down. By considering where the action occurs, public space or user's own space, the extended approach (thing plus location) can correctly identify the activity. Thus, we need to consider the dynamically changing semantics of location.

This paper presents a method that uses the dynamically changing semantics of location to identify human activities. We refer to the dynamically changing semantics of location as activity spaces. We focus on the things present in the spaces to identify the activity spaces based on the association between things and human activities with the things.

We explain related works in Section 2. Sec-

† NTT DoCoMo, Inc.
†† DoCoMo Technology, Inc.

tion 3 clarifies the definition, characteristics, and technical issues of activity spaces. It also describes our approach; it utilizes ontology and the multi-class naive Bayesian technique. Section 4 explains the preliminary experiments conducted in an ideal environment to verify the effectiveness of ontology, the key issue of our proposal. Section 5 describes the simulations and experiments conducted to verify our proposal in an actual environment.

2. Related Work

Various approaches have been proposed to tackle the essential issue of human activity recognition. Most of the basic approaches focus on location. Location is a key part of context and various location-based services have been proposed [16],[26]. They focus more on how to identify the spatial position of users and less on how to specify the semantics of a spatial position because they assume that the semantics of a spatial position are static and predefined such as a map.

Taking another direction, other research has focused on the user's interaction with things. For identifying things that the user touches or grasps, Moore, et al. [9] utilize a camera, Nishida, et al. [13] utilize an ultrasonic sensor, Tapia, et al. [22] employ environmental state change sensors, while Fishkin, et al. [5] utilize RFID tags. They recognize the user's activities based on the series of things that the user touches or grasps.

However, we consider that all of these approaches are unable to recognize human activities with sufficient accuracy. The location-based approach works well to some extent. For example, if the user stays in the kitchen for some time, s/he may be cooking food. However, the effectiveness of this approach is limited because most locations offer several activities. As an example, consider that upon occasion a living room can support various activities such as studying, working, eating, and playing TV games. A multipurpose room and a park are other examples. Prior work fails to handle these changes in semantics.

The thing-based approach [5],[9],[13],[22] can identify dynamically changing activities by using remote sensors to periodically identify the things that the user is interacting with. However, these approaches have implementation issues with regard to the sensors used and recognition accuracy. As for the issue of sensors, the costs of cameras and ultrasonic sensors are high and the approaches that use these sensors work only in limited areas such as laboratories.

An RFID tag is a low cost sensor that consists of a microchip that stores a unique ID and a radio antenna [4]. They are seen as replacing the barcode in the area of logistics. Some companies or governments now require suppliers to attach RFID tags to every item [20]. EPC Global [3] and Ubiquitous ID center [24] have proposed an ID scheme that makes it possible to put a unique serial number on every item. Considering this background, we can assume that everything will soon have its own RFID tag. This means that RFID tags are the most promising approach to achieving the ubiquitous computing environment.

RFID tags raise two issues with regard to recognition accuracy. One is that the system cannot always detect the things that the user is interacting with since the detection rate of commercial RFID systems is not high. Therefore, the system may wrongly recognize human activities. The other is that the situation is insufficiently identified from the things that the user is interacting with. For example, leaving your camera on the bus is completely different from leaving it at home.

The ultimate goal of our research is human activity recognition. To recognize a user's activity, it is important to identify the actual semantics of the user's location, that is, the activity space as well as the things that the user is interacting with. A system that can identify the actual semantics can distinguish between unintentionally dropping a possession and deliberately leaving it at home by recognizing whether the activity space is the user's own domain or a public space. Our approach is to identify the activity space from the things forming the user's immediate environment.

The things present in a space can be easily detected by RFID tags. Though the detection rate of individual tags is not high, our approach is realistic since the detection of even a few things provides enough context information. The use of RFID tags demands the use of RFID tag readers; we assume that they are either located in the environment or carried by the user. From the perspective of hardware cost, if the target space is small and the number of users is many, the former method is better, otherwise, the latter is better. In this paper, we assume RFID tag readers are placed in the

environment.

3. Activity Space Detection

This section clarifies the definition, characteristics, and technical issues of activity spaces. It also describes our approach to the automatic detection of activity spaces.

3.1 Activity Space

An activity space (AS) is a logically defined space that offers the user some particular activity. For example, "shopping AS" such as supermarkets, flea markets, and stalls, are where users buy commodities. "Eating AS" such as a dining room, restaurants, and cafeterias are where we eat and drink. "User's own domains AS" such as the user's own room in his/her house or hotel where s/he is staying and the user's desk at the office are where the user keeps his/her possessions. Activity spaces are not just spaces defined in terms of X-Y-Z coordinates with no regard for semantics; activity spaces are inherently associated with semantics. Activity space is a subconcept of place. With regard to place, Tuan[23] stated that "place is space infused with human meaning", and Curry[2] mentioned the several ways in which places are created: naming, categorizing, making a symbol, telling stories, and performing activities. In his categorization, an activity space is a place at which activities are performed.

Activity spaces have the following characteristics.

Dynamics of existence Activity spaces are dynamically generated, move, and disappear. For example, "a shopping AS" such as a flea market is dynamically generated, moves, and disappears in parks or squares depending on the action of the booth owners. "An eating AS" is dynamically generated during a meal and disappears after the meal. Each activity space has a different period of existence. While the activity spaces such as a bathroom in a house exist for long periods, the activity spaces such as eating activity spaces or flea markets exist for short periods. This characteristic raises a technical issue: the latter activity spaces cannot be preliminarily identified by maps.

Spatial relationships Several activity spaces can exist at the same spatial position. For example, a living room supports meeting, eating, and working. Therefore, there are spatial relationships among activity spaces such as inclusion, overlap, and adjacency.
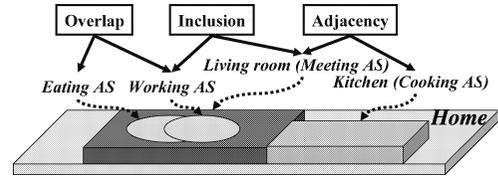

Fig. 1 Spatial relationships among activity spaces.

Figure 1 shows an example. This characteristic raises the point that multiple activity spaces can occupy the same spatial position.

Therefore, a key technical issue is automatic detection of multiple activity spaces.

3.2 Thing-oriented Activity Spaces Detection and Its Difficulties

People can generally identify an activity space simply by "looking at" it. For example, if people look at a kitchen in a house, they can recognize it as a cooking AS, the activity space that enables them to cook. If people look at a market stall, they can recognize it as a shopping AS, the activity space that enables them to buy products.

We focus on things in the space to describe the reason for these assessments. Things, especially artifacts, inherently have functions that enable users to do particular activities. For example, a knife offers people the function of cutting and they can recognize that a knife can be used to cut objects by just looking at the knife. This idea is relatively similar to the concept of affordance introduced by Gibson[6]. Based on this idea, enabled activities can be identified by the functions of things in the space. Therefore, we focus on sets of things to identify activity spaces.

An activity space is identified by four environmental entities: spatial position, time, human, and things. An example of an activity space that is unambiguously identified by spatial position is a praying area. This kind of activity space is relatively static. By comparison, a sunset viewpoint is a watching activity space that exists only when the sun is setting.

However, identifying activity spaces from sets of things suffers from several difficulties. We list them below based on the characteristics of things.

Massiveness People are surrounded by a huge number of things. Therefore, the processing load is high when activity spaces are detected from surrounding things (P1).

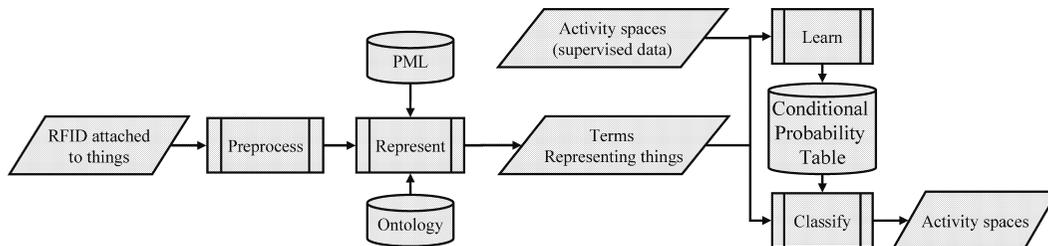**Irrelevance to activities** While some things

Fig. 2 Overall process flow of activity space detection.

are effective in identifying the activity space, others such as lamps, trash, and user's clothing are useless since people do not use these things to perform particular activities. Therefore, tolerance to the noise of irrelevant things is required to correctly identify activity spaces (P2).

Mobility Things can be moved for several reasons. The things that are moved due to the user's intention such as food or dishes for preparing meals are important in identifying the activity space. Since things in a target space can change rapidly, nearly real-time identification of activity space is required (P3).

Multiple representation Each thing has multiple representations. For example, the thing 'pencil' can represent a writing tool and stationery, and at the store, it has the meaning of commercial goods. Thus, being able to handle multiple representations of things is needed (P4).

Variability Things that form the same kind of activity space will be different in different activity spaces. For example, each house has a cooking AS that includes different things. Therefore, manual creation of detection rules is difficult (P5). Furthermore, even if some learning approaches are utilized to automatically extract inference rules, the system cannot deal with things that have not been learned (P6).

While some of the above difficulties arise from the characteristics of things, other problems are created by RFID attributes: RFID tag detection is not completely reliable because of collision and different ID transmission intervals [1].

3.3 Ontology and PMM for Detecting Activity Spaces with Things

To solve the difficulties caused by multiple representations (P4), unlearned things (P6), and some part of thing mobility (P3), we utilize ontology. As for the difficulties caused by the massiveness of things (P1), irrelevant things (P2), thing mobility (P3), the variability of things (P5), and the technical issue of multiple activity space detection, we employ the parametric mixture model (PMM) [25], a text classification method, because we draw an analogy between documents composed of words and activity spaces composed of things. **Figure 2** shows the overall process flow of the proposed system. Though our description of the proposed system assumes the use of EPC Global, this assumption is not a definitive requirement. At the preprocess stage, the system aggregates detected RFID tags and extracts distinct things. For example, the system extracts only things that appeared recently to detect newly generated activity spaces. At the representation stage, the system acquires all terms associated with each thing. We acquire the attribute information of each detected thing from Physical Markup Language servers (PML servers) [14]. Utilizing this information, all terms representing the things are acquired through ontology. At the learning stage, the probability of a thing being in an activity space is specified by utilizing the terms and supervised activity space data. At the classification stage, the system uses PMM to classify a set of terms into activity spaces.

3.3.1 Ontology to Manage Representations

Ontology has a long history in philosophy; it refers to the subject of existence. Although there is no agreed definition of ontology, one definition involves the specification of terms in each domain and the relations among them. Ontology sets "basic concept" to represent underlying concept such as *pencil* and *writing instrument*. In addition, it also sets "is-a relation" to represent the sub concept between two terms. For example, "A pen is-a writing tool" means *a pen* is a sub concept of *a writing tool* [10]~[12]. We call the hierarchy based on these is-a relations the abstraction level. Utilizing these concepts and relations makes it pos-
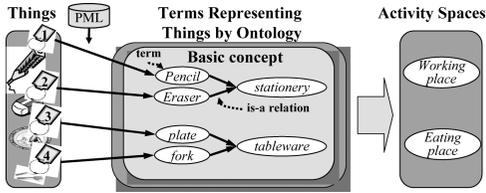
Fig. 3   Representing things by utilizing ontology.

sible to acquire all terms related to a thing by tracing the relations, which resolves the multiple representation problem (P4). **Figure 3** shows an example. Acquired terms of the thing whose ID is 1 are *pencil* and *stationery* by the is-a relation through its basic concept. Here, the terms at the lowest abstraction level in each concept are preliminarily linked to the ID of each thing in PML.

Among all terms related to a thing, we need to identify those that permit activity space detection. The problem of identifying a proper "term" or "abstraction level" is solved in 3.3.2. To solve unlearned things (P6), we transform the terms that have not been learned into the terms that have been learned by raising ontology's abstraction. For example, in Fig. 3, if the thing *eraser* has not been learned but the thing *pencil* has, we can treat both as *stationery*, which has already been learned. In addition, raising the ontology's abstraction reduces the number of kinds of terms. Therefore, the processing load of PMM can be reduced, which resolves the difficulty of thing mobility (P3).

3.3.2   Activity Space Identification via Topic Detection

Many schemes for tackling the identification of the topics of documents or web contents have been proposed. The characteristics of their target are very similar to those of our objective: a document consists of a set of words that includes noise such as stop words[18]; each document on the same topic consists of different words, but people can identify the topic of a document at a glance. Among the many approaches proposed for topic detection, most assume that a document has only one topic; the parametric mixture model (PMM), however, allows one document to have multiple topics. It employs a probabilistic approach, which is efficient and robust against noise; it offers the highest accuracy in detecting multiple topics[25]. Since it is highly likely that multiple activity spaces will be detected from one set of things, we employ PMM. PMM represents documents

as a set of words, known as "Bag of Words" (BOW). It is based on the naive Bayes model[8]. Since the naive Bayes model only provides binary classification, PMM extends this model to provide multi topic detection. PMM assumes that a multi-topic document is composed of a mixture of words typical of each topic. Based on this assumption, a multi-topic document can be represented as the linear summation of the word occurrence probability vector of each topic as shown in Eq. (1). Here, conditional probability $p(t_i|c_l)$ is calculated by MAP estimation. By replacing (words, topics) with (things, activity spaces), we can use Eq. (1) to detect multiple activity spaces from a set of things.

$$p(d|c) = p(t_1, \cdots, t_n|c)$$

$$= \prod_{j=1}^{n} \left(\sum_{l=1}^{L} h_l(y)p(t_i|c_l) \right)^{x_i} \quad (1)$$

$where \ h_l(y) = \frac{y_l}{\sum_{l'=1}^{L} y_{l'}}, \ l = 1, \cdots, L,$

$y_l = 1(y_l \ belongs) \ or \ y_l = 0(y_l \ does \ not \ belong),$
$d : document, \ c : topic, \ x_i : frequency \ of \ word \ t_i,$
$\quad L : \# \ of \ topics, \ n : \# \ of \ word \ kinds$

To select the appropriate abstraction level of is-a relation, conditional probability $p(t_i|AS_l)$ is learned utilizing the thing lists of each abstraction level. PMM then acquires the classification accuracy of activity spaces through the learned conditional probability of each abstraction level. Finally, the abstraction level with the highest classification accuracy is employed to classify test sets of things.

4.   Preliminary Experiment

This section describes the preliminary experiment conducted in an ideal environment to verify the effectiveness of the proposed method, especially the effectiveness of ontology in handling unlearned things. Since there is no related conventional method against which the proposed method can be benchmarked, this paper compares the proposed method with the method without ontology. As for the method without ontology, there are three ways of representing things as follows.

(1)   RFID, attached to things, or manufacturer code, object code, and serial number defined by EPC Global (ex., *32FB34FE* or *X company's ultrafine ballpoint 124515*).

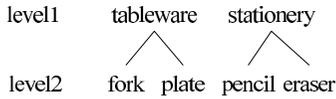(2)   Manufacturer code and object code defined by EPC Global (ex., *X company's*

Fig. 4 Ontology used in preliminary experiment.

ultrafine ballpoint).
(3) The lowest terms in abstraction level of ontology (ex., ballpoint).
If either (1) or (2) is adopted, the classification accuracy of the method without ontology is obviously low since the system can identify activity spaces from the things only when exactly the same thing in case of (1) or the same company's same model thing in case of (2) are learned. Therefore, this paper adopts (3), the lowest terms in abstraction level of ontology in the preliminary experiment, simulations, and full experiments.

To clarify the effectiveness of ontology in handling unlearned things, we simply used 4 kinds of things: forks and plates to identify eating AS and pencil and eraser to identify working AS. The ontology of these things is shown in **Fig. 4**. For learning data, we used 1 plate for eating AS, and 1 pencil for working AS. For test data, we created 11 data sets, each containing 10 things; each had a different mixture ratio of unlearned things to all things. For example, the data set of eating AS whose mixture ratio of unlearned things is 100% contains 10 forks, and the set whose mixture ratio of unlearned things is 40% contains 4 forks and 6 plates.

F-measure is often employed to evaluate the classification accuracy. Therefore, this paper uses F-measure in simulations and experiments. However, the preliminary experiment uses the probability ratio of an activity space to all activity spaces, not the F-measure, to evaluate the estimation result in more detail. The probability ratio of an activity space specifies the level of confidence that can be ascribed to the estimation result. When the probability ratio of an estimated activity space is much larger than that of other activity spaces, the activity space can be identified with confidence. In this case, the system can identify the activity space even if noisy things are added to the thing dataset. On the other hand, when the probability ratio of an estimated activity space is almost the same as that of other activity spaces, the activity space cannot be identified with confidence. In this case, the system may identify the activity space incorrectly if noisy things are added to the thing dataset.
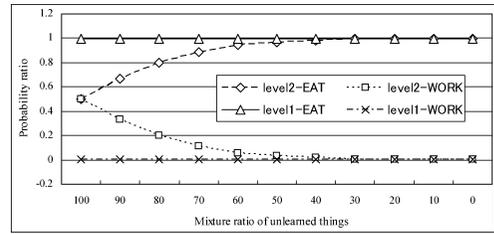

Fig. 5 The probability ratio of abstraction level and activity space for each mixture ratio of unlearned things.

Figure 5 shows the probability ratio of each abstraction level and activity space versus the mixture ratio of unlearned things for the classification of the datasets of eating AS. The results show the effectiveness of ontology in handling unlearned things. When the mixture ratio of unlearned things is 100%, which means all things are unlearned, the probability ratio of both eating AS and working AS is 0.5 for abstraction level 2. This means that the system cannot say which activity space is most probable. By decreasing the mixture ratio, the probability ratio rises since the conditional probability of leaned things is added. On the other hand, the probability ratio of eating AS is almost 1 for abstraction level 1 for all mixture ratios of unlearned things. This is because forks, which are unlearned, and plates, which were learned in level 2 were translated into tableware, which was learned in abstraction level 1.

While the preliminary experiment demonstrated the effectiveness of ontology in an ideal environment, actual environments raise many barriers such as the massiveness of things (P1), irrelevant things (P2), thing mobility (P3), variability of things (P5), and low RFID detection accuracy. Therefore, we conducted simulations and experiments to demonstrate the effectiveness of the proposed method in actual environments.

5. Simulations and Experiments

This section explains the simulations using actual things that were manually collected, and experiments in an actual environment where RFID tags were attached to each thing. The simulations evaluated the performance of the proposed method in terms of activity space detection; we focused on the difficulties caused by the massiveness of things (P1), multiple representations (P4), variability of things (P5), and unlearned things (P6). The experiments con-
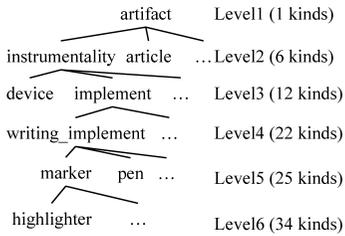
Fig. 6 Term example and the number of kinds of terms at each abstraction level in WordNet.

firmed the feasibility of the proposed method in the face of the difficulties caused by RFID characteristics. To represent things at each abstraction level, we surveyed existing ontology bases. We decided to employ WordNet [28] since it contains the most kinds of terms at each abstraction level and is easily handled by external programs. We set "artifact" in WordNet as abstraction level 1, the highest abstraction level. Instead of utilizing the attribute information stored in PML, we manually set the terms of abstraction level 6, the lowest abstraction level, to directly represent each thing. The terms on abstraction levels 2 to 6 were acquired by utilizing the is-a relation in WordNet (**Fig. 6**).

We implemented PMM in Java and ran the programs on a Pentium 4.3 GHz, 2 GB RAM PC. F-measure, defined as the harmonic mean of precision and recall to evaluate the accuracy of activity space detection. Precision is the ratio of the number of activity spaces correctly identified to the total number of activity spaces identified. Recall is the ratio of the number of activity spaces correctly detected to the number of correct activity spaces.

5.1   Simulations

For the simulations, we considered two cases: sim.1) frequently changing activity space detection to tackle multiple representations (P4), variability of things (P5), and unlearned things (P6), and sim.2) detection of activity spaces that contain a large number of things to address the massiveness of things (P1). For the former, we focused on a dining table in a dining room since it is the site of several activities as described in 3.1. Based on an observation of the activities at a dining table in one author's house, we identify three activity spaces: a meeting AS, an eating AS, and a working AS. Since a meeting AS always exists even if a working AS and an eating AS is generated, the activity spaces that we try to detect are three types: a meeting AS, both a meeting AS and a working AS, and both a meeting AS and an eating AS. We assume that RFID tag readers are put on the table and detect things on or near the table.

Though activity spaces at the dining room table frequently change, they contain relatively few things (26 kinds of 94 things). Therefore, simulation #2 (sim.2.) focused on rooms in a home since each room has many things (836 things: 472 kinds). We examined four activity spaces: a dining room (a meeting AS), a kitchen (a cooking AS), a bath room (a bathing AS), and a study room (a working AS). We assume that each room had several RFID tag readers.

Things datasets acquisition

As for simulation #1 (sim.1), we manually identified all things present on an actual dining table in a dining room. PMM must be informed of the things in each activity space if it is to learn the conditional probability $p(thing_i|AS_l)$. However, when we collect the things of an eating AS or a working AS, the collections will also include the things of a meeting AS since a meeting AS always exists where a working AS or eating AS can exist. Therefore, we acquired the things of just an eating AS or a working AS by eliminating the things of just a meeting AS from those of an eating AS or a working AS. The collections made for the activity spaces are shown in **Table 1**. As for sim.2, we used the sets of things in an actual Korean family's house as collected by the National Museum of Ethnology [21]. They listed all the possessions of an actual family. Since the Korean family's house did not have a study room, we manually identified all things in and on an office desk from photos taken at various angles (**Fig. 7**).

We then added noise to the abstracted data sets with noise ratios of 0%, 25%, and 50%. In more detail, as one half of the noise ratio, we added the things of another activity space to reflect the characteristics that some existing things are not related to the activity space. In addition, as the other half of the noise, we randomly eliminated some things from the data sets to reflect RFID detection errors and things not equipped with RFID tags. By randomly adding noise, we created 1,000 data sets for each activity space. To include unlearned things in the test data for evaluating unlearned things (P6), we used the eating AS of breakfast as learning data and those of lunch and dinner as test data in sim.1.

Table 1 Activity space collections for sim.1.

Activity space	Things
Meeting AS	1 table, 4 chairs and cushions, 4 newspapers, 1 vase, 5 window envelops, 5 ballpoints, 1 in-basket, 1 wastepaper basket, 2 coasters, 1 jotter,
Eating AS (breakfast)	6 dishes, 2 chopsticks, 2 table spoons, 2 mugs, 2 table linens
Eating AS (Lunch)	6 dishes, 2 chopsticks, 2 forks, 2 table knives, 2 glasses, 2 table linens
Eating AS (Dinner)	6 dishes, 2 chopsticks, 2 forks, 2 table knives, 2 glasses, 2 beer cans, 2 table linens
Working AS	2 ballpoints, 4 highlighters, 1 commonplace book, 1 digital computer, 1 power cord, 1 mouse, 7 files
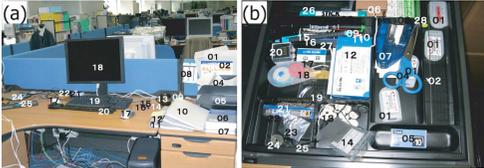
Fig. 7 Manual collection for an actual office desk for sim.2.

Fig. 8 Things equipped with RFID tags.

5.2 Experiments in an Actual Environment

To check the feasibility of the proposed method, we implemented the system and repeated sim.1 in an actual environment. An active RFID tag (RF CODE Spider Ⅲ [15]) was attached to each thing listed in Table 1. RFID transmission interval was 0.4 seconds. RFID tags were detected by TAVIS Concentrator [19]. **Figure 8** shows the things equipped with RFID tags. As the learning data and test data, we acquired RFID datasets for 1 hour each in the various combinations of activity spaces. We then split the datasets into 1 second or 6 second intervals to create learning and test datasets, which means that the system learnt or estimated every 1 second or 6 seconds, respectively.

5.3 Results
5.3.1 Simulations

Table 2 shows the F-measure of each activity space for detecting three types of activity spaces

Table 2 F-measure of detecting each activity space at each abstraction level: sim.1.

AS	Noise	Lev1	Lev2	Lev3	Lev4	Lev5	Lev6
Meet	0%	100	100	100	100	100	100
	25%	100	99.9	100	100	100	100
	50%	100	98.3	100	100	100	100
Eat	0%	0.0	100	100	100	100	100
	25%	0	100	100	100	100	100
	50%	0.0	100	100	100	99.8	100
Work	0%	0.0	100	100	100	100	100
	25%	0.0	75.9	93.5	94.2	96.4	90.5
	50%	0.0	62.9	81.0	79.3	86.1	78.0

in sim.1. The proposed method successfully detected each activity space with a high degree of accuracy. Detection accuracy was lowest, slightly, for the working AS since some of the things that are discriminative of a working AS are also discriminative of a meeting AS such as ballpoints and jotters. On the other hand, a meeting AS and eating AS have discriminative things. Therefore, while the multiple activity spaces of working and meeting can be successfully detected, the single meeting AS was classified as the multiple activity spaces of working and meeting.

This result also demonstrates the noise tolerance of the proposed method since the accuracy of activity space detection did not drop as the noise ratio was raised. Furthermore, the accuracy of working AS detection increased when the ontology's abstraction was raised. Raising the ontology's abstraction decreased the number of kinds of terms: 1 kind in level 1, 6 kinds in level 2, and 34 kinds in level 6 (Fig. 6). This means that the information amount decreased and the accuracy of activity space detection generally fell when the ontology's abstraction rose. Ontology can provide an explanation: each activity space has many kinds of discriminative terms but only a few instances of each; the use of ontology raised the ontology's abstraction which yielded fewer kinds of discriminative terms but a larger number of each. Note that it makes sense that the F-measure is 0 at level 1 in most activity spaces since the
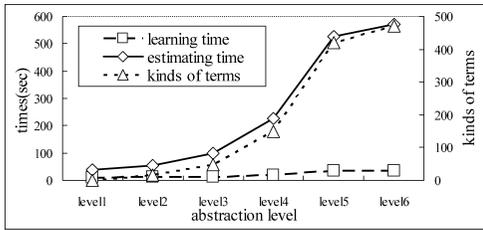
Fig. 9 Processing time for learning and estimating 4,000 sets of things, and average number of kinds of terms contained in the sets of things with regard to each abstraction level.

Table 3 F-measure of detecting each activity space at each abstraction level: sim.2.

AS	Noise	Lev1	Lev2	Lev3	Lev4	Lev5	Lev6
Bath	0%	0.0	100	100	100	100	100
	25%	0.0	97.1	95.6	99.8	98.9	100
	50%	0.0	86.6	92.2	99.4	97.8	100
Cook	0%	40	100	100	100	100	100
	25%	40	91.1	96.5	98.7	99.4	100
	50%	40	82.5	92.5	97.6	96.9	100
Meet	0%	100	100	100	100	100	100
	25%	100	94.6	97.2	98.1	98.8	100
	50%	100	89.9	94.0	97.5	97.7	100
Work	0%	0.0	100	100	100	100	100
	25%	0.0	90.6	97.5	99.3	99.4	100
	50%	0.0	83.9	95.1	98.4	97.4	100

Fig. 10 Detection rate of things in each activity space combination (Numbers in parentheses indicate the number of things).

Table 4 F-measure of detecting each activity space at each abstraction level and each interval in the experiment.

Intvl	AS	Lev1	Lev2	Lev3	Lev4	Lev5	Lev6
1 sec	Meet	99.7	75.8	92.7	96.5	96.7	95.7
	Eat	1.2	90.7	90.7	91.1	86.0	86.7
	Work	0.0	65.0	85.0	90.5	90.3	90.8
6 sec	Meet	100	100	100	100	100	100
	Eat	0.0	100	100	100	100	100
	Work	0.0	91.4	100	100	100	100

term of abstraction level 1 is just "Artifact". As for unlearned things data, we did not learn forks, table knives, and glasses. WordNet transformed forks, table knives, and tablespoons into cutlery, which had been learned, at abstraction level 5. Glasses and mugs that were learned were also transformed into containers at abstraction level 3. Therefore, ontology can utilize unlearned things for activity space detection by raising ontology's abstraction. The appropriate abstraction level to identify activity spaces depends on the target activity spaces. The results of sim.1 show that abstraction level 5 yielded the highest accuracy while for sim.2, abstraction level 3 offered the highest accuracy. Therefore, the appropriate abstraction level needs to be specified in the learning phase.

Figure 9 shows the processing time taken for learning and estimating 4,000 sets of things data and the number of kinds of terms at each abstraction level. This demonstrates that the proposed method can rapidly handle large sets of things and that increasing the abstraction level makes it possible to reduce the processing time. Furthermore, though 472 kinds of things were aggregated into 17 kinds at abstraction level 2, the F-measure of each activity space in **Table 3** only slightly decreased,
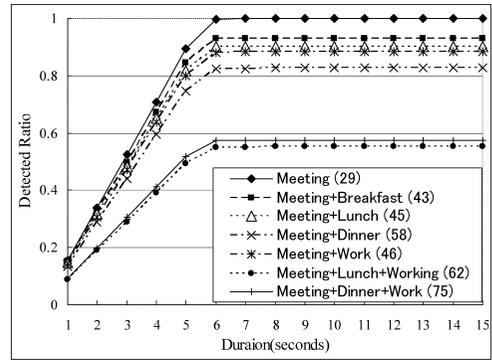
which clearly demonstrates the effectiveness of ontology. The reason for the slight decrease in F-measure when the ontology's abstraction was raised is because learning data and test data cover almost the same things since there is only 1 dataset for each activity space.

5.3.2   Experiments

Figure 10 shows the detection rate of things in each combination of activity spaces. Though each RFID sent its ID every 0.4 seconds, it took 6 seconds to stabilize at a certain ratio. Furthermore, the detection rate stayed under 60% when over 60 things were present on the table. The reason for this is that some RFID tags were attached to metallic things such as beer cans and others to things that were placed under other things. The detection rate was 9 – 15% (3 – 7 things) and 54 – 100% (18 – 42 things) for the intervals of 1 second and 6 seconds, respectively. **Table 4** shows the F-measure of each activity space in detecting combinations of activity spaces for the intervals of 1 second and 6 seconds. The results demonstrate that the proposed method can identify activity spaces in actual environments. As in the case of sim.1, the F-measure of working AS was slightly lower than those of the other activity spaces. In addition, abstraction by ontology raised the F-

Table 5 Cosine similarity of the terms' conditional probabilities between two activity spaces (M:Meeting AS, E: Eating AS, W: Working AS).

ASs	Lev1	Lev2	Lev3	Lev4	Lev5	Lev6
M - E	0.9993	0.1833	0.0798	0.0048	0.0002	0.0050
M - W	0.9994	0.7995	0.1238	0.0382	0.0144	0.0218
E - W	0.9994	0.2966	0.0003	0.0003	0.0003	0.0003

measure of eating AS.

At the interval of 6 seconds, most activity spaces were successfully detected. For the interval of 1 second, the F-measure is too high since the system detected only 3 – 7 things. The reason is that, in this experiment, we used only things distinctive of the activity space. For example, we used the distinctive things of Working AS such as a digital computer and a mouse for learning and estimating Working AS. In addition, each activity space has a different set of discriminative things. To prove this fact, we calculated the cosine similarity of feature vectors among the activity spaces. The feature vector is provided by the term and its conditional probability for each activity space. In **Table 5**, the similarity at abstraction levels 3, 4, 5, and 6 is pretty low, which demonstrates that the activity spaces have completely different discriminative things. Therefore, even if just one thing belonging to Meeting AS was detected, the system identified the Meeting AS. Since the dataset has few things, the influence of one thing on activity space detection was relatively high.

5.4 Discussion

The above simulations and experiments demonstrate the ability of our proposal to rapidly detect multiple activity spaces and resist noise. Though the findings of this experiment are meaningful and interesting, activity space detection is just the first step in achieving the ultimate goal, that is, human activity recognition. The remaining issues to achieve human activity recognition are as follows.

Human activity recognition with multiple ASs By identifying the activity spaces where the user exists, the system can identify the user's activity. However, since multiple activity spaces can overlap, the system needs to discriminate between them to accurately identify the user's activity. This issue can be solved by identifying the activity space generated most recently since a user often intentionally moves things to perform a particular action such as preparing a meal. Another approach is identifying the thing that the user is interacting with and selecting the activity space whose conditional probability of the thing is highest.

Target activity spaces While this experiment focused on 3 or 4 activity spaces, the variety of target activity spaces must be expanded and refined. With regard to expansion, we need to acquire as many activity spaces as types of human activities. With regard to refinement, we need to split each activity space if possible into more detailed activity spaces. For example, a meeting AS has the subconcepts of director's meeting AS and group meeting AS. Ontology would be helpful in achieving this.

Concepts of ontology In this paper, we utilized only the basic concept to acquire the terms representing things. However, ontology also offers other concepts such as "role concept". "Role concept" represents the role that a thing plays in a particular domain such as *product*. By utilizing the role concept, the system can identify other kinds of activity spaces such as selling AS. Since no existing ontology base defines role concept, we need to build the ontology of role concept.

PMM for activity space detection When PMM is employed for topic detection from text, stop word elimination is an efficient way of improving performance since so many words remain. When PMM is employed for activity space detection, on the other hand, the amount of information available may be very little. For example, an eating AS can be created anywhere the user brings takeout food. This issue can be solved by weighting things in a preprocessing step.

6. Conclusion

This paper proposed a novel approach to detect activity spaces, spaces that offer a human a particular activity, for achieving the ultimate goal of human activity recognition. Focusing on the association between things and enabling activities with the things, activity spaces are identified through the things that exist there. We utilize ontology to specify terms representing things and the multi-class naive Bayesian approach to identify activity spaces from the terms. Simulations and experiments have demonstrated that the proposed system

has excellent noise tolerance, high accuracy in activity space detection, and the ability to rapidly handle large amounts of data. Though we focused on human activities with things, other activities that are independent of things remain to be recognized. However, they may depend on other real world entities such as human context or time context and in that case, our approach is still applicable.

References

1) Floerkemeier, C. and Lampe, M.: Issues with RFID Usage in Ubiquitous Computing Applications, *Proc. Pervasive Computing 2004* (*PERVASIVE*), Vol.3001 of Lecture Notes in Computer Science, pp.188–193 (2004).

2) Curry, M.: *The Work in the World — Geographical Practice and the Written Word*, University of Minnesota Press, ISBN 0-8166-2665-0 (1996).

3) EPC Global. http://www.epcglobalinc.org/

4) Finkenzeller, K.: *RFID Handbook: Radio-Frequency Identification Fundamentals and Applications*, John Wiley & Sons (2000).

5) Fishkin, K., Jiang, B., Philipose, M. and Roy, S.: I Sense a Disturbance in the Force: Unobtrusive Detection of Interactions with RFID-tagged Objects, *Proc. 6th Int'l. Conf. on Ubiquitous Computing* (*UbiComp2004*), pp.268–282 (2004).

6) Gibson, J.: *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Assoc Inc., ISBN: 0898599598 (1979).

7) Gruber, T.: A translation approach to portable ontologies, *Knowledge Acquisition*, Vol.5, No.2, pp.199–220 (1993).

8) McCallum, A. and Nigam, K.: A Comparison of Event Models for Naive Bayes Text Classification, *Proc. Int'l. Workshop on Learning for Text Categorization in AAAI-98* (1998).

9) Moore, D., Essa, I. and Hayes, M.: Exploiting Human Actions and Object Context for Recognition Tasks, *Proc. 4th Int'l. Conf. on Computer Vision* (*ICCV'99*) (1999).

10) Mizoguchi, R.: Tutorial on ontological engineering Part1: Introduction to Ontological Engineering, *New Generation Computing*, OhmSha & Springer, Vol.21, No.4, pp.365–384 (2003).

11) Mizoguchi, R.: Tutorial on ontological engineering Part2: Ontology development, tools and languages, *New Generation Computing*, OhmSha & Springer, Vol.22, No.1, pp.61–96 (2004).

12) Mizoguchi, R.: Tutorial on ontological engineering Part3: Advanced course of ontological engineering, *New Generation Computing*, OhmSha & Springer, Vol.22, No.2 (2004).

13) Nishida, Y., Kitamura, K., Hori, T., Nishitani, A., Kanade, T. and Mizoguchi, H.: Quick Realization of Function for Detecting Human Activity Events by Ultrasonic 3D Tag and Stereo Vision, *Proc. 2nd IEEE Int'l. Conf. on Pervasive Computing and Communications* (*PerCom2004*), pp.43–54 (2004).

14) PML Core Specification 1.0. http://www.epcglobalinc.org/standards_technology/Secure/v1.0/PML_Core_Specification_v1.0.pdf

15) RF CODE. http://www.rfcode.com/

16) Schilit, B., Adams, N., Gold, R., Tso, M. and Want, R.: ParcTab Mobile Computing System, *Proc. 4th Workshop on Workstation Operating Systems* (*WWOS-IV*), pp.34–39 (1993).

17) Seon-Woo, L. and Mase, K.: Activity and Location Recognition Using Wearable Sensors, *Pervasive Computing*, pp.10–18 (Sep. 2002).

18) Stop list. ftp://ftp.cs.cornell.edu/pub/smart/english.stop

19) Tavis Concentrator. http://www.rfcode.com/tavis.asp

20) The American Department of Defense. http://www.defenselink.mil/releases/2003/nr20031023-0568.html

21) The National Museum of Ethnology: Seoul Style 2002, ISBN: 4915696465 (2002).

22) Tapia, E., Intille, S. and Larson, K.: Activity Recognition in the Home Using Simple and Ubiquitous Sensors, *Proc. 2nd Int'l. Conf. on Pervasive Computing 2004* (*Pervasive2004*), pp.158–175 (2004).

23) Tuan, Y.: Space and Place: The Perspective of Experience, ISBN: 0816638772 (1977).

24) Ubiquitous ID Center. http://www.uidcenter.org/

25) Ueda, N. and Saito, K.: Singleshot detection of multi-category text using parametric mixture models, *Proc. 8th Int'l. Conf. on Knowledge Discovery and Data Mining* (*SIGKDD2002*), pp.626–631 (2002).

26) Ward, A., Jones, A. and Hopper, A.: A New Location Technique for the Active Office, *IEEE Personal Communications*, Vol.4, No.5, pp.42–47 (1997).

27) Weiser, M.: The Computer for the 21st century, *Scientific American*, pp.94–104 (Sep. 1991).

28) WordNet. http://wordnet.princeton.edu/

Naoharu Yamada received his B.E. and M.I. degrees from Kyoto University in 2001 and 2003, respectively. He has been with NTT DoCoMo, Inc. since 2003. His current research interests are context-awareness in the ubiquitous computing environment. He is a member of the Information Processing Society of Japan.

Kenji Sakamoto received his B.E. and M.E. degrees in Electrical Engineering from Keio University, Yokohama, Japan in 1999 and 2001, respectively. Currently he is working in NTT DoCoMo, Inc. He is mainly engaged in research on wireless LAN communication systems. He received 2000 TAF Telecom System Technology Student Award. He is a member of IEICE.

Goro Kunito received his B.E., M.E., and Dr.E. degrees in electrical engineering from the University of Tokyo in 1995, 1997, and 2000, respectively. He is currently a research engineer at NTT DoCoMo's Research Laboratories. His research interests include distributed, mobile, and ubiquitous computing. He is a member of IEEE, IPSJ, and IEICE.

Yoshinori Isoda received a master's degree from the Department of Systems Engineering, the University of Osaka in 1993. Since joining NTT in 1993, he has been researching sensor processing systems and ubiquitous computing systems. He is currently a Manager of Application Systems Development Group, NTT DoCoMo, Inc. He is a member of the Information Processing Society of Japan and the Robotics Society of Japan.

Kenichi Yamazaki received his M.S. degree from Tohoku University in 1986 and his Dr.Eng. degree from the University of Electro-Communications in 2001. He is currently with Research Laboratories at NTT DoCoMo, Inc. His interests include ubiquitous computing, mobile service provisioning, and transport protocol for mobile communications. He is a member of the Information Processing Society of Japan and the Association for Computing Machinery.

Satoshi Tanaka received a B.E. and M.E. degrees from Keio University, Japan, 1985 and 1987, respectively. Since joining NTT in 1987, his research interests include object-oriented systems, distributed systems, and ubiquitous computing systems. He is currently a Senior Manager of Multimedeia Division in DoCoMo Technology, Inc. He is a member of ACM and the Information Processing Society of Japan.