

## Activity Recognition based on Intra and Extra Manipulation of Everyday Objects

Dipak Surie<sup>1</sup>, Fabien Lagriffoul<sup>1</sup>, Thomas Pederson<sup>1</sup>, and Daniel Sjölie<sup>2</sup>

<sup>1</sup>Department of Computing Science, Umeå University,  
S-901 87 Umeå, Sweden  
{dipak, fabien, top}@cs.umu.se

<sup>2</sup>VRlab / HPC2N, Umeå University,  
S-901 87 Umeå, Sweden  
deepone@hpc2n.umu.se

**Abstract.** Recognizing activities based on an actor's interaction with everyday objects is an important research approach within ubiquitous computing. We present a recognition approach which complement objects grabbed or released information with the object's internal state changes (as an effect of intra manipulation) and the object's external state changes with reference to other objects (as an effect of extra manipulation). The concept of Intra manipulation is inspired by the fact that many everyday objects change their internal state when manipulated by the human actor, while extra manipulation is motivated by the fact that humans commonly rearrange the spatial relations between everyday objects as part of their activities. A detailed evaluation of our prototype activity recognition system in virtual reality (VR) environment is presented as a "proof of concept". We have obtained a recognition precision of 92% on the activity-level and 81% on the action-level among 15 everyday home activities. Virtual reality was used as a test-bed in order to speed up the design process of our activity recognition system, allowing us to compensate for the limitations with currently available sensing technologies and to compare the contributions of intra manipulation and extra manipulation for activity recognition.

**Keywords:** Activity Recognition, Context Awareness, Ubiquitous Computing, Wearable Computing, Virtual Reality.

### 1 Introduction

In the recent years, there have been many research efforts that attempt to use computers for supporting human activities performed in the real world [1], [2]. Such systems are not only useful in providing assistance to elderly people and those with cognitive impairments in performing their activities of daily living, but also in other

application areas like providing training to newly employed staffs, providing assistance in specialized activities like surgery, etc. Advancements in sensing, processing, communication and storage technologies have played an inspiring role for such research efforts. To build such systems is a challenging task due to the number and variety of activities performed by human actors in the real world. One interesting challenge is to come up with a general approach for modelling and recognizing human activities with finer-granularity at not only activity level, but also at action and operation level. There have been several attempts in recognizing the actor's current activity based on object manipulation information, in particular based on the objects that are grabbed and released information [3], [4]. This is an interesting approach since most of the physical activities performed by humans in the real world are mediated by objects [5]. We take a similar research stance, but extend this approach to also consider the object's internal state changes and the object's external state changes as an effect of the actor's intra manipulation and extra manipulation respectively for the following reasons:

- a) Recognizing activities based on objects grabbed or released information alone is a promising approach, but there are no previous work to our knowledge that has used this information alone in recognize activities at a lower abstraction action level [3], [4], and [6]. However an extension of such an approach has shown promising results at the action level in our previous work [6].
- b) Recognizing activities based on objects grabbed or released information alone has also had difficulties in recognizing the activities during the initial phase of an activity, when the activity has just begun<sup>1</sup>. This introduces long temporal delay between the actual starting of an activity and the moment the system makes a guess about the actor's current activity, which we have addressed in this paper.
- c) Such an approach has also had difficulties in recognizing the end of an action or an activity due to the unavailability of sharper events that could be generated using intra manipulation and extra manipulation information channels to be discussed later in this paper. This issue is also discussed in [6].

The wearable computing community has investigated activity recognition based on wearable accelerometers [7], [8], [9] microphones [7] and even cameras [10]. Approaches based on accelerometer data are restricted to activity recognition of simple activities like walking, running, etc. that involve the actor's body movements. Approaches using microphones and cameras have had difficulties in extracting high-level features without extensive computation. The ubiquitous computing community has investigated activity recognition using simple state change sensors [11] and using objects grabbed or released information [3], [4]. The approach using objects grabbed or released information is interesting because complex activities like *preparing the table for lunch* or *preparing breakfast* could be recognized with high precision and recall values. However such an approach does not address the three challenges described in a), b) and c). We consider the objects' internal state change data in the context of an actor's interaction with that object concerned (similar to [11]) and also its changed relationships to other objects from a perspective centred on how the actor literally perceives the world. The work within this line of research focus has mainly

---

<sup>1</sup> We need to recognize activities during the initial phase of an activity in order for our activity support application to provide assistance before the activity reaches an irreversible state.

been driven by currently available sensor technology. In this paper, we introduce a conceptual design platform based on intra and extra manipulation that could survive and handle the generations of changes in the field of sensor technology, and present a detailed evaluation of the system as a “proof of concept”.

**Application Area: Activity Support for People Suffering Dementia.** The prototypical activity recognition system discussed in this paper is part of a larger system that aims to provide assistance to people suffering early stages of dementia disease in completing their activities of daily living (ADL) [2]. ADL include getting dressed, preparing breakfast and activities related to personal hygiene. Typical problems include the forgetting of performing an activity or an action within an activity; not being able to get started in the first place; not being able to continue after having been interrupted; or missing some operations that are mandatory for the completion of an activity. A system that could help overcome the above mentioned problems would enable patients to stay in their home for a longer period of time, have a normal independent life, and also reduce the burden on family members and caregivers. The long-term goal is to build a dementia tolerant home environment using ubiquitous and wearable computing technologies<sup>2</sup>.

**Structuring Human Activities: Activity, Action and Operation.** According to activity theory [5], human activities have an objective and are mediated through tools. We consider the objects present in the actor’s environment as tools for the actor to accomplish his/her activities. This theory introduces a 3-level hierarchy of activity, action and operation. An activity takes place in several situations, where each situation is comprised of a set of actions under certain conditions like, location, time, etc. An action is a conscious goal-directed process performed by an actor to fulfil an objective and is comprised of a set of operations. Operations are unconscious processes that depend on the structure of the action and the environment in which it takes place. We follow the above mentioned definitions and the 3 levels of granularity in modelling and recognising human activities.

## 2 Intra and Extra Manipulation of Everyday Objects

**Theoretical Background:** Distributed Cognition moves the boundary of human cognition outside of the head of an individual to include his/her body parts and the environment as part of a functional system [12]. According to this theory, human cognition is distributed by placing memories, facts, or knowledge in the objects, individuals, and tools in our environment. Within our research, we keep track of the state changes to everyday objects in the actor’s environment based on an actor’s object manipulation as part of performing an activity. Intra manipulation is inspired

---

<sup>2</sup> We use VR as a test-bed to develop a system that can assist dementia patients in performing their ADL (Fig. 1a.). Dementia patients will not be asked to perform the experiments in a virtual reality environment. Based on our initial results in VR (activity recognition system described in this paper) we are currently developing an actual hardware prototype which will be evaluated by patients suffering early stages of dementia.

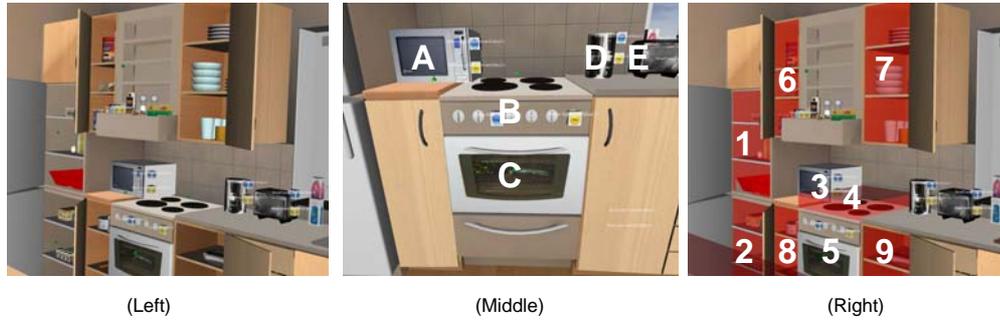
by the fact that many everyday objects change their internal state when manipulated by the human actor, while extra manipulation is motivated by the fact that humans commonly arrange and rearrange the spatial relations between everyday objects as part of their activities [13]. For example, a fridge could be considered as a container that contains objects like milk packet, juice bottle, cake box, etc. When the actor removes the milk packet from the fridge, then he/she has actually extra manipulated the milk packet with reference to the fridge and intra manipulated the fridge. Similarly, an actor might turn on the stove, where the actor has intra manipulated the stove by changing its internal state.

## 2.1 Applying the Concept of Intra and Extra Manipulation

Our operational definition of intra manipulation and extra manipulation is as follows:

- **Intra manipulation (IM).** Any operation that changes the internal state of an object is known as intra manipulation. When an actor interacts with everyday objects, some objects might change their internal state resulting in the following events:  $\langle objectID, \{is\_grabbed, is\_released, is\_activated, is\_deactivated, is\_opened, is\_closed\} \rangle$ . Refer to Fig. 1 (Middle) for all the objects that can change their internal state based on an actor's interaction with it. We consider the *objectID* of all the objects the actor is holding in his/her hands (we do not make a difference between the left hand and the right hand) every 1 sec between their respective *is\_grabbed* and *is\_released* events. This information is complemented with the everyday objects' internal state information between *is\_activated* and *is\_deactivated* events or between *is\_open* and *is\_close* events (when the object is in operation) every 1 sec to obtain what we refer to as the IM information channel.
- **Extra manipulation (EM).** Any operation that changes the external state of an object is known as extra manipulation. When an actor interacts with everyday objects, some objects might change their external state with reference to other objects resulting in the following events:  $\langle objectID, containerID, \{has\_entered, has\_left\} \rangle$  also referred to as the EM information channel. *ObjectID* refers to the object the actor is currently interacting with, while *containerID* provides information about the object that contained or will contain the object the actor is currently interacting with. Refer to Fig. 1 (Right) for all the objects that can contain other objects. Container objects include fridge, freezer, cupboard, dining table, etc. in our VR simulated home environment. The external state change information includes the relationship change between the object the actor is currently interacting with and the container object in terms of if the object has

entered the container or has left the container. The volumes sensitive to extra manipulation are shown in Fig. 1 (Right).<sup>3</sup>



#### Internal state change events:

- A. *<Microwave Oven, {is\_activated, is\_deactivated, is\_opened, is\_closed}>*
- B. *<Stove, {is\_activated, is\_deactivated}>*
- C. *<Oven, {is\_activated, is\_deactivated, is\_opened, is\_closed}>*
- D. *<Coffee Maker, {is\_activated, is\_deactivated}>*
- E. *<Bread Toaster, {is\_activated, is\_deactivated}>*
- F. *<Rice Cooker, {is\_activated, is\_deactivated}>*
- G. *<Tap, {is\_activated, is\_deactivated}>*

#### External state change events:

- 1. *<objectID, Fridge, {has\_entered, has\_left}>*
- 2. *<objectID, Freezer, {has\_entered, has\_left}>*
- 3. *<objectID, Microwave Oven, {has\_entered, has\_left}>*
- 4. *<objectID, Stove, {has\_entered, has\_left}>*
- 5. *<objectID, Oven, {has\_entered, has\_left}>*
- 6. *<objectID, Cupboard1, {has\_entered, has\_left}>*
- 7. *<objectID, Cupboard2, {has\_entered, has\_left}>*
- 8. *<objectID, Cupboard3, {has\_entered, has\_left}>*
- 9. *<objectID, Cupboard4, {has\_entered, has\_left}>*
- 10. *<objectID, Sink, {has\_entered, has\_left}>*
- 11. *<objectID, Table, {has\_entered, has\_left}>*

**Fig. 1.** VR home environment (Left) with volumes sensitive to extra manipulation marked in red colour (Right) and objects that possess internal states (Middle).

<sup>3</sup> In parallel to VR simulation we are currently working on a hardware prototype based on passive RFID technology for sensing EM events. Passive RFID tags are attached to everyday objects, while RFID readers are worn on the actor's wrists for sensing *<objectID, {is\_grabbed, is\_released}>* events and attached to a selected set of containers for sensing *<objectID, containerID, {has\_entered, has\_left}>* events in a real home environment used for ubiquitous computing research. Simple state change sensors like on-off switches, light sensors, pressure sensors, temperature sensors, etc. are attached to selected objects in the home environment for sensing *<objectID, {is\_activated, is\_deactivated, is\_opened, is\_closed}>* events. Information about EM events and IM events are communicated to a wearable computer using ZigBee communication protocol.

When objects are manipulated by an actor, the following two events  $\langle objectID, \{is\_grabbed, is\_released\} \rangle$  alone are considered for activity recognition in [3] and [4]. In our previous work [6], we have compared the contributions of such an approach with two other information channels (observable space and manipulable space) and have discussed the limitations of considering an approach based on objects grabbed or released information alone. In this paper we extend such an approach to include IM and EM information channels for recognizing activities with higher precision values and sharper events that are some of the implications for building a system that can provide reliable assistance to people suffering mild dementia in completing their activities of daily living.

### 3 Activity Recognition System

#### 3.1 Virtual Reality as a “Test-Bed”

VR was used as a test-bed [6] in order to speed up the design process of our activity recognition system, allowing us to compensate for the limitations with currently available sensing technologies and to compare the contributions of intra manipulation information channel and extra manipulation information channel for activity recognition. A VR model, developed using the Colosseum3D real-time physics platform [14] is used to simulate a physical home environment with wearable sensors and sensors embedded on selected everyday objects to capture an actor’s intra manipulation (IM) events and extra manipulation (EM) events. Fig. 1 (Left) shows a snapshot of our VR environment. Refer to Fig. 2 for the activity recognition system architecture. We have experimented with 78 object types. Object types include simple object types like *fork*, *knife*, *plate* etc. that does not change their internal state, complex object types like *microwave oven*, *stove*, *oven*, *tap*, etc. that can potentially change their internal states and container object types like *fridge*, *freezer*, *cupboard*, *dining table*, etc. that can contain other objects. 7 objects have internal states and 11 objects are container objects. We only consider the *type* of object in recognizing activities, not the identity (e.g. *fork\_1* and *fork\_2* are both considered as *fork* type).

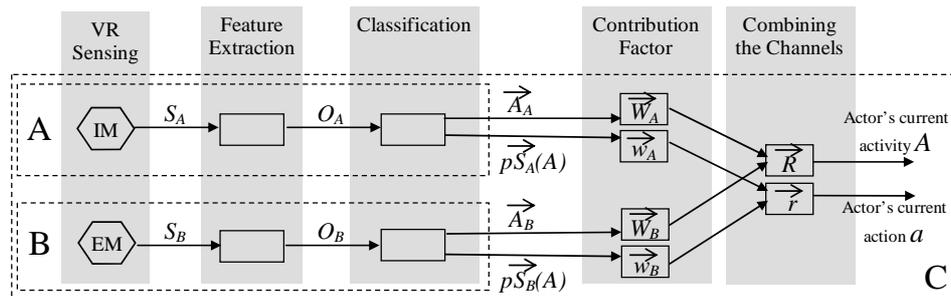
There are many objects that overlap for several activities. For instance, simple objects like *fork*, *knife*, *plate*, etc. are used for several activities like *preparing table for lunch*, *having lunch*, *having coffee-break*, *doing the dishes*, etc. Similarly complex objects like *stove*, *oven*, *microwave oven*, etc. and container objects like *fridge*, *freezer*, *cupboard*, etc. are also used for many activities. This makes the classification problem harder compared to taking an approach where the recognition system is strongly characterised by one or two objects that are unique to the activity. We do have some activities like for instance *preparing\_rice*, where the *rice\_bag* and the *rice\_cooker* are unique objects for this activity. But this does not simplify the classification problem for the following reasons also discussed in [6]: 1) we are not only recognizing the actor’s current activity, but also the actor’s current action. The *rice\_bag* and the *rice\_cooker* are not unique objects for all the actions within the activity of *preparing\_rice*, but only for some actions; 2) the *rice\_bag* manipulation or

the *rice\_cooker* manipulation might be a noise created by the actor while performing another activity, and 3) the recognition system should recognize the activity and the action before they are actually completed to provide appropriate assistance to the actor. Hence the system cannot wait until the unique object is manipulated to recognize the activity and the action.

### 3.2 Feature Extraction and Classification

IM and EM information channels consist of sets of events that need to be quantified every second. Our quantification scheme builds  $\vec{S}_A$  and  $\vec{S}_B$  as shown in Fig. 2, where  $\vec{S}_A$  represent the set of distinct *eventIDs* calculated using *objectIDs* and the object's internal state, while  $\vec{S}_B$  represent the set of distinct *eventIDs* calculated using *objectIDs*, *containerIDs* and the object's external state change.

The probabilistic generative framework of hidden-markov model (HMM) [15] is used because of its clear Bayesian semantics, its ability to handle time-varying signals and the availability of efficient algorithms for state and parameter estimation. HMMs reduce the system's configuration space into a number of finite discrete states together with the probabilities for transition between the states. One limitation of HMMs is that the model structure has to be user-defined, which includes the number of states and the connections between the states. The model structure cannot be determined by standard learning methods. This should not pose a major problem since the activities recognized are user-defined (also discussed in [6]). The actor provides ground truth for both activities and actions. Each activity is modelled using a separate HMM with the number of states corresponding to the number of actions within that activity. Similarly, the transitions between states correspond to the transitions between different actions within that activity. HMMs have shown good results in many activity recognition systems including [16], [17], [18] and [4].



**Fig. 2.** The proposed activity recognition system architecture: A and B corresponds to information channels IM and EM respectively. C corresponds to a combination of IM and EM with automatically generated weights based on their contributions in recognizing individual activities.

The activity recognition system uses two information channels (refer to Fig. 2). Each information channel produces a sequence of observations that are fed into fifteen HMMs (one for each activity). For each information channel, the outputs from the fifteen HMMs are used to build an activity probability vector ( $\vec{A}_A$  and  $\vec{A}_B$ ) containing the probabilities for each possible activity. The element of the activity probability vector with the highest value gives the actor's current activity and its most probable state gives the actor's current action.

### 3.3 Combining the Information Channels

The two information channels are combined using activity contribution factors ( $\vec{W}_A$  and  $\vec{W}_B$ ) and action contribution factors ( $\vec{w}_A$  and  $\vec{w}_B$ ).  $\vec{W}_A$  and  $\vec{W}_B$  consist of the recognition precision values for each activity using IM information channel and EM information channel respectively, while  $\vec{w}_A$  and  $\vec{w}_B$  consist of the recognition precision values for each action using IM information channel and EM information channel respectively. The contribution factors are automatically generated from the training data. We first determine the actor's current activity by computing  $\vec{R}$ , the weighted sum of both the information channels using the following formula:

$$\vec{R} = \vec{W}_A * \vec{A}_A + \vec{W}_B * \vec{A}_B \quad (1)$$

Where \* represents element-by-element multiplication. The element of  $\vec{R}$  with the highest value gives the actor's current activity  $A$ . Once the activity is known, we determine the actor's current action by calculating  $\vec{r}$  using the following formula:

$$\vec{r} = \vec{w}_A * p\vec{S}_A(A) + \vec{w}_B * p\vec{S}_B(A) \quad (2)$$

Where  $p\vec{S}_A(A)$  and  $p\vec{S}_B(A)$  are the vectors of state probabilities of the HMM representing the actor's current activity  $A$  for intra manipulation and extra manipulation respectively.  $\vec{w}_A$  and  $\vec{w}_B$  is equal to zero if their respective channel is not supportive to the activity  $A$  determined previously. The element of  $\vec{r}$  with the highest value gives the actor's current action. We have not combined the two information channels before classification since we want to evaluate their contributions independently and combine them based on their contribution in recognizing individual activities. Such an approach also provides the possibility to include additional channels without affecting the overall infrastructure of our activity recognition system.

## 4 Evaluation

The experiments were performed by 5 subjects (none of them are affiliated to the system development team) in a virtual reality home environment<sup>4</sup>. 15 activities of daily living were included as shown in Table 1. The activities were performed 10 times as part of various scenarios. A scenario comprises of a few related activities performed in some sequence. We have used 7 scenarios: *lunch scenario (1)*, *lunch scenario (2)*, *coffee-break scenario (1)*, *coffee-break scenario (2)*, *baking scenario (1)*, *baking scenario (2)* and *cleaning scenario*. Some activities were common for several scenarios like the activity of *preparing table for lunch* which is common to both the *lunch scenario (1)* and the *lunch scenario (2)*. All the subjects were allowed to perform the activities in their own way (often in many different ways).

**Table 1.** List of activities and actions based on the AMPS framework [19].

Activities	Actions within individual activities
Preparing rice	Bring rice, Pour rice, Pour water, Add salt, Switch ON rice cooker, Replace rice
Preparing vegetables	Bring vegetables, Cut vegetables, Place pan on the stove, Switch ON stove, Fry vegetables in oil, Add spices and salt, Switch OFF stove
Baking cake	Bring baking dish, Switch oven ON, Add eggs, Add milk, Add sugar, Add cake powder, Place baking dish into oven, Switch oven OFF, Remove baking dish
Preparing cake	Get cake, Put cake into microwave oven, Switch ON microwave oven, Remove cake from microwave oven
Preparing coffee	Take coffee powder, Pour water into coffeemaker, Switch ON coffeemaker
Having coffee-break	Pour coffee, Drink coffee, Cut cake, Eat cake
Having lunch	Have main meal, Have dessert, Have coffee
Preparing table for coffee-break	Bring cake and coffee, Bring cutlery
Preparing table for lunch	Place mats, Bring cutlery, Bring plates, Bring juice glasses, Bring the food, Place napkins
Doing the dishes	Bring the dishes, Brush the dishes, Rinse the dishes, Wash hands
Preparing apple pie	Bring the baking dish, Bring the butter, Bring flour, Bring apples, Cut apples, Switch oven ON, Place baking dish into oven, Switch oven OFF, Remove baking dish
Preparing pasta	Get pasta, Switch stove ON, Fill casserole with water, Add salt, Set the timer and switch stove ON, Drain with the colander
Preparing pasta sauce	Place pan on the stove, Add oil, Switch stove ON, Bring onions and cut them, Place onions in the pan, Bring tomatoes and cut them, Place tomatoes in the pan, Add spices and salt, Switch stove OFF
Preparing tea	Fill casserole with water, Switch stove ON, Add tea bags, Switch stove OFF
Cleaning kitchen	Take sponge and spray, Clean the microwave oven, Clean the oven, Clean the stove, Clean the sink, Clean the table, Replace stuff and clean hands

When a subject begins performing his/her activity, each object is in the location where it was last placed in the subject's previous activities. This makes our experiments realistic compared to having a fixed initial location for each objects. Cases when the subjects dropped an object on the floor or grabbed the wrong object or performed an inappropriate object state change were also included in our dataset. A real chair was used for the subjects to perform the activities of *having coffee-break* and *having lunch* that obliged them to sit down. Subjects' body postures and

<sup>4</sup> The subjects were initially taught how to perform activities in a virtual reality environment and then given a time period to practice in this environment. Only when the subjects were comfortable with the environment, they were allowed to perform the activities.

locomotion within the VR environment were realistic. For instance, the subjects were not allowed to pass through a table, even though it is possible in a VR environment.

#### 4.1 Precision, Recall and Confusion Matrix

The average *number of events* generated by IM information channel is 42 for each action, while that of EM information channel is 3.9 for each action. The *observation sequence lengths* were empirically determined for individual information channels based on a trade-off between the precision and recall values. An optimal *observation sequence length* of 6 is used for both IM and EM information channels. We used the “Leave-One-Out Cross-Validation” (LOOCV) scheme to obtain the precision and recall figures. Cross-validation was used to validate our classification considering our limited, but sufficient datasets from the 5 subjects. Refer to Table 2 and Table 3 for precision, recall and confusion matrix. We define precision and recall as follows:

$$precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

**Table 2.** Precision (P) and recall (R) in percentage (%) for each activity (Act) and action (An) using the two information channels (A and B) and by combining the two information channels (C). The last row represents global values (G) in percentage (%).

Act #	Intra Manipulation (IM)				Extra Manipulation (EM)				Combining IM and EM			
	Activity		Action		Activity		Action		Activity		Action	
	P	R	P	R	P	R	P	R	P	R	P	R
1	81	100	66	100	90	100	90	100	86	100	40	100
2	47	99	35	99	73	100	73	100	96	100	81	100
3	92	100	76	100	81	94	81	94	89	95	62	93
4	76	98	45	96	45	60	45	60	97	100	78	100
5	89	96	89	96	67	99	67	99	91	100	66	100
6	94	98	91	98	79	99	79	99	87	100	71	100
7	69	100	69	100	94	99	94	99	87	100	83	100
8	57	100	57	100	93	100	93	100	98	100	94	100
9	46	99	46	99	91	100	91	100	95	100	77	100
10	63	98	46	95	98	97	83	97	94	100	69	100
11	66	99	55	99	96	87	96	87	95	100	52	100
12	72	99	44	99	82	100	81	100	88	91	64	89
13	57	98	42	98	90	98	90	98	92	100	60	100
14	93	96	89	95	62	84	62	84	91	99	86	99
15	73	98	26	94	94	99	89	99	95	100	61	100
G	72	98	58	98	82	94	81	94	92	99	70	99

## 5 Discussion

**Information Channels Reinforce Each Other at the Activity Level.** By combining the two information channels, we obtain a recognition precision of 92% at the activity level. Such a high precision is possible due to the combination of the information channels that represent different and complementary aspects of the actor’s activities. In Fig. 3, for the activity of *preparing cake*, action 2 (*Put cake into microwave oven*) and action 3 (*Switch ON microwave oven*) are recognized using IM as the information channel. This is due to the fact that the internal state of the *microwave oven* changes

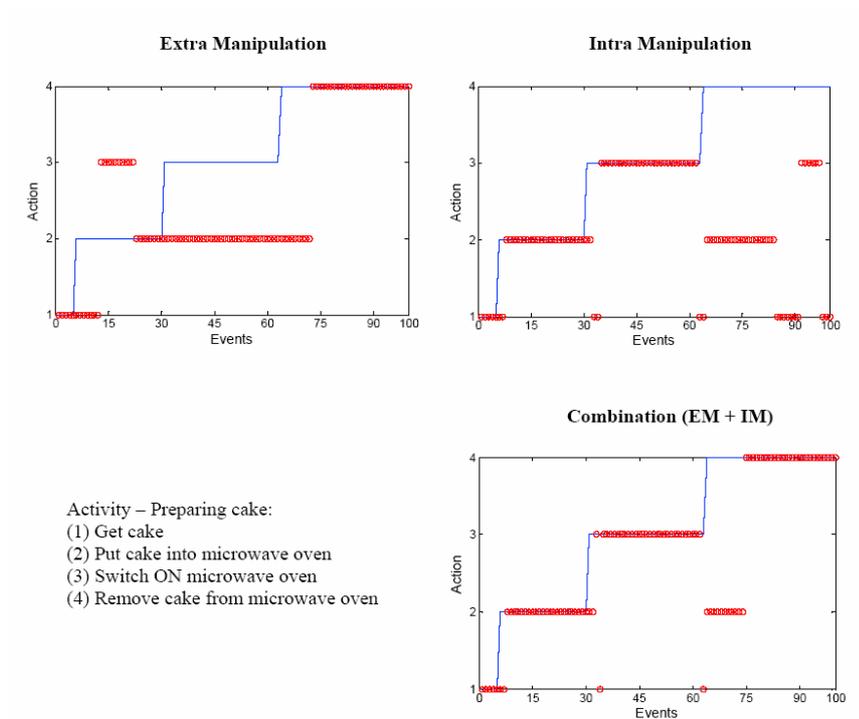
from *is\_close* to *is\_open* and then to *is\_close* once again during action 2, while the *microwave oven* changes its internal state from *is\_deactivated* to *is\_activated* and then to *is\_deactivated* during action 3. Action 4 (*Remove cake from microwave oven*) is confused with action 2 because during both these actions, similar events are generated. In this case EM provides more reliable information since *<Cake, Microwave Oven, has\_left>* event is generated that is unique for action 4. Action 1 (*Get cake*) is recognized equally well using both IM and EM as information channels.

Table 3. Confusion matrix.

		Recognized Activities															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Actual Activities	1	595	0	0	0	0	0	0	0	0	0	0	0	0	0	95	
	2	31	954	0	4	0	0	0	0	0	0	0	0	0	3	0	0
	3	0	56	648	0	0	0	0	0	0	0	2	0	25	0	0	0
	4	0	0	17	481	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	23	236	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	31	233	0	5	0	0	0	0	0	0	0	0
	7	0	0	0	0	9	107	972	0	32	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	10	460	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	70	1351	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	204	3226	0	0	0	17	0	0
	11	0	0	13	0	0	0	0	0	0	19	669	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	50	745	0	54	0	0
	13	0	102	0	0	0	0	0	0	0	0	0	37	1643	0	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	20	209	0	0
	15	0	0	0	0	0	1	0	0	0	0	0	0	0	68	1320	0

**Information Channels Complement Each Other at the Action Level.** At the action level, EM shows a good precision, but the recall value of 94% indicates that some actions may not be detected at all. A closer look reveals that actually 21 actions among the 15 activities are not detected even once, which is not acceptable in building a reliable activity assistive system. But by combining IM and EM information channels, the recall value is increased from 94% to 99%, there by allowing all the actions within the 15 activities to be detected at the cost of a lower overall recognition precision.

**Temporal Delay in Recognizing Activities and Actions.** The output of the activity recognition system may sometimes be unstable, especially around the transitions between two activities or between two actions. In order to provide a reliable assistance based on the proposed activity recognition system, the output of the activity recognition system must be smoothed which introduces temporal delays between the actual starting of an activity or action and the moment when the system makes a guess about the actor's current activity or action. An activity is recognized on an average after 30 events. At the action level, this delay varies between 7 and 24 events (13 on an average), which means that in some cases (especially if the duration of an action is short), the system may guess the actor's current action with a delay of one or two actions.



**Fig. 3.** IM and EM information channels complement each other for activity recognition.

**Implications for building an Activity Assistive System.** In our previous work [6], we have experimented with different information channels and have obtained an activity recognition precision of 89% among 10 activities of daily living. One main reason for building over our previous efforts by exploring complementary information channels is to recognize activities with higher precision (we have achieved an activity recognition precision of 92% among 15 activities of daily living) and sharper events that can be used for both activity recognition and in providing real-time assistance based on the actor’s current activity. Examples of sharper events include *turning ON a stove* or *taking out juice packet from the fridge*. Events that are generated using EM and IM information channels are sharper compared to mere objects grabbed or released events, since EM and IM information channels contain information about a specific object’s state change information. Events generated using observable space and manipulable space information channels (refer to our previous work [6]) also does not yield sharper events that can directly be used to provide assistance to the user. Even though we obtain sharper events using EM and IM information channels, there exists to some extent uncertainty and delay in recognition. To address this limitation which is inherent due to the probabilistic nature of our activity recognition system, we perform data mining on the training data and extract some events that are mandatory for each activity. Such events are referred to as mandatory events. For

instance, if an actor *switches ON the stove* during the activity of *preparing vegetables* in all the training data, then *switching ON the stove* is considered to be a mandatory event. Such mandatory events are also used along with the actor's current activity and action in providing assistance to the actor.

## 5 Future Work

**Combining IM and EM Information Channels.** At present the combination between IM and EM information channels is performed based on simple weights (generated considering the individual information channels' activity recognition precision), improving the recognition precision at the activity level to 92% (see Table 3). However, the combination procedure at the action level has reduced the recognition precision from 81% (EM information channel alone) to 70% (combining IM and EM information channels). We are currently investigating the cause to this negative effect and intend to address it. At present, we have also not considered the temporal relationship between the two information channels. We intend to include such relationships and investigate the possibility of improving the recognition precision at the action level using the combination of the two information channels.

**Improving the Granularity of IM Information Channel.** Complex objects that can change their internal states are important information to capture in recognizing user activities. However, in our study we have used only a selected number of complex objects to avoid the necessity of too much sensing and computation in the environment. Among those selected complex objects, objects that are unique for individual activities like rice cooker, coffee maker, etc. have contributed well for activity recognition. However devices like stove and oven that are common for several activities provide less information for our activity classifier and introduce noise in recognition. This is because we sense the internal state of the stove or the oven at a coarse granularity (only on/off states). We believe that by improving the sensing of the complex objects' internal states with finer granularity, like for instance, sensing the temperature of the stove, there is a reasonable chance in improving the recognition at both the activity and the action level for many of the activities included for experimentation. For instance, the temperature of the stove changes differently during the activity of *preparing tea* compared to the activity of *preparing pasta sauce*. However, this does not mean that improved granularity always contribute to the performance improvement of recognizing activities. Further work needs to be done to compare the relationship between performance improvement and granularity of sensing the IM information channel.

**Exploring the Spatial Relations between Everyday Objects.** At present, the EM information uses simple relationship between the object the actor is currently interacting with and the container changes due to such interaction. However there may be some relationship between the object the actor is currently interacting with and the other objects that are inside or on the container object. Such relations will be explored in the future. Also when an object enters or leaves a container object, the

internal state of the container changes, which will be considered as being part of IM information channel in the future.

**Transferring to Real-World Applications.** Our approach of using virtual reality as a test-bed introduces the issue of how this translates to real-world applications. Virtual-reality simulation implies that there is no noise and uncertainty in the collected signals, which is an important factor in real-world applications. As mentioned earlier, we are targeting on using passive RFID technology [20] considering its reliability in identifying the objects the actor is currently interacting with [21], [3], and [4]. Sensing EM information channel introduces many challenges including the difficulties in limiting the volume of the container object that is sensitive to EM events and in attaching RFID readers on devices like oven that might be used at high temperatures. We are aware of some passive tags that can handle high temperatures, and currently investigating on RFID readers that can handle high temperatures. Similarly there are issues that need to be solved in sensing intra manipulation information channel. However the focus of this paper is not to get too much carried away by the technology that is existing today, but instead to check the contributions of intra and extra manipulation with the assumption that sufficient technology will be available to sense these information channels in the near future. Even though the ecological validity cannot be guaranteed, our approach is a novel one and is primarily intended for guiding the development of ubiquitous and wearable computing systems capable of assisting human activities. Other issues like scalability of our approach and adaptation of our system to variations in activity patterns are discussed in [6].

## 7 Conclusions

In this paper we have presented a prototype activity recognition system developed based on an actor's interaction with everyday objects. Our activity recognition approach includes the objects grabbed or released information with the objects' internal state changes and their external state changes with reference to other objects. When evaluated in a virtual-reality simulated home environment: 1) activity and action recognition accuracies using EM has shown promisingly better results compared to currently dominant IM based approaches [3], [11]; 2) by combining both the information channels we have obtained a recognition precision of 92% at the activity-level among 15 activities of daily living. Our work provides an activity-aware platform for further investigations into the development of personal and user-defined activity assistive systems. As a secondary focus, we have also presented the approach of developing ubiquitous and wearable computing systems using virtual-reality simulation [6].

## Acknowledgement

We would like to thank Anders Backman, Björn Sondell, Gösta Bucht, Kenneth Bodin, Lars-Erik Janlert, and Marcus Maxhall from Umeå University, Sweden. This work is partially funded by the EC Target 1 structural fund program for Northern Norrland, Sweden.

## References

1. Fishkin, K., Consolvo, S., Rode, J., Ross, B., Smith, I., Souter, K. Ubiquitous Computing Support for Skills Assessment in Medical School, UbiHealth Workshop at the Sixth International Conference on Ubiquitous Computing (2004)
2. Backman, A., Bodin, K., Bucht, G., Janlert, L.-E., Maxhall, M., Pederson, T., Sjölie, D., Sondell, B., & Surie, D. easyADL – Wearable Support System for Independent Life despite Dementia. Workshop on Designing Technology for People with Cognitive Impairments, CHI2006, April (2006) 22-23
3. Philipose, M., Fishkin, K., Perkowitz, M., Patterson, D., Fox, D., Kautz, H., Hähnel, D. Inferring Activities from Interactions with Objects, IEEE Pervasive Computing, October (2004) 50-57
4. Patterson, D., Fox, D., Kautz, H., Philipose, M. Fine-Grained Activity Recognition by Aggregating Abstract Object Usage, Ninth IEEE International Symposium on Wearable Computers, (2005)
5. B. Nardi. (ed.): Context and Consciousness: Activity Theory and Human-Computer Interaction. Cambridge: MIT Press, (1995)
6. Surie, D., Pederson, T., Lagriffoul, F., Janlert, L.-E., & Sjölie, D. Activity Recognition using an Egocentric Perspective of Everyday Objects. In Proceedings of IFIP 2007 International Conference on Ubiquitous Intelligence and Computing (UIC2007), Springer LNCS 4611, July (2007) 246-257
7. Ward, J. A., Lukowicz, P., Troster, G., Starner, T. Activity Recognition of Assembly Tasks Using Body-Worn Microphones and Accelerometers, Pattern Analysis and Machine Intelligence, IEEE Transactions on Vol. 28, No. 10, October (2006) 1553-1567
8. Bao, L., Intille, S. Activity Recognition from User-Annotated Acceleration Data, Second International Conference, PERVASIVE 2004, Linz/Vienna, Austria. LNCS 3001, April (2004) 1-17
9. S.W. Lee and K. Mase. Activity and location recognition using wearable sensors. IEEE Pervasive Computing 1(3), (2002) 24-32

10. Mayol, W., Murray, D. Wearable hand activity recognition for event summarization, Ninth IEEE International Symposium on Wearable Computers, October (2005) 122-129
11. Tapia, E., Intille, S., Larson, K. Activity Recognition in the Home Using Simple and Ubiquitous Sensors, Second International Conference, Pervasive 2004, Linz/Vienna, Austria. LNCS 3001, April (2004) 158-175
12. Hutchins, E. *Cognition in the Wild*, MIT Press, ISBN 0-262-58146-9 (1995)
13. Pederson, T. From Conceptual Links to Causal Relations — Physical-Virtual Artefacts in Mixed-Reality Space. PhD thesis, Dept. of Computing Science, Umeå university, report UMINF-03.14, ISBN 91-7305-556-5, (2003)
14. Backman, A. Colosseum3D – Authoring Framework for Virtual Environments. In Proceedings of EUROGRAPHICS Workshop IPT & EGVE Workshop, (2005) 225-226
15. Rabiner, L. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, Vol. 77, No. 2, February (1989)
16. Chen, J., Kam, A., Zhang, J., Liu, N., Shue, L. Bathroom Activity Monitoring Based on Sound, Third International Conference, Pervasive 2005. LNCS 3468, April (2005) 47-61
17. Lukowicz, P., Ward, J., Junker, H., Stäger, M., Tröster, G., Atrash, A., Starner, T. Recognizing Workshop Activity Using Body Worn Microphones and Accelerometers, Second International Conference, Pervasive 2004, Linz/Vienna, Austria. LNCS 3001, April (2004) 18-32
18. J. Lester, T. Choudhury, N. Kern, G. Borriello, B. Hannaford. A Hybrid Discriminative/Generative Approach for Modeling Human Activities. 19<sup>th</sup> International Joint Conference on Artificial Intelligence, (2005)
19. AMPS. <http://www.ampsintl.com/> as on 2<sup>nd</sup> January (2007)
20. Finkenzeller, K. *RFID Handbook*. John Wiley and Sons, New York, NY, USA, Second edition, (2003)
21. Pederson, T. Magic Touch: A Simple Object Location Tracking System Enabling the Development of Physical-Virtual Artefacts in Office Environments. Journal of Personal and Ubiquitous Computing, Vol. 5. Springer (2001) 54-57