A Sensor and Video Based Ontology for Activity Recognition in Smart Environments

D. Mitchell-*IEEE Student Member*, Philip J. Morrow, *IEEE Member*, and Chris D. Nugent, *IEEE Member*

***Abstract*— Activity recognition is used in a wide range of applications including healthcare and security. In a smart environment activity recognition can be used to monitor and support the activities of a user. There have been a range of methods used in activity recognition including sensor-based approaches, vision-based approaches and ontological approaches. This paper presents a novel approach to activity recognition in a smart home environment which combines sensor and video data through an ontological framework. The ontology describes the relationships and interactions between activities, the user, objects, sensors and video data.**

I. INTRODUCTION

Activity recognition is the process of recognising what activities are being performed within an environment. This is an essential task in a wide range of applications including the topics of security, surveillance and healthcare. In a healthcare scenario activity recognition can be used to monitor Activities of Daily Living (ADLs). These tasks are routinely carried out by a person on a daily basis and include for example the activities of preparing food, bathing and dressing. When dealing with the elderly or disabled, ADLs can be used to monitor health and wellbeing status and, to a certain extent, can be used to indicate whether support at home is required. In this work we present a novel ontological framework combining sensor and video data which aims to provide a more accurate approach to activity recognition. The approach developed can be used to monitor and recognise ADLs being performed within a kitchen environment. We believe that this approach will help to solve the problems associated with uncertain or incomplete sensor data in addition to the issues of occlusions when dealing with video data.

The remainder of the paper is organized as follows. Section 2 describes related work surrounding the topic of activity recognition. In Section 3 the Ontological Framework is presented with a brief overview of the relationships to be considered. Preliminary results and a discussion surrounding the work undertaken to date are considered in Section 4. Section 5 presents Conclusions from the study and opportunities for further work.

This research is supported by the Department of Employment and Learning, Northern Ireland.

Deborah Mitchell is with the School of Computing and Information Engineering, University of Ulster, Northern Ireland (e-mail: Mitchell-d4@email.ulster.ac.uk).

Philip J. Morrow is with the School of Computing and Information Engineering, University of Ulster, Northern Ireland (e-mail: pj.morrow@ulster.ac.uk).

Chris D. Nugent is with the School of Computing and Mathematics, University of Ulster, Northern Ireland (e-mail: cd.nugent@ulster.ac.uk).

II. RELATED WORK

There have been many different approaches used to undertake the task of activity recognition. These have been mainly based on the use of sensor based or video based data.

Sensor based activity recognition has been used to infer what activities have taken place in an environment [1]. Sensors can be attached to objects in the environment or they can be worn by the user. Wearable sensors have been used to collect physiological data, for example, pulse rate [2] and used to classify daily activities [3]. Multiple sensors can be used to provide more information about the activities taking place [4] where the sequence of the sensor activations is used to understand the activity being performed. Sensor data can, however, suffer due to uncertainty in the data or incomplete data, for example, due to sensor failure.

Video based activity recognition uses video cameras to monitor the activities being performed in the environment. These systems have been successfully used to detect and track objects and people within an environment [5] and [6]. Different types and numbers of cameras have been used in an effort to solve problems related to occlusions and changes in illumination.

The fusion of sensor and video data can provide more information about the environment and the activity being carried out. Sensor data can be used to provide rich information related to the time when an object is interacted with whilst video data can be used to provide further information about the objects within the scene, for example, the number of people, objects, colour and texture. As an example in [7], recorded sensor data was compared to video data to determine who was closest to a sensor when it was activated. This was used to infer who had performed a particular task.

In [8] an ontology was used for the task of activity recognition where various sources of information were available. The ontology accommodated for the relationships between sensors, objects and the activities being performed. The use of ontologies based on video data is important when taking into account security and surveillance systems. Ontologies to detect events from video were presented in [9] for a surveillance setting and in [10] for a soccer domain. In [11] and [12], a formal language was developed, Video Event Representation Language (VERL) to describe the events taking place within the recorded video data.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

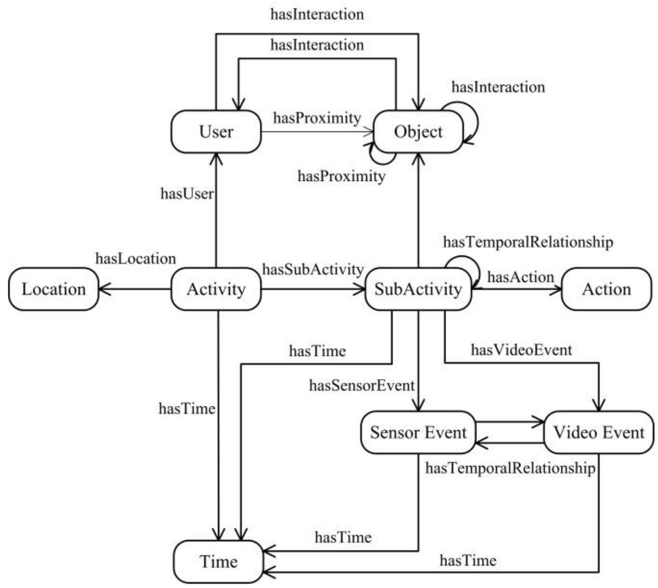


Figure 1 Sensor and Video Based Ontological Framework

A significant amount of work in activity recognition has been undertaken using these approaches; however, less has been done to fuse data from both sensors and video in an ontological framework.

III. ONTOLOGICAL FRAMEWORK

The aim of the current work is to create a sensor and video based ontology to be used for the purposes of activity recognition within a smart home. The ontological framework is shown in Figure 1 and provides a way in which to describe the relationships and interactions between the activity, video, objects and the users during an activity. In addition, the ontology represents the associations between the sensor and video data gathered from the smart environment during the activity. The integration of the information in this manner will be used to provide a more accurate representation of the activity being performed.

The activity is central to the ontology and has a direct relationship with the user, the location and the sub-activities. For the purposes of this paper an activity is considered to be a complex task, for example, *'Make a Cup of Tea'* and consists of a sequence of sub-activities. These sub-activities can be defined as: *'boil kettle'*, *'get cup'*, *'get teabag'* and *'pour water'*. Every activity is undertaken by a user in the environment. For the purposes of this study it is assumed that a single activity will be performed by a single user. The activity is linked to the location by the facilities that are required to complete the activity and objects required when performing the activity. For example, if the user is making lunch some of the facilities they will require may for example include access to a cooker, fridge and some food.

Consideration must also be given to the sub-activities that make up the general activity. These sub-activities can be described in terms of their essential or optional sub-activities. Essential sub-activities must be performed for the purpose of completing the general activity. The sub-activities are described by the actions and objects required to complete the sub-activity. Actions are simple events that are performed by the user during the sub-activity for example, *'lift kettle'*. The user will interact with various objects whilst performing the sub-activities. These objects can be categorized as either *'static'* or *'dynamic'* objects. For example, a tap is defined as a *'static'* object that the user will interact with during the *'boil kettle'* sub-activity and a cup is defined as a *'dynamic'* object that the user will interact with during the *'get cup'* sub-activity. In addition, the objects will have interactions with other objects, for example the kettle will interact with the tap when it is being filled with water.

The user will have *proximity* to the objects during the activity. The user has to be near the kettle or tap to enable them to interact with it. It is also important to consider the proximity of the objects to other objects during the activity. For example, the kettle will be '*over'* the cup when the user is pouring the water.

Temporal relationships exist between the sub-activities and describe the order in which one sub-activity is performed in relation to another sub-activity. Temporal relationships between these sub-activities can be described as being *'before', 'after'* or *'during'*. It can also be stated that a sub-activity may *'enable'* or *'depend on'* another sub-activity. For example, *'pour water'* depends on *'boil kettle'* and *'get cup'*.

A set of sensor and video events are produced during the performance of the sub-activities. The sensor and video events will have time-stamped data that will be used in the analysis of the activity. The activity, sub-activities and events are all related by time and they will all have a start time, an end time and duration. Temporal relationships exist between the sensor and video events. A video event may happen at the same time as a sensor event. Alternatively, a sensor event may occur during, before or after a video event. For example, if the cupboard door is opened the sensor will change state and this is recorded as a sensor event. At the same time, the video cameras will record the cupboard door opening as a video event. The associated time-stamps will be used in the analysis of the activity.

IV. RESULTS AND DISCUSSION

The sensor and video ontology was created using the Protégé ontology editor[1]. The purpose of the ontology is to combine the sensor and video data to improve the process of identifying what activity is being undertaken, the sequence of the sub-activities being performed, who has performed the activity and what objects were used during the activity. This initial ontology has been based on a single person performing a single activity, *'Make a Cup of Tea'*. Although this considers one ADL it is the intention that the ontology will be extensible. The initial process was undertaken to prove that the ontology had the ability to combine sensor and video data. For the purposes of simplicity it is assumed that the sub-activities will be performed in a sequence, one sub-activity will follow the other until the activity is completed. The ontology describes the relationships between the activity, the user, the objects and the sensor and video events.

[1] http://protege.stanford.edu

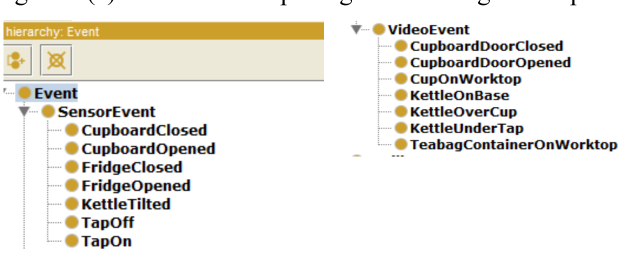
The sensors and video cameras installed in the environment are used to observe and record the activities being performed. We believe that the combination of the sensor and video data provides a better representation of the activity being performed; however, it is important to define these relationships in the ontology. During the initial stages of planning the activities were considered in terms of the essential sub-activities that must be performed in order to complete the activity. These sub-activities were then studied in further detail to determine a) what objects are required to perform this sub-activity? b) what actions will be undertaken when performing the sub-activity? c) what are the temporal relationships between the sub-activities? d) what events are produced during the sub-activities?

A. The Environment

Sensor and video datasets were recorded in the Smart Environments Research Group laboratories within the School of Computing and Mathematics on the Jordanstown campus, University of Ulster[2]. The kitchen is fitted with high and low level units, a sink, a refrigerator and a microwave. In addition, contact sensors have been installed on the kitchen access doors, the cupboard doors, the refrigerator door and the microwave door. These sensors are used to indicate when the doors are opened and closed. A tilt sensor is attached to the kettle and is used to indicate when the kettle has been used, for example when pouring boiled water into a cup. A contact sensor is also attached to the tap and is used to indicate when the tap has been turned on/off. Each sensor has a unique ID which is used to determine what object has been interacted with. Multiple cameras are installed within the kitchen to record the user when they are performing activities. Two cameras are mounted at either end of the kitchen worktop, under the wall mounted cupboards, to record the activity being performed along the worktop area. A third camera is mounted on the wall to record the kitchen cupboards, the fridge and the microwave.

B. Events

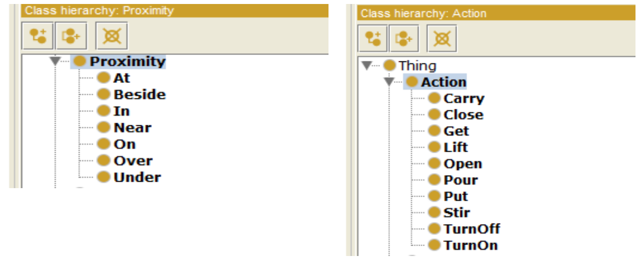
Figure 2 presents a snapshot of sensor and video events that have been defined within the ontology for each sub-activity. These events can be described by the objects of interest during the sub-activity. For example, when performing the *'boil kettle'* sub-activity the objects of interest are the kettle and the tap. The sensor events are detected by the change in state of the sensor. Therefore the actions of tilting the kettle or turning the tap on/off will produce a sensor event. The sensor events are shown in Figure 2(a) and include opening and closing the cupboard



(a) (b)

Figure 2 Events defined in the ontology (a) Sensor Events (b) Video Events

[2] Serg.ulster.ac.uk



(a) (b)

Figure 3 Definitions in the Ontology (a) Proximity of Objects (b) Actions

door and the fridge door, turning the tap on and off and tilting the kettle. A video event can be identified when an object has moved, for example, the change in position of the kettle when it is moved from the base to under the tap generates a video event. These video events are shown in Figure 2(b) and include the kettle events: *'under the tap'*, *'on the base'*, *'over the cup'*. There are also events to include the cupboard doors, the cup and the teabag container. Additionally, the events also consider the relationships that exist between these objects. In the ontology these relationships are defined by the proximity between the objects during the sub-activities. Figure 3(a) presents the proximity of objects defined in the ontology. For example, the kettle will be placed *'under'* the tap during the *'boil kettle'* sub-activity and the kettle will be *'over'* the cup during the *'pour water'* sub-activity. Figure 3(b) shows a list of actions defined in the ontology. These actions include *'lift'* and *'carry'* given that the user will *'lift'* the kettle or *'carry'* the kettle. Together these events can be used as cues to determine the activity being performed and the status of that activity. For example, if the kettle is moved from the base to under the tap, the tap sensor has been activated and deactivated and the kettle is returned to the base we can infer that the *'boil kettle'* sub-activity has been completed.

C. Scenario

Consider the scenario of a user making a cup of tea. The activity is performed in the kitchen and requires the use of a kettle, a cup and a teabag. The cup is stored in cupboard 1 and the teabag container is stored in cupboard 2. On completion of the activity the sensor and video events were obtained from the recorded datasets. A subset of these events is presented in Table 1.

TABLE 1 AN EXCERPT OF RECORDED SENSOR AND VIDEO EVENTS

Timestamp	Device Id(object)	Sensor Event	Video Event
12:38:18			Lift kettle
12:38:20			Kettle Under Tap
12:38:20	35125(tap)	Alarm	Interaction with Tap
12:38:29	35125(tap)	Door Closed	Interaction with Tap
12:38:32			Kettle on Base
12:38:33	35509(cupboard1)	Alarm	Open Cupboard 1
12:38:35			Retrieve Cup
12:38:36			Close Cupboard 1
12:38:36			Cup on Worktop
12:38:39	35509(cupboard1)	Door Closed	
…	…	…	…
12:39:10			Lift kettle
12:39:11			Kettle over Cup
12:39:13	35485(kettle)	Door Closed	
12:39:16			Kettle on Base

The video events for the *'boil kettle'* sub-activity are shown in Figure 4. This activity is initiated by the *'lift kettle'* action shown in Figure 4(a). No sensor events are recorded given that the kettle has not been tilted to activate the sensor. This video event is valuable in determining that an activity has begun. The video event *'kettle under tap'* is shown in Figure 4(b) and shows the importance of the proximity of the kettle with the tap in deciding what sub-activity is taking place. The tap sensor is activated at 12:38:20 and deactivated at 12:38:29. These sensor events provide evidence that the tap has been turned on and off. The video shows the user interacting with the tap; however, it is impossible to tell from the video if the user has turned the tap on or off. The video event *'kettle on base'* is shown in Figure 4(c) and indicates the sub-activity is finished.

Figure 5 presents the sequence of video events recorded during the *'get cup'* sub-activity. The sensor data on its own provides evidence that the cupboard door has been opened and closed. The *'open cupboard'* event is shown in Figure 5(a) and is supported by the sensor event recorded at 12:38:33. In Figure 5(b) we can view that the user has lifted an object out of the cupboard. Since we know that the cup is stored in this cupboard we can infer that the user has removed the cup. The sensor is deactivated at 12:38:36 which indicates the cupboard has been closed. This is represented as a video event in Figure 5(c). In addition, we can see that an object is now on the worktop.

Video events for the *'pour water'* sub-activity are shown in Figure 6. The *'lift kettle'* action is presented in Figure 6(a) which indicates either the *'boil kettle'* or *'pour water'* sub-activity is about to start. Figure 6(b) presents the proximity of the kettle *'over'* the cup and this determines that the *'pour water'* sub-activity is taking place. In addition, the kettle tilt sensor is activated at 12:39:13 which supports the video event. The *'kettle on base'* video event is shown in Figure 6(c) and this ends the *'make tea'* activity.

(a) 12:38:18 (b) 12:38:20 (c) 12:38:32

Figure 4 Video Events during the 'boil kettle' sub-activity (a) life kettle (b) kettle under tap (c) kettle on base

(a) 12:38:33 (b) 12:38:35 (c) 12:38:36

Figure 5 Video Events during the 'get cup' sub-activity (a) open cupboard door (b) retrieve object (c) close cupboard door and place object on worktop

(a) 12:39:10 (b) 12:39:11 (c) 12:39:16

Figure 6 Video Events during the 'pour water' sub-activity (a) lift kettle (b) kettle over cup (c) kettle on base

V. Conclusion

In this paper we have introduced an ontological framework to combine sensor and video data for the purpose of activity recognition in a smart home environment. This novel approach uses an ontology to describe how the sensor and video data are related to the activity, the user and the objects within the environment. The sensor and video data were compared to determine how the sensor and video events were related. The results from preliminary experiments show that video events provided additional information relating to the location or proximity of the objects. In addition, the video events can be used to enhance the sensor events and provide a greater understanding of the activity being performed. Furthermore, the video events can be used to overcome problems associated with anomalies in the sensor data such as missing data. The proposed approach has concentrated on a single activity being performed by a single user, however, this solution is limited and the work will be extended in the future to incorporate a solution for a single person performing multiple activities, multiple people performing a single activity and multiple people performing multiple activities.

References

[1] D. J. Cook and S. K. Das, "How smart are our environments? An updated look at the state of the art," *Pervasive Mob. Comput.*, vol. 3, no. 2, pp. 53–73, Mar. 2007.

[2] H. Medjahed and D. Istrate, "A pervasive multi-sensor data fusion for smart home healthcare monitoring," *2011 IEEE Int. Conf. Fuzzy Syst. FUZZIEEE 2011*, pp. 1466–1473, 2011.

[3] J. Parkka and M. Ermes, "Activity classification using realistic data from wearable sensors," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128, 2006.

[4] B. Chikhaoui, S. Wang, and H. Pigot, "A Frequent Pattern Mining Approach for ADLs Recognition in Smart Environments," *2011 IEEE Int. Conf. Adv. Inf. Netw. Appl.*, pp. 248–255, Mar. 2011.

[5] F. M. Khan, V. K. Singh, and R. Nevatia, "Simultaneous inference of activity, pose and object," *2012 IEEE Work. Appl. Comput. Vis.*, pp. 281–288, Jan. 2012.

[6] E. J. Pauwels, P. M. de Zeeuw, and P. H. N. de With, "Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 255–263, May 2012.

[7] Z. Uhríková, C. D. Nugent, and V. Hlavác, "The use of computer vision techniques to augment home based sensorised environments.," *Conf. Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, vol. 2008, pp. 2550–2553, Jan. 2008.

[8] L. Chen and C. Nugent, "Ontology-based activity recognition in intelligent pervasive environments," *Int. J. Web Inf. Syst.*, vol. 5, no. 4, pp. 410–430, 2009.

[9] A. G. J.C. SanMiguel, J.M. Martinez, "An ontology for event detection and its application in surveillance video," *2009 Sixth IEEE Int. Conf. Adv. Video Signal Based Surveill.*, 2009.

[10] L. Bai, S. Lao, G. J. F. Jones, and A. F. Smeaton, "Video Semantic Content Analysis based on Ontology," *Int. Mach. Vis. Image Process. Conf. (IMVIP 2007)*, pp. 117–124, Sep. 2007.

[11] R. Nevatia, J. Hobbs, and B. Bolles, "An Ontology for Video Event Representation," *Comput. Vis. Pattern Recognit. Work.*, pp. 1–10, 2004.

[12] J. R. Smith, R. J. Alexandre, J. Hobbs, and R. C. Bolles, "Standards VERL : An Ontology Framework for Representing and Annotating Video Events," *IEEE Multimed.*, vol. 12, no. 4, pp. 76–86, 2005.