Woznowski, P. R., King, R., Harwin, W., & Craddock, I. (2016). A Human Activity Recognition Framework for Healthcare Applications: ontology, labelling strategies, and best practice. In Proceedings of the International Conference on Internet of Things and Big Data 2016. (pp. 369-377). Rome, Italy: SciTePress. DOI: 10.5220/0005932503690377

A Human Activity Recognition Framework for Healthcare Applications: ontology, labelling strategies, and best practice

Przemyslaw R. Woznowski[1], Rachel King[2], William Harwin[2] and Ian Craddock[1]
[1]*Faculty of Engineering, University of Bristol, Woodland Road, Bristol, UK*
[2]*School of Systems Engineering, University of Reading, Whiteknights, Reading, UK*
{*p.r.woznowski, ian.craddock*}*@bristol.ac.uk,* {*rachel.king, w.s.harwin*}*@reading.ac.uk*

Keywords: Activity recognition; ontology; annotation; video.

Abstract: Human Activity Recognition (AR) is an area of great importance for health and well-being applications including Ambient Intelligent (AmI) spaces, Ambient Assisted Living (AAL) environments, and wearable healthcare systems. Such intelligent systems reason over large amounts of sensor-derived data in order to recognise users' actions. The design of AR algorithms relies on ground-truth data of sufficient quality and quantity to enable rigorous training and validation. Ground-truth is often acquired using video recordings which can produce detailed results given the appropriate labels. However, video annotation is not a trivial task and is, by definition, subjective. In addition, the sensitive nature of the recordings has to be foremost in minds of the researchers to protect the identity and privacy of participants. In this paper, a hierarchical ontology for the annotation of human activity recognition in the home is proposed. Strategies that support different levels of granularity are presented enabling consistent, and repeatable annotations for training and validating activity recognition algorithms. Best practice regarding the handling of this type of sensitive data is discussed.

1 Introduction

Healthcare needs have changed dramatically in recent times. An ageing population and the increase in chronic illnesses, such as diabetes, obesity, cardiovascular and neurological conditions, have influenced research, directing it towards Information Communication and Technology (ICT) solutions. Technology has also advanced enabling low cost sensors and sensing systems to become widely available, and rapid developments in the Internet of Things (IoT). These technologies complement research in field of Ambient Intelligent (AmI) spaces, including Activity Recognition (AR) and Ambient Assisted Living (AAL) for healthcare applications.

The design and implementation of AmI applications pose many challenges including, but not limited to: the selection of suitable hardware (sensors and gateways); system architecture and infrastructure design; software design, including the training and validation of data mining/data fusion algorithms; system testing and deployment; and the appropriate dissemination of collected knowledge. Each of these steps requires the consideration of the needs and preferences of particular stakeholders. This paper will focus on a specific aspect of the training and validation process: the acquisition of usable ground-truth data.

AmI spaces process large quantities of sensor-derived data and require robust and accurate data mining strategies in order to recognise activities of interest to monitor health, well-being, or other personal benefits such as fitness level. To validate these strategies, the output from these algorithms is usually compared with *ground-truth* or *benchmark* data i.e labels or semantics that describe what the data actually represents in the same form as the algorithm's output. Ground-truth data is also often used to train algorithms when using supervised training strategies.

Acquiring ground-truth data has three main stages: 1) collecting suitable information upon which to base the ground-truth data; 2) determining the appropriate labels to be applied to the raw data; and 3) annotating or labelling the data. The first step is challenging and requires a suitable strategy to ensure there is enough known about the activities being conducted, to enable precision and detailed annotation. Often studies are focused on specific aspects of activity, such as gait and ambulation, transitions, postures, or specific diseases. As such, for the second stage, labels are often research specific and often not globally applicable. The final step is to annotate or transform this data to carry the same labels/semantics as

output of the algorithms. However, accurate and consistent annotation of video to provide labels for the design and validation of algorithms for activity recognition is expensive and time consuming. The task is not trivial and if performed by multiple annotators can result in vastly different and inconsistent results. Researchers want to ensure consistency and high-quality of annotations, while annotators need clear and concise instructions on how to perform this task. Multiple strategies can be applied, yet there is no definitive solution to this problem.

In this paper, background on methods to collect and annotate ground-truth data are described. A method for collecting data for annotation is provided based on a script including prescriptive and informal activities (stage 1). An ontology for the annotation of activities in the home environment for healthcare applications is then described with an emphasis, not only on the activity a person is performing, but the context of the activity (stage 2). The translation of the ontology into a framework for annotating human activities in the home is described including the development of strategies to provide consistent and repeatable annotation (stage 3). This paper concludes after a discussion on the efficacy of the annotation framework presented in this paper for the acquisition of high quality ground-truth data.

2 Background

Ground-truth data acquisition involves three stages: collecting data upon which to base the ground-truth, deciding what labels are appropriate to describe the data, and applying these labels annotating the data to obtain the ground-truth. These stages are not necessarily applied linearly, but are often developed as a result of an iterative process to determine the best data collection methods, most appropriate labels to use, and a suitable way to annotate the data to produce consistent and informative ground truth.

Self report or diaries are imperfect as they rely purely on participant's compliance and their subjective perception and may not be accurate enough to use for developing activity recognition algorithms and validation. It is also unrealistic to expect detailed activity diaries with the exact timings of the activities. Allen et al. (Allen et al., 2006) collected unsupervised activity data in the home using a computer set up to take participants through a routine. Input from the user was in the form of a button press from which the data was annotated. Bao and Intille (Bao and Intille, 2004) used a semi-supervised strategy to collect annotated data for AR based on scripts. Unobserved participants labelled the beginning and end time of the activities. Kasteren et al. (van Kasteren et al., 2008) asked participants to wear a Bluetooth headset that used speech recognition to label the ground-truth data. This method is inexpensive, however there is a limit to the amount of detail that can be captured.

Another strategy is for the researcher to record the activity and context during data collection (Pärkkä et al., 2006; Maurer et al., 2006). Pärkkä et al. (Pärkkä et al., 2006) adopted a semi-supervised approach to collecting data for AR classification based on realistic activities. A single researcher followed the participant during the experiment and used an annotation app to record the activities. Even with this approach it was noted that there were annotation inaccuracies that might explain inaccuracies in classification.

Method using video recordings provide an objective reflection of participant's activities enabling a far more accurate and detailed activity ground-truth data, however these require additional attention in the form of *ontology*. Data can also be collected in an unsupervised environment, encouraging natural behaviour; however it can also be perceived as intrusive and will only capture actions with no room for participant interpretation. Atallah et al. (Atallah et al., 2009) used video to annotate activities during laboratory experiments to train AR classifiers. Tsipouras et al. (Tsipouras et al., 2012) used video recordings to annotate data to develop a system for the automatic assessment of levodopa-induced dyskinesia in Parkinson's disease.

Video annotation can be costly and time consuming. Active learning is a technique that can reduce the amount of annotated data needed for training a classifier. In this approach, classifiers are initially trained on a small set of data. When the classifier is applied to unlabelled data, the user is asked to label only data that, for example, is classified with low confidence or those with disagreement between classes (Stikic et al., 2008). Hoque and Stankovic (Hoque and Stankovic, 2012) employed a clustering technique to group smart home environmental data into activities and the user labelled the clusters. Another application for active learning, is to update classifiers or personalise them (Longstaff et al., 2010). While attractive, these methods rely on a collaborative effort on the part of the user.

An alternative to attempting to reduce the amount of annotated data required could be to increase the number of annotators to label the data. By crowdsourcing, video annotation tasks can be opened up to the online community. Vondrick et al. (Vondrick et al., 2013) developed an open platform for crowd-

sourcing video labeling, VATIC (Video Annotation Tool from Irvine, California). This platform enables the labelling and tracking of objects of interest within the scene and associate attributes to the object using bounding boxes.

There are a number of available software tools which are suitable for video annotation, such as ANVIL[1] video annotation tool (Kipp, 2012) or ELAN[2] (Brugman and Russel, 2004) developed by the Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands.

The labels used to annotate data is another annotation consideration. Labels are often application specific, e.g. Pärkkä et al. (Pärkkä et al., 2006) used a hierarchical list of labels, aimed at capturing the context of the activities, where as, (Tsipouras et al., 2012) focused purely on a specific disease and the associated symptoms. Logan et al. (Logan et al., 2007) presented a detailed activity ontology for the home using a custom tool that enabled annotators to label foreground and background activities for when the participant's attention is focused on another activity, addressing the fact that humans naturally multitask. Roggen et al. (Roggen et al., 2010) used a four "track" annotation scheme for annotating human activities based on video data including tracks for locomotion, left and right hand activities (with an additional attribute that indicates the object they are using), and the high level activity.

In the SPHERE project data mining algorithms are initially trained and validated against recordings from a head-mounted video camera worn by participants. The data originates from the three different sensing modalities (depth-camera, wearable accelerometer and environmental sensors) deployed in a real house, which constitutes the testbed (Woznowski et al., 2015). The same ontology, presented in this paper, underpins system-generated data and the controlled vocabulary used in video annotation. In this paper we propose two annotation strategies – based on a controlled vocabulary derived from SPHERE ontology of activities of daily living (ADL) – which provide simple frameworks for video annotation of ground-truth video data.

3 Data Collection

The script-based data collection took place in the SPHERE testbed – a typical two storey, two bedroom house with two bedrooms on the upper flow and two rooms on the ground floor. The platform deployed in this testbed is based on three sensing technologies: a Body Sensor Network made up of ultra low-power wearable sensors; a Video Sensor Network focusing on recognition of activities through video analysis of home inhabitants; and an Environment Sensor Network made up of hardware sensing the home ambience. More detailed description of the sensing platform, together with system architecture can be found in (Woznowski et al., 2015). The presented taxonomy applies more widely and is not limited to the testbed setup. Scripted data has been collected for 11 participants.

3.1 Script

The script for data collection has been designed based on several objectives and involves one participant at a time. Firstly, it aims at exercising all types of sensors deployed in the house – from electricity monitoring of individual appliances to RGB-D cameras in the hallway. Secondly, the participant is asked to visit all locations in the house which allows us to observe sensor activations, their coverage, temporal relationships, etc. Finally, the script consists of a representative set of activities and posture transitions which provides data mining algorithms with training data.

3.2 Protocol

Prior to data collection each participant was issued with a consent form and a copy of the script. Copy of the script was provided in order to ensure participants were fully aware of the tasks they would be asked to do and also to resolve any questions or ambiguities before the start of an experiment. Each participant was asked to execute the script twice with a short break in between the two runs. This break was not only to allow the participant to rest/ask further questions but also to reassure that all the systems were working correctly and no data was lost.

Participants were asked to wear three devices. A head-mounted, wide-angle, 4K resolution camera (Panasonic HX-A500E-K[3]). A Motorola TLKR T80 walkie talkie consumer radio[4] with hands-free VOX setup over an earpiece headset. The third device, worn on the dominant hand, was the wearable device incorporating two three-axial accelerometers (Fafoutis et al., 2016). The full setup in Fig. 1.

[1] http://www.anvil-software.org/

[2] https://tla.mpi.nl/tools/tla-tools/elan/

[3] http://www.panasonic.com/uk/consumer/cameras-camcorders/camcorders/active-hd-camcorders/hx-a500e.html

[4] http://www.motorolasolutions.com/en_xu/products/consumer-two-way-radios/t80e.html

Figure 1: Data collection participant setup: head mounted camera (head and left arm), wearable accelerometer (dominant, right arm) and walkie talkie (belt + earpiece).

Experiments involved two people: participant and experimenter. The experimenter remained on site, out of sensors' view, and read instructions over the walkie talkie radio to the participant. Instructions were precise enough for the participant to understand what was required from them, however allowed for some degree of variability since participants were only instructed on what to do but not precisely on how to do it i.e. prepare yourself a cup of tea (without explicit instructions on how to do it, where cups and teabags are, etc.). Each experiment started and finished with jumping in front of the mirror activity in order to provide alignment point between and within the three sensing modalities and the head-mounted camera. Although the sensing platform is NTP-synchronised, this was seen as a good practise and a backup strategy if any of the sensing modalities was out of sync. There are more sophisticated techniques for time synchronisation in AAL spaces e.g. (Cippitelli et al., 2015) used a series of LED lights, however having a unique stamp of participant's jump (clear accelerometer and video signatures) was seen satisfactory.

4 Ontology for Activities of Daily Living

The SPHERE ontology for activities of daily living lists and categorises activities occurring in the home environment. No model is fully complete, hence this ontology is expected to evolve over time. The initial dictionary of ADLs was compiled during project meetings between researchers and clinicians involved on the project. The result of this collaborative effort has been extended for completeness mainly with activities found in the Compendium of Physical Activities[5]. The final stage involved merging it with BoxLab's Activity Labels[6] and thus extending their model. Compliance with existing models ensures interoperability and applicability of collected datasets beyond the project.

In-line with BoxLab's model, ACTIVITY has the following properties: *activity, physical state* (posture/ambulation in BoxLab's model), *social context* and *room*. Furthermore, as represented by Fig. 2, the ontology has been extended with additional properties, namely *physiological context, (at) time, (involve) object, (involvedAgent) person, ID,* and *sub-activity* with a self referencing relation. Thus, each activity occurs at a certain point in time, is identifiable by some unique ID. Moreover it can involve physical objects, people and might be made up of a number of *sub-activities*. Physiological context is also of importance due to application of this ontology for healthcare and monitoring people living independently in their own homes. Overall, AAL technologies have to ensure user's safety and monitoring physiological signs is one possible method. The growing wearable sensor market and smart-phone apps reflects this interest, as more products offer heart rate monitoring features. Fig. 2 depicts the structure of the proposed ontology, the latest version of which, in the OBO[7] format, is available from http://data.bris.ac.uk (DOI: 10.5523/bris.1234ym4ulx3r11i2z5b13g93n7).

4.1 Activity Hierarchy

In the proposed ontology, ADL are organised hierarchically. *Activity* has 20 sub-classes out of which 15

[5] https://sites.google.com/site/compendiumofphysicalactivities/Activity-Categories/home-activity

[6] http://boxlab.wikispaces.com/Activity+Labels

[7] http://oboedit.org/

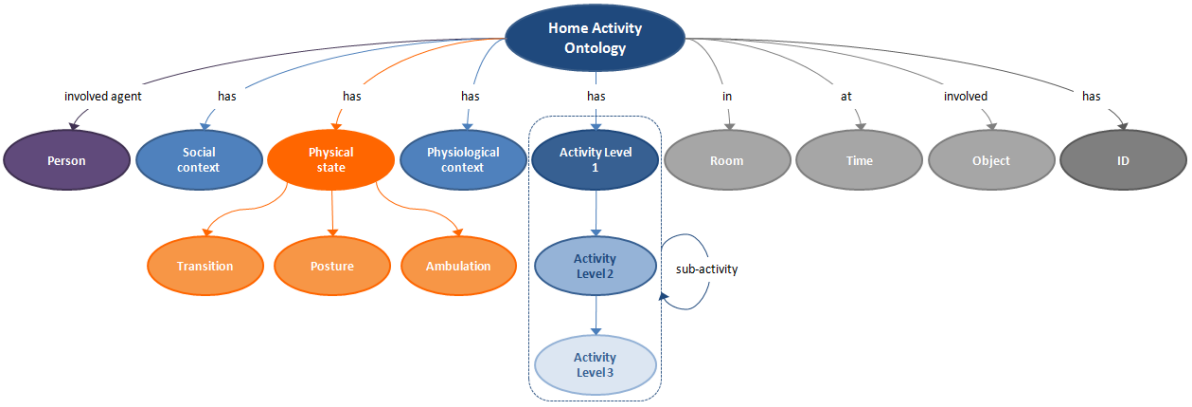

Figure 2: High-level view of proposed ADL ontology.

are present (albeit some names may differ slightly) in the BoxLab ontology. Five categories were added to the original ontology to capture additional ADL and to reflect aspects related to health. These include *Atomic home activities, Health condition, Social interaction, Working*, and *Miscellaneous* (shortened to "Misc"). These categories are further described in the reminder of this section. Detailed description of the 15 sub-classes which are present in both ontologies, can be found at BoxLab website[6].

Activities often involve interactions with one or more object. These interactions/activities have been reflected in the ontology in the *Atomic home activities* class and its subclasses. These capture the low-level activities or simple actions which form the basic building blocks for other activities (evidencing sub-activities). Video annotators can use these labels to identify short actions for use in AR algorithms. With the increasing sophistication of wearable technology and sensing, research into identifying these types of activities will become more prominent. In the current version of the ontology, *Atomic home activities* has the following subclasses: *door interaction, window interaction, object interaction, tap interaction, cupboard interaction, draw interaction,* and *electrical appliance interaction*, each with a further level of subclasses (omitted for brevity).

The *Health condition* class is essential to describe activities and behaviours in the context of a persons health. By training algorithms for AmI or AAL applications and associating activities (or lack of) with a persons well-being, early warning signs that someone is unwell or in need of assistance or medical treatment could be predicted. This is especially important given the health challenges and the inherent socioeconomic impact facing society today. This category currently includes: *coughing, fall, fever/infection, shaking* and *sweating*.

Social interaction is comprised of: *receive visitors, social media, talking* (with subclasses), and *video calling* activities. Finally, *Working* is further divided into *intellectual* and *physical work*. Every subclass of *Activity* has a *misc* member to enable annotation of knowledge which the ontology does not explicitly capture. *Misc* is for activity labels which do not fit into any of the existing classes and currently has *smoking tobacco*.

4.2 Ambulation, Posture & Transition

The structure of *Physical state* directly reflects BoxLab's ontology *posture/ambulation* category, yet has been extended with additional entities. It has three subclasses, namely *ambulation, posture*, and *transition*. Since activities do not always describe a person's posture (with some exceptions e.g. *running* where the posture is inherent in the activity), it is important to capture this information separately.

4.3 Contextual Information

Room/location, social context and physiological context make up contextual information. For any activity it is beneficial to know the context in which it occurred. Some activities are associated with a particular location (bathing activity in bathroom location) where some can occur anywhere inside or outside the home environment. From the healthcare point of view, it is also important to capture social context as people's behaviour can be affected by presence of other individuals. Finally, physiological context such as blood pressure or glucose level have influence on our well-being and behaviour. Information captured without context is of limited value as it does not fully reflect reality.

In addition, activities consist of *(involved) object* and *(involvedAgent) person* properties, which capture object(s) and people involved in a particular activity. Since some activities can be made up of shorter (in duration) activities, *sub-activity* relation was introduced. For completeness, (has) ID attribute was introduced to allow to differentiate between activities. All these properties and relations are captured in Fig. 2

5 Proposed Annotation Strategies

Given the proposed ontology, a structure was developed to map the ontology into a form that could be intuitively used by a human annotator. Video and audio annotation software, such as Anvil[1] and ELAN[2] provide a tier-based solution that lends itself well to the hierarchical activity structure of the proposed ontology. These tiers can also be used to provide the context of the activities, e.g. the time, location, and social context. The work presented in this paper uses ELAN for annotating the script video data.

Three tiers were assigned to describe activities in terms of detail from high level activities to low level activities called *ADL: Tier1* e.g. cleaning, *ADL: Tier2* e.g. hoovering, and *ADL: Tier3* e.g. object interactions. For each tier a controlled vocabulary (CV) was defined using ontology labels. Further tiers were assigned with CVs based on the ontology, to annotate physical state (*Posture/ambulation*), room (*Location*), and social context (*Social Interaction*).

Two annotation strategies have been explored. The first annotation method directs the annotator to label the activity the participant is focused on at the time. Some researchers report that "Humans (...) don't do lots of things simultaneously. Instead, we switch our attention from task to task extremely quickly"(Hamilton, 2008). With this view in mind, this strategy assumes no concurrency in annotated activities. In this strategy, the three activity tiers, *ADL: Tier 1 - 3*, are used. Such an approach may result in annotating a large number of short and potentially unfinished activities. The second annotation method allocates two sets of activity tiers enabling the capture of concurrent activities.

To demonstrate the impact of these different strategies consider a situation in which a person is cooking (kitchen) and watching TV (living room) at the same time – as illustrated in Fig. 3. Using the first strategy i.e. annotating based on the user's attention, *cooking* would have a total duration of 1h with the same duration of 1h for *watching tv*. Using the second strategy, i.e. labelling concurrent activities, *cooking* activity could be annotated as lasting 1.5h and *watching tv* as 2h long activity.

Figure 3: Example of concurrent activities.

In the strategy 1, the user's attention shifts between the two activities with a physical change in location. Without further information or context these tasks are considered independent. For many applications, this may be suitable and is often the annotation strategy used to annotate much of the research in AR. However, if additional information is present that indicates the bouts of cooking are related, i.e. the same meal being prepared, annotating both activities concurrently maybe more suitable. This would be especially useful for modelling higher level behaviour routines and trends.

6 Video Annotation

Video annotation task is not simple due to multiple factors such as complexity of human actions, multitasking/switching between tasks, complex interaction with surrounding environment, angle/orientation of the camera, etc. Moreover, human annotators use their own subjective judgment and can interpret the same actions/activities in different ways. Video (with audio) is a very rich source of information and hence some clues/pieces of information are not always apparent. Furthermore, people execute the same activities in different ways.

6.1 Recruitment

Annotators for the video annotation task were recruited via an internal job advertisement. A half-day workshop was organised to train and select undergraduate students from the University to perform the annotation. During the workshop students were given an overview of the SPHERE project, the ADL ontology, and the annotation task at hand. The two different annotation strategies described in this paper, and the relationship of the controlled vocabulary to the ontology, were explained with examples of each. The workshop concluded with two annotation tasks to be completed using the ELAN video and audio annotation tool where attendees had the opportunity to apply these strategies. For each task, students were provided with a 10min long video, annotation templates with preloaded controlled vocabularies, and 40 minutes to annotate as much as possible of the video.

Prior to attempting these tasks they were given an opportunity to practise simple annotations and ask questions. At the end of the workshop, attendees submitted their video annotations for both tasks which were visually evaluated for attention to detail, application of appropriate annotation labels, and consistency of annotations.

6.2 Annotation of Scripted Experiments

The annotation of scripted experiments to provide ground-truth serves several purposes. One of which being to provide researchers addressing data mining/activity recognition problem with a base-line to compare the results from their algorithm's to. Validation of AR algorithms is required to demonstrate how well they work and to calculate accuracy metrics such as precision and recall. Providing annotators the ADL ontology via controlled vocabularies opens up an opportunity to test the integrity of the ontology. Annotators were asked to report on any problems faced including missing labels which then could be added to the ontology. Since every video is annotated by two different annotators, there is scope for comparing the two annotations to highlight conflicts (and remove bias) and/or merge the two to provide a more complete picture of the real world.

Annotation of videos, performed by students using two different strategies, allowed us to compare the usefulness of each technique and identify their shortcomings. In the provided workshop, annotators were asked to make a note of any difficulties experienced while performing the task, such as missing activity labels, inability to capture some information, etc. Qualitative data collected in this way has been analysed and some suggestions were adopted.

The ontology of ADL was found to be fairly complete with few amendments made based on annotators' feedback. Most of the annotators' comments contributed to the growth of classes in the *atomic home activities* category. Moreover, the ongoing video annotation task provides continuous feedback and annotators are reporting (as they get to know the dictionary better) to spend less time on every subsequent annotation.

7 Discussion

The model described in this paper and visualised in Fig. 2 captures all the necessary information required to describe some activity. It provides a hierarchical structure of activities, together with all the relevant and important contextual information. Hence, when the AR algorithm recognises a particular activity, it can produce an output in the format conforming to the proposed ontology. Therefore, the inferred activity would happen in a particular time interval, in a particular room, involving certain object(s) and people, in a given social context with the subject being in a particular physical state and physiological context, and could be composed of a number of sub-activities – all of which can be captured via use of available properties.

Both annotation strategies presented in this paper reflect the hierarchical structure of the activity classes allowing for different levels of precision when annotating videos. For example, with limited human or financial resources, or based on the application of the AR algorithms, one may wish to annotate videos only at *Tier 1* to know the classes of the activities occurring. Mirroring the ontology structure allows for direct comparison between annotation labels and activities recognised by AR algorithms. The choice of the annotation strategy can be based on the type of the algorithm which is to be validated and its underlying assumptions. The first annotation method is well suited for e.g. (Filippaki et al., 2011) which uses the concept of activity resources (tangible and intangible). In their work, "two complex activities are in conflict, if their time-intervals overlap and they use common resources' (Filippaki et al., 2011). Since user attention is an intangible resource, *cooking* and *watching tv* activities would be identified as conflicting cases and the conflict resolution strategy would resolve it resulting in activities which do not overlap in time. The second annotation method is more annotator-friendly, in the sense that fewer annotations are required, and is well suited for AR algorithms which work with the assumption that humans can multitask and/or look for unique sensor signatures which indicate start and finish of the activity (and hence assume that the activity continued from start to finish time) e.g. (Woznowski, 2013). Holding the assumption about people's ability to multitask, a limit has to be set in order not to have three or more tracks to capture concurrent activities (as this would complicate the task and result in far too many annotation tracks). However, if the subject in the annotated video switches between three or more tasks, the first annotation strategy (task-switching) can be applied to the two activity tracks of the second strategy.

8 Future Work

The work presented in this paper can be taken further in multiple directions. Firstly, since every video

is annotated twice by randomly selected annotators, it is possible to carry out intra-annotation correlation analysis. Such analysis would take into consideration agreement between individual labels in a given tier (Activity, Physical state, etc.) as well as start and finish time (and duration) of each annotation. Labels with low agreement score, could then be revisited and resolved.

One way to improve, or rather make activity labels more intuitive, would be to provide more user-friendly activity labels by considering their synonyms. This information can either be acquired from dictionaries or by running a user study with annotators. Currently the naming of activity categories and actual activities is more scientific/encyclopedic than commonly used language. More colloquial naming would enable annotators to navigate quicker and more freely in the ontology hierarchy.

Since only scripted experiments' data has been considered in the presented study, analysis of free-living recordings is the next logical step. This would stress-test the ADL ontology for missing labels. Moreover, data collection subjects could be asked to annotate their own video. Such approach would remove ambiguities, yet would require participants to undertake some training in order to familiarise themselves with the video annotation software and the controlled vocabulary.

9 Conclusion

The aim of this work and presented video annotation strategies was to design two approaches which provide good and easy to follow annotation frameworks. The purpose of annotating videos captured during scripted experiments is to provide ground-truth data against which activity recognition algorithms can be trained and validated. Based on the hypothesis undermining AR algorithms (whether multitasking is considered or not), one can chose an annotation strategy which supports that assumption. Furthermore, algorithms can work at different levels i.e. from providing a very fine level (at the *atomic home activities* level) of detail to only producing high-level activity labels (*cleaning*). The two presented annotation approaches work at different level of granularity and hence e.g. only Tier 1 activities, which correspond to the high-level activity categories, might be annotated if required.

The two annotation approaches were tested on a group of students with different backgrounds. The analysis of feedback from the video annotation workshop reassured us that the selected annotation strategies are easy to comprehend and work with. Moreover, it highlighted areas which needed improving i.e. missing activity labels. The approach to map the controlled vocabulary to the ontology of ADL was found to be appropriate as it enables direct comparison of output of AR algorithms against annotated activity labels. Moreover, adopted approaches provide other contextual information such as location, social context, physiological context, physical state (posture/ambulation/transition) and to some extent interaction with objects (via ADL Tier 3). All these pieces of information captured during the annotation process are important in the context of validating AR algorithms. Moreover, annotations resulting from either of the presented approaches can validate other algorithms such as those sitting behind various services that intelligent spaces can provide e.g. real-time location service (RTLS). In order to accelerate research in AmI and AAL spaces, sensors' data together with ground-truth data from the SPHERE project will be made available online to the research community. Since design, implementation and collection of intelligent spaces' data (with ground-truth data for validation) is time consuming and expensive, we share the view that results of efforts in this space should be made publicly available.

Ethical Approval

Ethical approval for this study has been granted from the Faculty of Engineering Research Ethics Committee (FREC), University of Bristol (ref: 14841).

Acknowledgements

We thank all the volunteers for participating in our trials and video annotators for their hard work. Many thanks to the SPHERE team for implementing a sensor-rich testbed in which all the data was collected. This work was performed under the SPHERE IRC, funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1.

REFERENCES

Allen, F. R., Ambikairajah, E., Lovell, N. H., and Celler, B. G. (2006). Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. 27(10):935–51.

Atallah, L., Lo, B., Ali, R., King, R., and Yang, G.-Z. (2009). Real-time activity classification using ambient and wearable sensors. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 13(6):1031–9.

Bao, L. and Intille, S. S. (2004). Activity Recognition from User-Annotated Acceleration Data. *Proceedings of PERVASIVE 2004*, pages 1–17.

Brugman, H. and Russel, A. (2004). Annotating multimedia / multi-modal resources with ELAN. *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC 2004)*, pages 2065–2068.

Cippitelli, E., Gasparrini, S., Gambi, E., Spinsante, S., Wahslen, J., Orhan, I., and Lindh, T. (2015). Time synchronization and data fusion for rgb-depth cameras and wearable inertial sensors in aal applications. In *IEEE ICC Workshop on ICT-enabled services and technologies for eHealth and AAL*.

Fafoutis, X., Tsimbalo, E., Mellios, E., Hilton, G., Piechocki, R., and Craddock, I. (2016). A residential maintenance-free long-term activity monitoring system for healthcare applications. *EURASIP Journal on Wireless Communications and Networking*, 2016(31).

Filippaki, C., Antoniou, G., and Tsamardinos, I. (2011). Using constraint optimization for conflict resolution and detail control in activity recognition. In *Ambient Intelligence*, pages 51–60. Springer.

Hamilton, J. (2008). Think you're multitasking? think again. *Morning Edition, National Public Radio (2 October 2008)*.

Hoque, E. and Stankovic, J. (2012). AALO: Activity recognition in smart homes using Active Learning in the presence of Overlapped activities. *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare*, pages 139–146.

Kipp, M. (2012). Annotation Facilities for the Reliable Analysis of Human Motion. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, pages 4103–4107.

Logan, B., Healey, J., Philipose, M., Tapia, E. M., and Intille, S. (2007). A Long-term Evaluation of Sensing Modalities for Activity Recognition. In *Proceedings of the 9th International Conference on Ubiquitous Computing*, UbiComp '07, pages 483–500, Berlin, Heidelberg. Springer-Verlag.

Longstaff, B., Reddy, S., and Estrin, D. (2010). Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In *Proceedings of the 4th International ICST Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7.

Maurer, U., Smailagic, A., Siewiorek, D., and Deisher, M. (2006). Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions. In *International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pages 113–116. IEEE.

Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., and Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*.

Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Forster, K., Troster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., Doppler, J., Holzmann, C., Kurz, M., Holl, G., Chavarriaga, R., Sagha, H., Bayati, H., Creatura, M., and Millan, J. d. R. (2010). Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh International Conference on Networked Sensing Systems (INSS)*, pages 233–240. IEEE.

Stikic, M., Van Laerhoven, K., and Schiele, B. (2008). Exploring semi-supervised and active learning for activity recognition. *2008 12th IEEE International Symposium on Wearable Computers*, pages 81–88.

Tsipouras, M. G., Tzallas, A. T., Rigas, G., Tsouli, S., Fotiadis, D. I., and Konitsiotis, S. (2012). An automated methodology for levodopa-induced dyskinesia: assessment based on gyroscope and accelerometer signals. *Artificial intelligence in medicine*, 55(2):127–35.

van Kasteren, T., Noulas, A., Englebienne, G., and Kröse, B. (2008). Accurate activity recognition in a home setting. In *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, New York, New York, USA. ACM Press.

Vondrick, C., Patterson, D., and Ramanan, D. (2013). Efficiently Scaling up Crowdsourced Video Annotation. *International Journal of Computer Vision*, 101(1):184–204.

Woznowski, P. (2013). *Rule-based semantic sensing platform for activity monitoring*. PhD thesis, Cardiff University.

Woznowski, P., Fafoutis, X., Song, T., Hannuna, S., Camplani, M., Tao, L., Paiement, A., Mellios, E., Haghighi, M., Zhu, N., et al. (2015). A multi-modal sensor infrastructure for healthcare in a residential environment. In *IEEE ICC Workshop on ICT-enabled services and technologies for eHealth and AAL*.