DANTE: A Video Based Annotation Tool for Smart Environments

Federico Cruciani[1], Mark P. Donnelly[2], Chris D. Nugent[2], Guido Parente[2], Cristiano Paggetti[1], and William Burns[2]

[1] I+ s.r.l., 50144 Florence, Italy
f.cruciani, c.paggetti@i-piu.it
http://www.i-piu.it/
[2] University of Ulster, Faculty of Computing and Engeneering, Shore Road, Newtonabbey, Co. Antrim, BT370QB, Northern Ireland

Abstract. This paper presents a novel system, which uses a set of video cameras to support the manual annotation of object interaction and person to object interaction within smart environments. The DANTE (Dynamic ANnotation system for smart Environments) system uses two stereo-based cameras to monitor and track objects, which are tagged with inexpensive paper based fiducial markers. Offline, the DANTE annotation module allows users to navigate, frame-by-frame, through recorded sessions and to annotate interesting and relevant events. The generation of annotated datasets has significant utility within the domain of smart environments to support the development of data driven context-aware systems which require large training datasets. The current paper presents the rationale for this work, provides a technical description of the system and highlights its potential opportunities. Early experimental work, investigating a comparison with accelerometer based sensors is also briefly introduced.

Keywords: Data Acquisition, Multi sensor systems, Video Recording, Optical Tracking.

1 Introduction

Worldwide, changing demographics in relation to increased life expectancy and falling birth rates are leading to an ageing society. Coupled closely with these changes are the documented rises in chronic illnesses such as cardiovascular disease, cancer and dementia [1] for which, age is the single biggest risk factor [1]. Studies have shown that persons with chronic illnesses often prefer to remain living within their own home for as long as is possible [1]. Facilitating this can assist to alleviate some of the strain faced by health services, however, a challenge is to provide equal levels of health support from within the home environment. An emerging area of interest lies in the provision of home-based care through the application of assistive technologies and integrated intelligent algorithms, to monitor and support independent living. In this paper, we describe a novel

G. Parr and P. Morrow (Eds.): S-Cube 2010, LNICST 57, pp. 179–188, 2011.

system for recording and interpreting person-object interactions within a home environment with the aims of providing labelled datasets to promote the development of data driven context-aware intelligent algorithms. The remainder of the paper is structured as follows. In Section 2, we introduce related work in this domain. Section 3 presents the DANTE system and describes its operation. Section 4 discusses future opportunities for the system and Section 5 formulates our conclusions by presenting some experimental results.

2 Related Works

Integration of technological solutions and services within the home environment offer the potential to support independent living through automated monitoring and support of activities of daily living (ADLs) [2]. Within this domain a challenge lies in "learning" how people interact with household objects to complete their ADLs. Once the learning process is complete, intelligent algorithms can then be used to detect deviations from normal patterns and when required offer some support. This challenge is further complicated by the fact that different people perform ADLs in different ways hence leading to the requirement to tailor algorithms so that personalised support can be offered [3]. Indeed, only by referring to a large source of data can we begin to fully understand ADL patterns and develop intelligent algorithm which will be robust when deployed *in situ*. This has been the view of Intille *et al.* [4] and other researchers [5] [6] who have stated the need for shared sensor rich data repositories to accelerate the development of real world context-aware systems to support and monitor ADLs. To date, several publically available sensor datasets have been published [7][8], however, a challenge remains in providing sufficiently annotated datasets [5] [6]. In part, this issue is attributed to the complexity and error-prone process of annotating data collected from sensorised environments[6], which typically involving hours of manual hand labelling [9] or requesting participants to manually keep detailed ADL logs [10].

In 2009, Cook *et al.* [3] presented a comprehensive overview, outling four methods for labelling / annotating sensor data involving: annotation of raw sensor data; annotation of raw sensor data coupled with resident diaries; using a 3D environmental visualisation tool; and the latter coupled with resident diaries. They employed a naive bayes classifier to evaluate the labeled data and found that the visulisation tool, coupled with resident diaries provided the most rapid platform for annotating data and was the most accurate (>73%) for identifying target ADLs. They acknowledged, however, the invasiveness of this approach in terms of resident burden in addition to the requirement for custom 3D models to be generated for any new environments.

In a similar study, Coyle *et al.* [6] used video cameras coupled with participant diaries to annotate data within an office environment. They attempted to streamline the annotation process by placing pressure sensors within the environment to infer important periods (e.g. person entering a room) where the captured video data contained activity related information.

Other researchers considered semi-automatic approaches to the annotation process. van Kasteren *et al.* [11] asked participants to state activities verbally. These were recorded, via a bluetooth headset, and voice recognition software was employed to interpret key phrases to indicate the beginning and end of an activity. They reported that voice recognition was relatively accurate, however, participants became tired as they had to verbally discuss each activity. They concluded by stating that by focusing research upon the development of software tools for visualising and examining sensor data could lead to a standard data structure being adopted. Indeed, Hong *et al.* [12] presented the OpenHome concept as a potential method for structuring data within smart environments, however, did not address the issue of data annotation within their work.

3 System Overview

The rationale for this work stems from the aforementioned literature, which has highlighted the need for large annotated datasets, to allow researchers to more efficiently and accurately capture ADL behaviours. The DANTE (Dynamic ANnotation Tool for smart Environments) system (see `http://www.orthokey.com/`) provides a suite of software tools for recording and annotating ADLs.

Hardware

The system hardware consists of two stereo MicronTracker cameras[14]. The cameras (fig.1) are lightweight and portable and can be positioned anywhere within an environment to provide the maximal field of view for the ADL being examined. Each camera can identify objects (cup, door, appliances, people) which are 'tagged' with custom fiduciary markers (fig.2). Once detected the system can report on the location and orientation of each detected object. Markers are fully passive and consist of a pattern (or set) of high-contrast regions printed on paper. The number of unique markers which can be designed is practically unlimited as they can differ in both size and pattern. The DANTE system samples at a frame-rate between 1-10 Hz, however, in its current configuration, has been optimised to concurrently track up to 30 objects. A tolerance parameter is used to set sensitivity to object movement. The camera range has been optimised to between 0.5 to 5 metres, depending on the size and complexity of the markers being tracked. Using an unique reference point to define an environment-based coordinate system, two cameras can track the same object in a global coordinate space.

Fig. 1. MicronTracker Camera

Fig. 2. Fiducial markers which are used to tag objects

Recording Module

Markers are 'learned by' the camera system and labelled offline. During recording, each camera captures a number of scenes, where a scene describes a video frames during which tagged objects have appeared/disappeared or moved, more than the configurable tolerance parameter, within the field of view. The DANTE system automatically activates/suspends recordings, intelligently ignoring frames where no change in the environment occurs or where only non-tagged objects are moved, thus significantly reducing the amount of stored data and recorded video duration. Fig.3 presents a recorded scene and accompanying data which describes the scene. Although video frames are stored to provide assistance during the manual annotation process, the recording module does not use the video frames to track objects. Data stored for each scene include: the object(s) ID(s) being tracked; the position of the object with respect to the global coordinate system; the orientation of the object(s) and the timestamp.

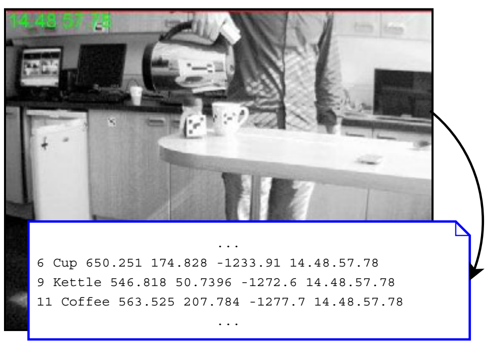

Fig. 3. A frame and the 3D tracking data indicating object IDs, labels and global coordinates with associated timestamps

The user interface (UI) of the recording software module shows the last identifed scene and the corresponding coordinate system (top and frontal) of the 3D representation based on coordinates of objects in the scene. Objects are colour coded and represented in form of dots in the 3D representation.

Annotation Software Module

The annotation module supports reviewing and labelling (annotating) of recording scenes. The annotation modules (Fig.4) presents a simple and user-friendly interface which allows users to load previously recorded sessions and play/pause/rewind or fastforward through scenes. The real-time interval between two consecutive scenes is variable depending on the presence of object movement within the environment, therefore it is possible to skip seconds or even minutes between consecutive frames.

Fig.4 indicates the range of services offered to users including top/frontal 3D visualisation of tagged objects (upper left), video frames from both cameras for the present scene (upper right) and a list of personalisable objects and associated actions to support annotation (centre). The list of objects and actions are easily configurable offline through a plain text file that enumerates and classifies the set of actions into an activity-subactivity scheme. For each object, two base actions are defined to indicate the active or inactive state of an action.

It is possible to replay recorded frames at a frame-rate of between 1-10 Hz to allow the user to efficiently progress through a session. At any point, the user can pause the playback and fastforward or rewind, frame-by-frame, until they locate the desired frame. At this point, it is possible to set/unset the object action by selecting the corresponding action's label on the UI. The visualisation of these states is shown in Fig.4 under 'Active Object State'. Alternatively, the system provides a free-text form to annotate an activity not represented. Following annotation, it is possible to replay the 'annotated session' and display the video frames in addition to highlighting those currently active actions, highlighted in the centre of the UI. Moreover, the system allows users to import 'filtered'

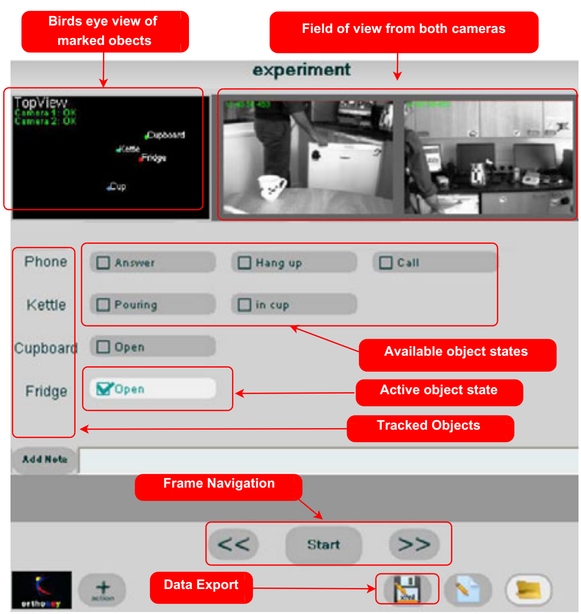

Fig. 4. Annotation Module User Interface

sessions: that is, scenes that include only a particular object or subset objects, aiding further to streamline the annotation process.

The output file of an annotated session consists of a list of events that can be the beginning or the end of an action (active or inactive) in addition to an optional free-text annotation. The annotation is saved in a plain-text file. The first line indicates which configuration file has been used to define the set of actions for the annotation. This is followe by a list of annotated events, described by a timestamp, activity ID (e.g. Phone), action (e.g. Pick-up), and boolean value to indicate object activity. An excerpt of a typical output file is shown below.

```
* myActionSet.txt
16.18.10.593 3 0 1
16.18.14.796 3 0 0
16.18.46.93 -1 -1 1 Some text
```

4 DANTE Opportunities

The DANTE system offers a number of additional opportunites with respect to improving the annotation process. We have already begun to examine the possibility of extending the system functionality to support automated annotation of particular scenarios. This is made possible by taking advantage of the global coordinate system, which enables DANTE to recognise person-object and object-object interaction based on their relative positioning and orientation. Focusing an example upon the ADL of drink preparation, DANTE was able to accuractely detect the actions, pick-up kettle, pour, and pour into cup. Although these are our initial findings, more experimentation is required to validate this concept. Another opportunity DANTE provides is assisting with identification of interleaved activities such as the ADL of cooking, interrupted by another ADL such as a telephone call. The challenge here lies in ensuring that the original tasks is returned to and completed following the interruption. Indeed, this issue is particularly relevant in assisting persons with cognitive impairments [15]. Coupled with the aforementioned automatic annotation, the DANTE system could be extended to act upon actions which are started (activated) but interrupted by another activity. Actions could be associated with a guideline duration and used to prompt prospective users if actions are not eventually completed (deactivated). Other environmental sensors or objects with sensors attached could be tagged with markers and DANTE used to indicate important scenes in the recorded data, which when compared with the timestamps of other sensors could be used to quickly 'mine' those sensor data. Additionally, DANTE could be used to manage battery life of other sensing technology. For example, accelerometer based sensors could receive 'wake' signals through DANTE detecting potential movement of objects to which the sensors are attached. Finally, platform independent APIs could be developed for DANTE to support the rapid development and evaluation of context-aware algorithms, based on data recorded and annotated by the DANTE system.

5 Conclusions

Video based systems for monitoring and supporting ADLs are not without their challenges. Indeed, occlusion of objects in addition to privacy considerations remain key challenges. In an attempt to address the former, we have undertaken some preliminary experiments, which examine the potential for coupling DANTE with other sensor platforms. Experiments were conducted at the Univeristy of Ulster within their smart lab environment [13]. The Smart Environment consists of a Kitchen and a Livingroom, as shown in Fig.5 . To date, the DANTE system has only been evaluated within the Kitchen, which covers an area of approximately 17m^2. The Kitchen comprises a sensor rich environment, from which behaviour analysis and activity assistance problems are investigated. A set of object were tagged for use with the DANTE system.

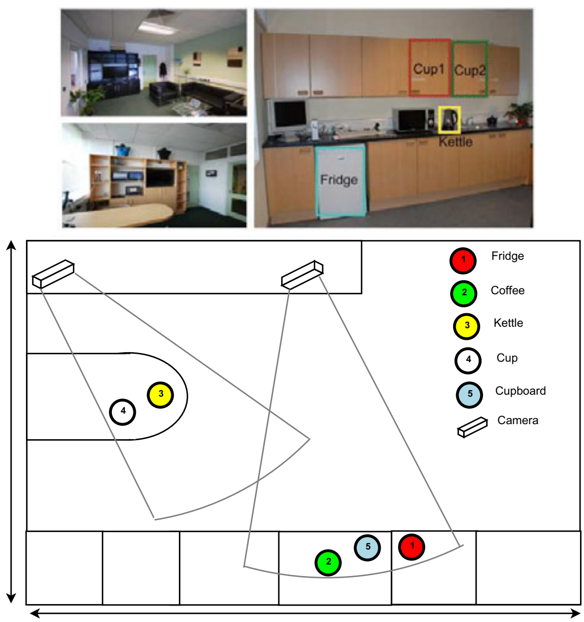

Fig. 5. Experiment Setup

The experimental setup included a wireless sensor network platform [16], coupled with DANTE markers, which were attached to a number of objects within the Smart Kitchen: fridge; 2 cupboards;a cup; and a kettle. Data was recorded using both Sunspots (sampled at 5Hz) and the DANTE recording module, and compared based on timestamp information. The summary findings, presented here, indicate that it is possible to detect a range of events through both modalities. As can be observed from Fig.6, both the Sunspots and DANTE system

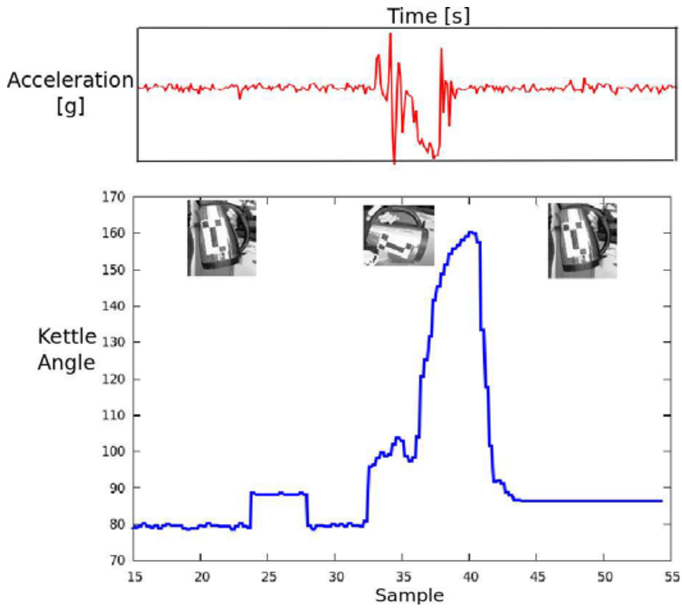

Fig. 6. (Upper Graph) Kettle acceleration measured by sunspot sensor, (Lower Graph) kettle orientation angle in respect to the floor

where able to independently detect that the kettle was ‘moving’. By comparison, the DANTE system was able to accurately detect the orientation of the object and thus infer when the kettle was being ‘poured’.

In our future work, we propose to use both the SunSpots and DANTE system for monitoring ADLs. The DANTE annotation software module can easily be extended to incorporate the annotation of both video and accelerometer data. We believe that we can utilise accelerometer based data to overcome potential occlusion scenarios where neither of the cameras can detect an ongoing activity. As previously discussed, one of our main focuses is to extend the DANTE system to support automatic annotation process. Indeed, preliminary experiments indicate that is possible to automaticaly annotate scenes. Activities can be detected based on the relative position between objects and their orientation. During our experiments, the system was able to detect, for example, interactions with the fridge and the pouring of the kettle by analysing the orientation of tracked objects.

In addition to reducing the amount of saved data, an automated process could significantly reduce potential privacy issues as recorded frames would not be required for manual annotation and therefore could be discarded. As a result we believe that the DANTE system could offer significant benefits to the research community in terms of facilitating the generation of sensor-rich datasets for supporting context-aware computing.

References

1. The Academy of Medical Sciences. Rejuvenating Ageing Research (September 2009), `http://www.acmedsci.ac.uk/p99puid161.html/` (accessed July 2, 2010)
2. Nugent, C.D., Fiorini, P., Finlay, D.D., Tsumaki, Y., Prassler, E.: Home Automation as a means of independent living. IEEE Transactions on Automation Science and Engineering 5(1), 1–9 (2008)
3. Szewcyzk, S., Dwan, K., Minor, B., Swedlove, B., Cook, D.: Annotating smart environment sensor data for activity learning. Technol. Health Care 17 3, 161–169 (2009)
4. Intille, S., Nawyn, J., Logan, B., Abowd, G.: Developing shared home behaviour datasets to advance HCI and Ubiquitous compuitng research. In: Intille, S., Nawyn, J., Logan, B., Abowd, G. (eds.) Proceedings oh the 27th Annual Conference Extended Abstracts and Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 04-09, pp. 4763–4766. ACM, New York (2009)
5. Cook, D., Schmitter-Edgecombe, M., Crandall, A., Sanders, C., Thomas, B.: Collecting and disseminating smart homes sensor data in the CASAS project. In: Proceedings of the CHI Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research (2009)
6. Coyle, L., Ye, J., McKeever, S., Knox, S., Stabeler, M., Dobson, S., Nixon, P.: Gathering Datasets for Activity Identification. In: Workshop on Developing Shared Home Behaviour Datasets to avdance HCI and Ubiquitous Computing Research at CHI 2009 (2009)
7. Logan, B., Healey, J., Philipose, M., Tapia, E.M., Intille, S.S.: A long-term evaluation of sensing modalities for activity recognition. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 483–500. Springer, Heidelberg (2007)
8. van Kasteren, T., Noulas, A., Englebienne, G., Krose, B.: Accurate activity recognition in a home setting. In: Ubicomp 2008: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 1–9. ACM, New York (2008)
9. Wren, C., Munguia-Tapia, E.: Toward scalable activity recognition for sensor networks. In: Hazas, M., Krumm, J., Strang, T. (eds.) LoCA 2006. LNCS, vol. 3987, pp. 168–185. Springer, Heidelberg (2006)
10. Tapia, E.M., Intille, S.S., Larson, K.: Activity recognition in the home using simple and ubiquitous sensors. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg (2004)
11. van Kasteren, T.L.M., Krose, B.J.A.: A sensing and annotation system for recording datasets from mutliple homes. In: Proceedings of the 27th Annual Conference Extended Abstracts and Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 04-09, pp. 4763–4766. ACM, New York (2009)
12. Hong, X., Nugent, C., Finlay, D., Chen, L., Davies, R., Wang, H., Donnelly, M., Zheng, H., Mulvenna, M.: Open Home: Approaches to Constructing Sharable Datasets within Smart Homes. In: Proceedings of the 27th Annual Conference Extended Abstracts and Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 04-09, pp. 4763–4766. ACM, New York (2009)
13. Nugent, C.D., Mulvenna, M.D., Hong, X., Devlin, S.: Experiences in the development of a smart lab. International Journal of Biomedical Engineering and Technology 2(4), 319–331 (2009)

14. Claron Technology Inc., "Micron Tracker" (2009), http://www.clarontech.com/measurement.php (accessed July 14, 2010)
15. Magherini, T., Parente, G., Nugent, C.D., Donnelly, M.P., Vicario, E., Cruciani, F., Paggetti, C.: Temporal logic Bounded Model-Checking for Recognition of Activities of Daily Living. Transaction on Biomedical Engineering Letters (July 2010) (submitted)
16. Sun Microsystem Laboratories, "Sun SPOT World", http://sunspotworld.com/ (accessed: July 14, 2010)