View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Explore Bristol Research

Hall, J., Hannuna, S., Camplani, M., Mirmehdi, M., Aldamen, D., Burghardt, T., ... Craddock, I. (2016). Designing a video monitoring system for AAL applications: the SPHERE case study. In 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016). [7801345] Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1049/ic.2016.0061

Peer reviewed version

Link to published version (if available):
10.1049/ic.2016.0061

Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at http://ieeexplore.ieee.org/document/7801345/. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms.html

Designing a Video Monitoring System for AAL applications: The SPHERE Case Study

J. Hall, S. Hannuna, M. Camplani, M. Mirmehdi, D. Damen
T. Burghardt, L. Tao, A. Paiement, I. Craddock

SPHERE, Faculty of Engineering, University of Bristol, Bristol, BS8 1UB

Keywords: Video Monitoring, Multi-cameras, RGBD sensor, eHealth.

Abstract

IT based Healthcare platforms have been widely recognized by research communities and institutions as key players in the future of home-based health monitoring and care. Features like personalised care, continuous monitoring, and reduced costs are fostering the research and use of these technologies. In this paper, we describe the design and implementation of the video monitoring system of the SPHERE platform (Sensor Platform for Healthcare in a Residential Environment). SPHERE aims to develop a smart home platform based on low cost, non-medical sensors. We present a detailed description of the hardware and software infrastructure designed and tested in real life scenarios, with particular emphasis on the design considerations employed to foster collaboration, the real time and budget constraints, and mid-scale deployment plan of our case study.

1 Introduction

Medical progress has increased life expectancy causing a dramatic increase in the elderly population. By 2050, 20% of the population will be over 60 years old [19]. Not only does old age impact on daily life with its associated chronic age-related diseases, but it also brings additional societal challenges. Our health-care services experience extra pressure, especially in terms of cost and shortages in caregivers [19]. These factors, and the fact that in general patients prefer to stay in the more familiar and comfortable environment of their home, has stimulated research and development into new technologies to facilitate health care practices in residential environments. Research interest in smart homes for eHealth and Active and Assisted Living (AAL) has continuously grown in the recent years. This research has examined different diseases and target groups [19, 4] and technologies ranging from wearable devices gathering movement information [27] and vital signals [4], to environmental sensors [1] and video based platforms [9].

An important feature of video based sensing is that it facilitates, in principle, the acquisition of useful and continuous information pertaining to human movement and activities that can be used to quantitatively and qualitatively assess the status of patients after specific treatments or after a rehabilitation period. For example, a depth camera based system was presented in [3], where video data was used to analyze patients' functional movements. A similar experimental setup was deployed in [22] where a movement quality assessment measure based on manifold learning and a Markovian assumption was presented. Video based systems are efficient for implementing alert systems to detect dangerous events like falls, as in [20]. Furthermore, video data analysis allows one to identify specific actions, long term activities, and behavioural patterns [9], with some exploiting contextual information [11]. While video based platforms offer the opportunity to extract unique, continuous, and rich information from the home environment, they also present a number of disadvantages, such as privacy issues [9], user acceptance and system cost and scalability. Finally, the accepted challenges of *computer vision*, such as arbitrary body poses, changing illumination, occlusion, and low cost/low resolution are still unsolved problems [23] even if depth data is combined with colour [6]. Furthermore, these issues are greatly amplified for AAL monitoring applications that operate in unconstrained environments and in long term scenarios. Different video acquisition architectures of widely varying complexities have been used for AAL applications. For example, a single camera was used in [16] to target specific functional mobility actions taking place in the living room area,with only snapshots and descriptions of detected actions stored. A centralised system was employed in the senior housing complex, TigerPlace [13], where depth data is stored in a central server which also contains labels, such as movement, inactivity, fall, etc. Researchers can access the stored data via a web interface. More flexible approaches based on smart cameras and distributed processing have been also employed, for example in [14, 15], where colour data was processed in each node and processing results transmitted to central storage nodes. As previously mentioned, RGBD cameras are becoming increasingly more deployed in AAL applications, such as [9, 20, 27, 21, 22, 11]. For a review of video based AAL systems, the reader is referred to [7, 8].

In this paper we present our video monitoring system developed for the SPHERE project (Sensor Platform for Healthcare in a Residential Environment). It aims to develop a low-cost, unobtrusive sensor platform for monitoring adults of any age in the home environment. We will focus on the main design challenges and choices, in terms of hardware (Section 2) and software (Section 3), for the development of the SPHERE video infrastructure. In particular, we shall highlight the importance and impact of real time considerations, budget constraints, and the reliability needed for our mid-scale development plan and user acceptance. Furthermore, we present a summary of the computer vision applications (Section 4) tested and implemented within the proposed system.

2 The SPHERE video monitoring system

SPHERE aims to develop a smart home platform comprising non-medical networked sensors, to obtain a rich description of the home environment and occupants' behaviour. Three main groups of sensors have been employed: (a) environmental sensors, which monitor temperature, humidity, luminosity, noise level, air quality, room occupancy, electricity metering, and water consumption, (b) RGBD vision sensors, and (c) low power wearable sensors that use accelerometers to measure body movements and to identify specific actions. SPHERE has already installed the first version of its system in a test house in Bristol (UK) used for short-to long-term user studies and architecture validation. The final goal of SPHERE is to deploy the sensor platform in up to 100 homes in Bristol for long-term, in-the-wild studies. The SPHERE system architecture (see Figure 1) includes a back-end (SPHERE Data Hub), which is made-up of a number of storage devices to collect and analyze all data collected from the 100 houses. Each house will be connected to the Data Hub through the SPHERE Home Gateway, that has several additional functions, including collecting data from the sensor network and monitoring the system status. Wireless and BLE links are used to ensure connectivity between the heterogeneous sensor networks and the Home Gateway. At the application layer, the SPHERE system makes extensive use of the MQ Telemetry Transport (MQTT) protocol for data collection as well as for system monitoring. More details about the overall SPHERE architecture can be found in [25, 27].

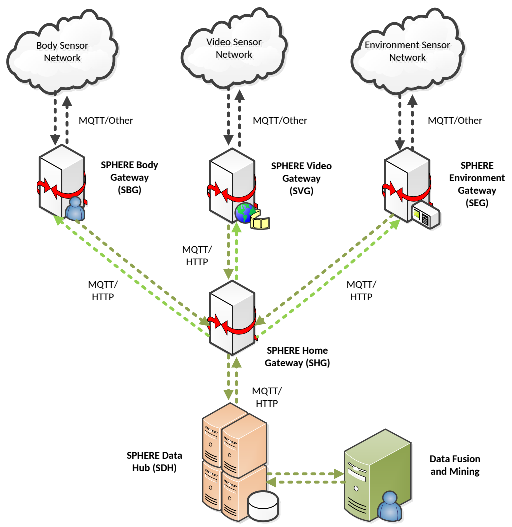

Figure 1. SPHERE system architecture [25].

2.1 Hardware platform

The video monitoring component of the SPHERE system is a real-time multi-camera system, which is tasked not only with tracking people in their home environment, but also with providing continuous quality of movement and activity recognition data. More details of these is reported in Section 4.

The selection of the hardware for the video monitoring system has been driven by different factors, such as ease of integration with other sensors, user acceptance, and deployment cost. Considering the 100 houses deployment plan and its financial viability, the low cost consumer RGBD camera, Asus Xtion, was selected. Furthermore, as previously mentioned, RGBD devices now represent the state of the art for indoor activity monitoring [9, 20, 21, 22, 11]. For the SPHERE project, the camera needs to be coupled with a machine with suitable processing capacity, minimal intrusion on the user, and minimal cost. The Intel Next Unit of Computing (NUC) with 8GB of RAM and an i5 processor meets these requirements. Its compact size and relatively low cost, when compared to most workstations, allows it to be placed strategically close to other sensors. Another important attribute of the NUC is its four USB 3.0 ports, which provide ample bandwidth for simultaneous capture from up to four RGBD cameras with an acquisition rate of 30fps for each camera at VGA resolution. This gives the system more versatility in deployment. Specifically, a range of configurations is possible: one NUC operating all of the cameras, (that was the first design choice for SPHERE [25]), one NUC per camera (Figure 2), or hybrid configurations, depending on users' needs as well as the circumstances of each deployment. We can store up to 64GB of data locally in each NUC, corresponding approximately to 2 hours in a single camera configuration, including RGBD and tracking results. This data can then be streamed using a buffering strategy to the central MySQL and Mongo databases (see Section 3). There is no ceiling on the number of NUCs that can be utilised in a single deployment given appropriate network capacity and server performance. In any eventual deployment in real homes, our configuration will have minimal storage requirements as only the processed results (i.e. tracking, segmentation) will be stored and not the raw video data.

Figure 2. NUC and Asus Xtion sensor in SPHERE house.

A vital consideration for all of the SPHERE sensors, and the video subsystem in particular, is installation time and long term reliability. Recalling that the current plan is for SPHERE to deploy its system into as many as 100 homes, with up to three NUCs and ASUS Xtions per home, the installation and management of such a large enterprise rapidly becomes impossible without optimised installation procedures and ultra reliable subsystems. In order to fulfil these requirements, each NUC runs Ubuntu with the video system set up to run as a service, when the machine boots up. This makes the system robust in the event of a temporary loss of power. To facilitate quick configuration, a standardised system image is employed. The set up process is streamlined such that a brand new NUC in its box can be ready for deployment in about 10 minutes.

Figure 3 illustrates how the NUC and Asus Xtion fit into the hardware infrastructure. Note that the NUCs' clocks are synchronised via NTP and the captured frames are time-stamped at the time of capture, before being buffered for subsequent transfer to the main gateway.

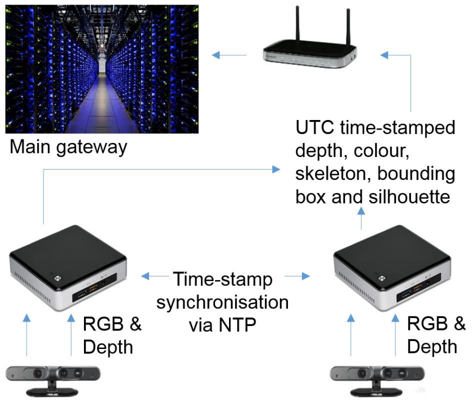

Figure 3. Video Monitoring System Architecture.

3 Collaborative Software Platform for Data Management and Analysis

SPHERE produces enormous amounts of video data and a variety of features associated with it. This data is used in a number of pure and applied computer vision research projects, including person re-identification, quality of movement assessment, calorific expenditure, robust real-time tracking, and activity recognition. The data recorded by the SPHERE system, as with any other vision system, is not perfect and sometimes contains false detections and tracking errors. Furthermore, for particular studies it is necessary to ground truth the data when supervised approaches to learning are employed. For example, the bounding boxes/tracklets relating to particular individuals need to be given a consistent identity for person re-identification as do the start and end times of particular activities.

Correctly labelled data sets are extremely valuable to the community for development of different algorithms and comparative evaluation. In a large project, with a number of collaborators, the risk of duplication of effort is high when producing these data sets. Different researchers commonly use different tools, languages and strategies for labelling. This may also include relatively ad hoc decisions regarding data formats. This fragmented approach produces barriers to collaboration as researchers have to not only obtain code and data from their peers, but also learn how to use it and incorporate it into

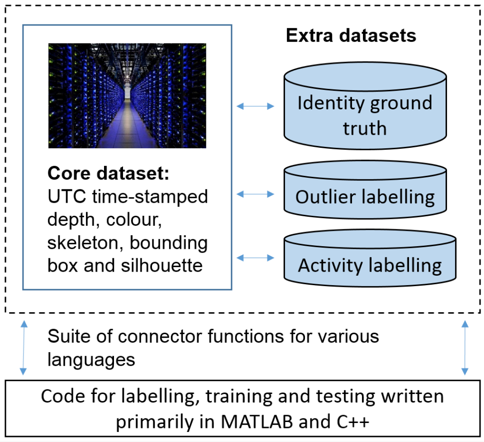

Figure 4. Collaborative infrastructure for data analysis.

their normal work-flow. To combat this fragmentation and foster synergies between individuals, work packages, and groups within SPHERE, we have developed a collaborative infrastructure for data analysis (Figure 4). Our infrastructure features a central data repository, consisting of MySQL and Mongo databases. During data capture, raw video and depth streams, bounding boxes, silhouettes, and skeletons are buffered and then stored in this central repository. For a particular capture session, metadata, including participant ids and start and end time-stamps, are stored in a separate database. When a researcher wishes to work on a particular data set, they simply look up the time-stamps and then enter them into either the C++ or MATLAB interfaces. Any labelling undertaken by that researcher, is stored centrally in tables that are accessible by collaborators and linked to the original data set. In this way teams can work on the same problem concurrently, on a platform and language of their choice.

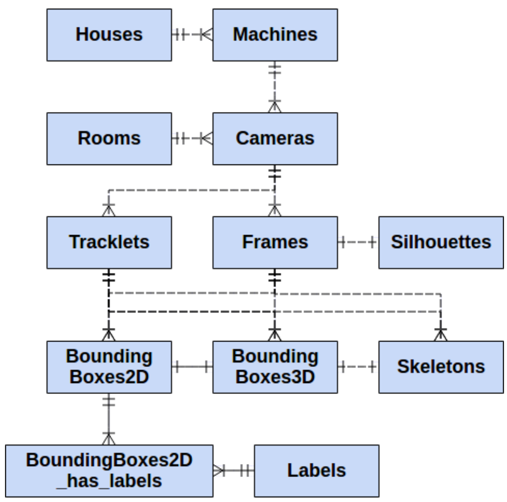

Figure 5. Database entity relationship diagram

Figure 5 illustrates this video database design. The top four rows represent the 'core' data which is obtained at the time of capture. The labels represent annotations added subsequently by researchers working with the data. In this case the labels refer to the identities of individuals navigating the SPHERE house, with outliers (false detections) being given a consistent label. This data may now serve as a ground truth for tracking and detection algorithms or research into person re-identification for example. We have found this has made very real productivity differences, with researchers discussing and prototyping ideas in hours rather than days.

3.1 Software architecture

Software reliability is critical for facilitating long term operational periods without regular physical access. To enable this, we have employed an object oriented design and popular software design patterns where possible. The design principles underpinning the software system reflect our ambitions to build a collaborative platform which naturally encourages cooperation and the development of reusable and easy to extend code.

The software system is centred around the *observer* design pattern. This is useful for decoupling the problems of data acquisition, processing, and storage into separate classes which can be developed independently. The acquisition is encapsulated in a camera class, that controls camera functionality providing RGBD, and basic tracking information as bounding boxes and skeleton joints. This class is the *subject* and it keeps an *observer* collection. These observers are notified when new data is available through a standard interface. This allows for rapid prototyping and deployment by researchers without needing to understand the underlying hardware, or the SPHERE infrastructure. In addition to employing this low-level design pattern, we add a set of simple operators which allow for a higher level designs to be produced, using observers as components. These operators are *combine, filter* and *par*.

$$combine : Observers \times Observers \rightarrow Observers$$
$$filter : Observers \times Observers \rightarrow Observers$$
$$par : Observers \rightarrow Observers$$

combine is a binary operator which takes two observers and wraps them in the interface of a single one. This newly created observer has the behaviour of the two given observers. *filter* is a binary operator which takes a pair of observers and creates a new observer. This newly created observer has the behaviour of the original observers, but sequences them so that the first observer can mutate the data before it is passed to the second. *par* is a unary operator which takes a single observer and creates a new observer which behaves in exactly the same way, except that its input is buffered and it is executed on another thread. This allows us to make use of all the computational resources available on the target hardware. The *par* operator is not strictly necessary at this level. We could have chosen for this decision to be made inside the observers, but this would introduce performance overheads that were not visible at the high level. Hence, we decided that observers' parallelisation should be exposed at a higher level so that the overhead introduced is explicit and the benefit is clear. This also means that researchers do not need to consider how, or when, their observer will be executed when developing their experiments.

Concrete applications of the described high-level design patterns used in the proposed system are shown in Figure 6. Depth based human detector (*DepthBasedDetector*) and unique user Id allocation (*UniqueUserID*) to the tracklets are examples of *filter*s as they modify the data stream adding additional information and passing to the next stage. For example *DepthBasedDetector* is used to boost the confidence of the basic tracking performed in the camera module, and the *UniqueUserID* is used to assign each tracklet a unique id among all of the cameras. Parallelisation is used to provide a buffer between the high performance processing pipeline and the potentially slow network and storage medium.

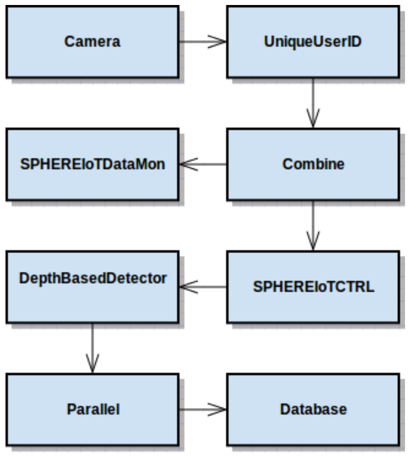

Figure 6. Data flow diagram of our video monitoring system

Given these operators and a set of observers we are able to quickly design and deploy complex applications which contain research experiments with minimal effort from the researchers themselves. The class hierarchy is shown in Figure 7. In the SPHERE video monitoring system the Observers are *UniqueUserID* and *DepthBasedDetector* discussed above, and *SPHEREIoTDataMon* that manages MQTT connection to SPHERE data HUB, *SPHEREIoTCTRL* that handles control commands from the SPHERE platform, such as start/stop recording and system initialisation, and *Database* that manages video storage. These, with our observer operators, are used to specify our software system as shown in Figure 6. The above application description illustrates how combining *combine* or *filter* and the *Observers* forms a tree. This combination of the observer design pattern and these operators comes close to describing functional reactive programming (FRP), specifically discreet push oriented FRP. With the addition of a single operator, to merge branches in the tree to create a graph, this would be equivalent to FRP [12]. In this model, our observers are behaviours and the data from the camera is an event stream.

Another practical example exploiting this design strategy is the way the video system is integrated into the SPHERE platform as an IoT device. This is achieved via the class *SPHEREIoTDataMon* which takes the video data, excluding the colour and depth frames themselves, and serialises all the processed data (such as bounding boxes and tracklets information) into JSON strings. These strings are then transmitted via

MQTT to the SPHERE data HUB. The choice to apply this structure at both the high and low level of the video monitoring system has given us a structure which is robust and flexible. This gives researchers the power to test and deploy research systems to our video system as shown in the following section.

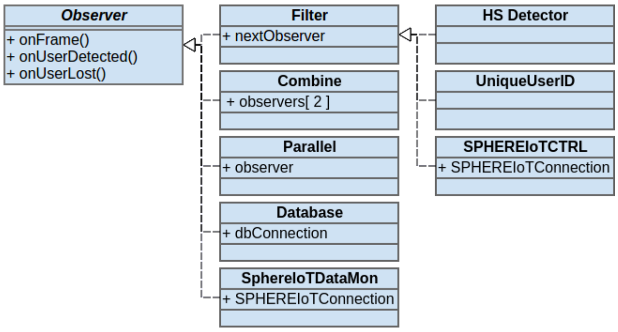

Figure 7. Video monitoring system class hierarchy

4 SPHERE vision based Applications

We now summarize how the HW and SW architectures presented above are employed to accomplish several eHealth applications within SPHERE. The complete vision based pipeline in each NUC device is shown in Figure 8. The data acquisition module gathers the camera stream synchronized with the other sensors' data. Depth and colour data are processed by a person detector and tracking module to continuously estimate the residents' position. The system currently used in the SPHERE house is the depth based OpenNI tracker [18] which also provides the estimation of user skeletal joints. This tracker and its enhanced output have been widely used in eHealth applications, e.g. [2, 20, 22]. Nevertheless, the proposed modular software system allows us to plug and play different detectors and trackers. Other real time detectors based on colour and depth data, such as [17] and real time trackers [17, 5], have been tested and tuned into the proposed system.

Even without the rest of the SPHERE platform's data, the video system's output is still a rich source of information. The aspect ratio, velocity, and location of the bounding boxes gives useful clues for detecting activities and behaviours. These bounding box characteristics are analysed by a cascade of SVM classifiers to identify basic low level functional movements, such as sitting, standing, lying, walking and their corresponding transitions. They allow us to accumulate statistics about speed of motion and hence to infer important clinical measurements *in the wild*. Some example studies that we have carried out so far are recognising actions [21], and quality and intensity of movement [22], and identifying typical indoor activities of daily living and routine modeling [26], amongst others. To overcome the unreliability of skeleton data, provided by the OpenNI tracker, for non-frontal views we have designed a new depth based pose estimation system that can support a large range of views [10]. Our low-level movement recognition work can be enriched by higher level analysis based on fusing video features with other sensor data, such as accelerometer data, and we hope to report on these in the near future.

Alongside these research activities, we have also released several datasets[1] for the research community. In [21], we presented a dataset, based on RGBD and wearable data, that contains actions of daily life collected in the SPHERE house. An RGBD and skeleton dataset for quality of motion estimation was released in [22]. Finally, multi-sensor data of single users living in the SPHERE house was released in [24].

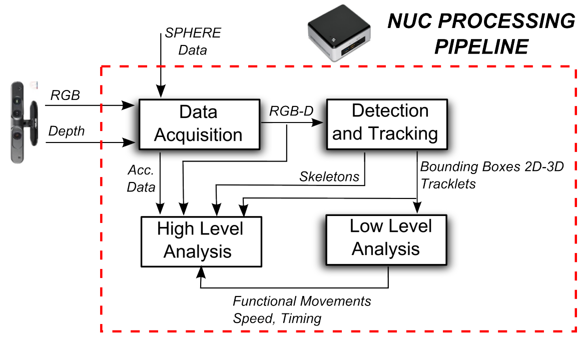

Figure 8. NUC processing pipeline for each camera.

5 Conclusions

In this paper we presented the video monitoring system developed within the SPHERE project. Designing a reliable video monitoring infrastructure for residential environments is a challenging task as many factors, such as cost, user acceptance, and application domains deeply affect design choices. We introduced a flexible hardware architecture based on NUCs which can deal with different camera configurations required for different installation setups during the deployment phase. Low cost RGBD devices are fundamental hardware components of the proposed system as the combination of colour and depth data provides a richer description of environments and humans. Furthermore, we proposed an efficient and highly modular and lightweight software architecture that allows to easily adapt the system to different recording devices, i.e new depth devices coming into the market, and to target different application scenarios, as the acquisition and the processing phase are decoupled and independent. Finally, our collaborative platform for data management and analysis ensures the huge amounts of data gathered are accessible to researchers in SPHERE and in the research community in a flexible way such that researchers can work on the same or different problems whilst seamlessly sharing annotations and other extra metadata with one another.

Acknowledgements

This work was performed under the SPHERE IRC project funded by the UK Engineering and Physical Sciences Research Council (EPSRC), Grant EP/K031910/1. We would also like to acknowledge the contribution of the SPHERE team.

[1]http://www.irc-sphere.ac.uk/

References

[1] G. Acampora, D. J. Cook, P. Rashidi, and A. V. Vasilakos. A survey on ambient intelligence in healthcare. *Proceedings of the IEEE*, 101(12):2470–2494, 2013.

[2] I. Ar and Y. S. Akgul. A Computerized Recognition System for the Home-Based Physiotherapy Exercises Using an RGBD Camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(6):1160–1171, 2014.

[3] T. Banerjee, M. Skubic, J. M. Keller, and C. Abbott. Sit-to-Stand Measurement for In-Home Monitoring Using Voxel Analysis. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1502–1509, 2014.

[4] S. Blackman et al. Ambient Assisted Living Technologies for Aging Well: A Scoping Review. *J. Intelligent Systems*, 25(1):55–69, 2016.

[5] M. Camplani et al. Real-time RGB-D tracking with depth scaling kernelised correlation filters and occlusion handling. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 145.1–145.11, 2015.

[6] M. Camplani et al. Multiple human tracking in RGB-D Data: A Survey. *CoRR*, abs/1606.04450, 2016.

[7] F. Cardinaux, D. Bhowmik, C. Abhayaratne, and M. S. Hawley. Video based technology for ambient assisted living: A review of the literature. *Journal of Ambient Intelligence and Smart Environments*, 3(3):253–269, 2011.

[8] A. A. Chaaraoui, P. Climent-Pérez, and F. Flórez-Revuelta. A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living. *Expert Systems with Applications*, 39(12):10873 – 10888, 2012.

[9] A. A. Chaaraoui et al. A Vision-Based System for Intelligent Monitoring: Human Behaviour Analysis and Privacy by Context. *Sensors*, 14(5):8895, 2014.

[10] Ben Crabbe, Adeline Paiement, Sion Hannuna, and Majid Mirmehdi. Skeleton-free body pose estimation from depth images for movement analysis. In *ChaLearn LaP workshop at ICCV*, 2015.

[11] C. F. Crispim-Junior et al. Semantic event fusion of different visual modality concepts for activity recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1598–1611, 2016.

[12] C. Elliott and P. Hudak. Functional reactive animation. In *International Conference on Functional Programming*, pages 163–173, 1997.

[13] M. Enayati et al. A novel web-based depth video rewind approach toward fall preventive interventions in hospitals. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4511–4514, 2014.

[14] S. Fleck and W. Strasser. Smart Camera Based Monitoring System and Its Application to Assisted Living. *Proceedings of the IEEE*, 96(10):1698–1714, 2008.

[15] Stefano Ghidoni et al. A distributed perception infrastructure for robot assisted living. *Robotics and Autonomous Systems*, 62(9):1316 – 1328, 2014.

[16] Kam-Yiu L et al. Activity tracking and monitoring of patients with alzheimer's disease. *Multimedia Tools and Applications*, pages 1–33, 2015.

[17] Matteo Munaro and Emanuele Menegatti. Fast RGB-D people tracking for service robots. *Autonomous Robots*, pages 1–16, 2014.

[18] OpenNI organization. *OpenNI User Guide*, 2010.

[19] P. Rashidi and A. Mihailidis. A Survey on Ambient-Assisted Living Tools for Older Adults. *IEEE Journal of Biomedical and Health Informatics*, 17(3):579–590, 2013.

[20] E. E. Stone and M. Skubic. Fall detection in homes of older adults using the microsoft kinect. *IEEE Journal of Biomedical and Health Informatics*, 19(1):290–301, 2015.

[21] L. Tao et al. A comparative home activity monitoring study using visual and inertial sensors. In *International Conference on E-health Networking, Application Services (HealthCom)*, pages 644–647, 2015.

[22] L. Tao et al. A comparative study of pose representation and dynamics modelling for online motion quality assessment. *Computer Vision and Image Understanding*, 148:136 – 152, 2016.

[23] N. S. Thakoor et al. People Tracking in Camera Networks: Three Open Questions. *Computer*, 48(3):78–86, 2015.

[24] N. Twomey et al. The SPHERE challenge: Activity recognition with multimodal sensor data. *CoRR*, abs/1603.00797, 2016.

[25] P. Woznowski et al. A multi-modal sensor infrastructure for healthcare in a residential environment. In *IEEE International Conference on Communication Workshop (ICCW)*, pages 271–277, 2015.

[26] Y. Xu, D. Bull, and D. Damen. Unsupervised daily routine modelling from a depth sensor using top-down and bottom-up hierarchies. In *Asian Conference on Pattern Recognition (ACPR)*, pages 056–060, 2015.

[27] N. Zhu et al. Bridging e-health and the internet of things: The sphere project. *IEEE Intelligent Systems*, 30(4):39–46, 2015.