Behavior analysis for elderly care using a network of low-resolution visual sensors

Mohamed Eldib
Francis Deboeverie
Wilfried Philips
Hamid Aghajan

Behavior analysis for elderly care using a network of low-resolution visual sensors

Mohamed Eldib,[a,*] Francis Deboeverie,[a] Wilfried Philips,[a] and Hamid Aghajan[a,b]
[a]Ghent University, TELIN-IPI-IMINDS, Sint-Pietersnieuwstraat 41, Ghent, Belgium
[b]Ambient Intelligence Research Lab, David Packard Building, Stanford, California 94305–9515, United States

Abstract. Recent advancements in visual sensor technologies have made behavior analysis practical for in-home monitoring systems. The current in-home monitoring systems face several challenges: (1) visual sensor calibration is a difficult task and not practical in real-life because of the need for recalibration when the visual sensors are moved accidentally by a caregiver or the senior citizen, (2) privacy concerns, and (3) the high hardware installation cost. We propose to use a network of cheap low-resolution visual sensors (30×30 pixels) for long-term behavior analysis. The behavior analysis starts by visual feature selection based on foreground/background detection to track the motion level in each visual sensor. Then a hidden Markov model (HMM) is used to estimate the user's locations without calibration. Finally, an activity discovery approach is proposed using spatial and temporal contexts. We performed experiments on 10 months of real-life data. We show that the HMM approach outperforms the k-nearest neighbor classifier against ground truth for 30 days. Our framework is able to discover 13 activities of daily livings (ADL parameters). More specifically, we analyze mobility patterns and some of the key ADL parameters to detect increasing or decreasing health conditions. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JEI.25.4.041003]

1 Introduction

Due to rapid improvement in the medical sector over the last decades, the average lifespan of seniors has been steadily increasing. According to the UN world population prospects,[1] by 2020, more than 25% of the European population will be over 60. During the same period, the number of people over 80 will double. As a person ages, the immune system is getting weaker and body organs begin to degrade. This results in a variety of chronic and degenerative diseases, such as Alzheimer's disease, diabetes, Parkinson's disease, heart attacks, and arthritis.[2]

Nowadays, aging-associated diseases have a significant impact on health care. The institutions that take care of the senior citizens will run into operational and financial problems in the approaching decade. As a financial consequence, it is estimated that the total number of workers is going to decrease from four workers per retiree to two workers per retiree. This reduction in the workforce is going to act as a drag on the growth and the per capita income, with a reduction risk in potential growth. This will increase the health care expenditure over the next couple of decades.[3–5] Also, the senior citizens in Europe use more than 54% of the hospital care followed by 19% of nursing home care in their last year of life.[6] Therefore, the current health care system is becoming strained as the aging population increases over time.

At the same time, a shortage of professional caregivers for the aging population is predicted. Therefore, family members acting as informal caregivers will play a more prominent role. Clinical observation[7] showed that supporting-dependent individuals at home create many complications for the informal caregiver, such as high levels of distress and depression. With an increase in the aging-associated diseases, the demographic old-age dependency ratio is predicted to increase in 2060 from 27.8% to 50.1% in the EU.[5] This raises concerns on the quality of the offered health care service for the senior citizen in the future.

Aging-in-place presents itself as a promising solution for health care systems. It is defined by the Center for Disease Control as "the ability to live in one's own home and community safely, independently, and comfortably, regardless of age, income, or ability level." Aging-in-place has gained a lot of attention in recent years due to the fact that many senior citizens prefer to age and maintain their independence as long as possible in their own homes[8] because of emotional and physical associations, memories, and comfort. Aging-in-place promotes the well-being of older people without sacrificing the quality of life in a familiar environment and maintains valuable social networks. The success of aging-in-place depends on ambient-assisted living (AAL) tools, which have witnessed tremendous improvements in the last few years. AAL tools provide supervision and assistance with activities of daily living (ADL) to prevent, cure, and improve wellness and health conditions of seniors.

In this paper, we utilize low-resolution visual sensors to build an in-home monitoring system. The system is installed in a service flat of a senior citizen. The results in this paper are based on data obtained from an elderly volunteer, 83 years old with diabetes and decreased mobility due to a little paralysis. The resident has a very clear mind. In previous work,[9] users' locations and mobility statistics were obtained from a robust people tracker based on recursive maximum

---

*Address all correspondence to: Mohamed Eldib, E-mail: mohamed.eldib@telin.ugent.be

likelihood principles. However, this people tracker requires accurate camera calibration. Camera calibration is a difficult task since it requires efforts to adjust the camera poses to have overlap between the cameras' fields of view (FOVs). Also, calibration is not practical in real-life monitoring environments because it may have to be repeated after accidentally moving the camera by a caregiver or the senior citizen. In this case, recalibration is needed.

The computer vision algorithms used in this paper are based on vision algorithms developed in the research project "Little Sister: low-cost monitoring for care and retail,"[10] which focuses on creating a sensor-based monitoring system that can match, in terms of performance, a combination of the body-worn devices and the high-resolution cameras at a much reduced cost. They are also one of the core components of the AAL Joint Programme project "SONOPA: social networks for older adults to promote an active life."[11] In SONOPA, the aim is to combine a social network with activity recognition in a smart home environment to stimulate and support activities and daily life tasks. SONOPA suggests suitable activities and social connections to the senior citizen automatically, proactively, and at the optimal time, while providing a simple bridge to the social network of the senior citizen. SONOPA achieves this by analyzing both physical and online activities of senior citizen users in their smart homes. We have an in-home monitoring system with visual sensors installed in Belgium, which has been operational for 10 months. This paper extends and improves the work of SONOPA and Little Sister with hidden Markov modeling and activity discovery techniques.

The main contributions of this paper are (1) the extraction of the elderly's location using visual sensor features and a hidden Markov model (HMM). This approach avoids the usage of tracking algorithms in a calibrated sensor network. We compare our approach with a k-nearest neighbor (kNN) classifier against collected ground truth for 30 days. (2) The introduction of a rule-based approach for activity discovery using spatial and temporal contexts. The ADL parameters span 10 months of real-life data in a service flat of a senior citizen. The data include many different activities, such as sitting, taking a nap, eating, cooking, taking a shower, going to the toilet, watching TV, sleeping, and being out-of-home. In contrast to earlier research,[12,13] we monitored real-life activity without resorting to simulations. Simulated data are obtained by people acting out senior citizens' life-style risk not being representative of real-life situations. Moreover, they are by necessity short, making it difficult to study the analysis of long-term trends. (3) the detailed analysis of some key ADL parameters is to detect some health changes.

The remainder of the paper is organized as follows. Related work in the literature is listed in the next section. Section 3 gives an overview of the service flat set-up with the in-home monitoring system. Section 4 explains the proposed behavior analysis approach. Section 5 shows the experimental results. Finally, Section 6 draws conclusions.

2 Related Work

The sensors used in AAL tools can be divided into two main categories: (1) wearable sensors and (2) ambient sensors. In the first category,[14–16] various wearable sensors, such as accelerometers, gyroscopes, proximity sensors, and e-textile sensors, are attached to the subject to monitor vital signs,

such as heart rate, respiration, blood pressure, glucose level, and muscle activity. Wearable sensors face a few disadvantages, such as limited battery life, high cost, missing data when the user forgets to wear the device, and the need to attach them to specific body parts to provide reliable measurements.

In the second category, ambient sensors are installed in the home environment by mounting them on the wall or the ceiling and/or embedding them in furniture and appliances. Passive infrared (PIR) motion sensors, visual sensors (including special technologies such as depth cameras), and radio frequency identification (RFID) are most popular in research.

Tables 1 and 2 summarize the different capabilities and properties of three in-home sensors: PIR, Kinect, and visual sensors. In Table 1, four capabilities such as location, presence, shape, and tracking of the three technologies are compared. PIR sensors have limited capabilities when they are compared to Kinect and visual sensors. PIR sensors can provide good presence detection accuracy, but they cannot provide very accurate information about the exact location (e.g., x and y positions). Also, PIR sensors cannot track multiple persons at the same time or do shape detection. On the contrary, Kinect and visual sensors have highly accurate location and presence detections, and both technologies can track multiple persons. Shape detection and skeleton extraction can be done more accurately using Kinect than visual sensors.

Table 2 shows several properties of PIR, Kinect, and visual sensors:

- Network density: The number of sensors required to be installed in an area to provide some specific service. In Ref. 17, the authors quantified the network density (ND) using the order of magnitude (in base 2) of the number of sensors. For instance, if a single camera can detect a person within area A, then the density of the camera solution is $\log_2(1) = 0$. PIR sensors require a high ND to provide accurate locations (ND = 4). A high ND requires a complex infrastructure, cumbersome to install and manage.

- Resolution: PIR sensors return a state "on" if human presence is detected within a certain sensing area, otherwise a state "off" is returned. Kinect has an infrared depth sensor with an image resolution of 640×480 pixels and a color camera sensor with an image

Table 1 Comparison between the different capabilities of PIR, visual, and Kinect sensors. H, M, and L stand for high, medium, and low values, respectively.

Technology	Location detection	Presence detection	Shape detection	Tracking
PIR sensors	L	M	Not possible	Single person
Visual sensors	H	H	L	Multiperson
Kinect sensors	H	H	H	Multiperson

Table 2 Comparison between the different properties of PIR, visual, and Kinect sensors. H, M, and L stand for high, medium, and low values, respectively.

Properties	PIR sensors	Visual sensors	Kinect sensors
ND	H	M	M
Resolution	Single pixel (on/off)	30 × 30 pixels	IR depth sensor: 640 × 480 pixels Color camera: 1280 × 1024 pixels
Space occupancy	L	L	H
Cost	L	M	H
Privacy concern	L	L	H
Operation (lighting)	No	Yes	IR depth sensor: No; Color camera: Yes
Applicability	Indoors	Indoors	Indoors/Outdoors
Battery life	H	L	Not possible

resolution of 1280 × 1024 pixels. Visual sensors provide an image resolution of 30 × 30 pixels.

- Space occupancy: The dimensions of Kinect, visual, and PIR sensors are ($w \times d \times h$): $37 \times 15 \times 12$ cm^3, $6.2 \times 4.1 \times 2$ cm^3,[18] and $3.2 \times 2.5 \times 2.8$ cm^3, respectively. The Kinect sensor clearly occupies more space than PIR and visual sensors.

- Cost: The Kinect sensor has advanced hardware components. This increases the price per unit (above 100 Euros), while the bill material of the visual sensor is under 25 Euros.[18] The PIR sensor is the cheapest solution.

- Privacy concern: User studies in the projects Little Sister and SONOPA indicated that the users attach high priority to privacy, and they agreed to install low-resolution cameras (e.g., visual sensors) or PIR sensors, but not high-resolution cameras (e.g., Kinect), which often raises privacy concerns. Visual sensors pose very little privacy issues since they are not capable of gathering detailed information.

- Operation: PIR sensors and the infrared depth sensor in Kinect do not require lighting conditions to operate, while visual sensors and the color camera in Kinect require sufficient lighting conditions to operate.

- Applicability: PIR and visual sensors can only be used in indoors scenarios (e.g., behavior analysis), while Kinect sensors can be used indoors and outdoors (e.g., car tracking).

- Battery life: PIR sensors have a longer battery life than Kinect and visual sensors, because PIR sensors consume less processing power. Kinect and visual sensors are installed in a wired setup and powered by mains electricity. Given the low power consumption of the visual sensors, it is possible to operate them on battery over prolonged periods of time.

From the detailed comparison in Tables 1 and 2, Kinect and visual sensors have similar and more powerful capabilities than PIR sensors. Furthermore, the properties of the visual sensors are more suitable than Kinect for in-home

monitoring systems, because of the affordable price and the preservation of privacy.[19] The images produced by the visual sensors are 30 × 30 pixels. In these images, privacy is maintained; thus, it is, for instance, hard to recognize faces. However, they are very useful in our in-home monitoring system to recognize activities and to detect behavior and behavioral changes of elderly. Examples of activities are going outside the home or receiving visits.[20] An example of a behavioral change is increased or decreased mobility measured from speed or walked distance.[21]

A single PIR sensor records the occupant's activities with only a binary state indicating whether there is a motion detected within its detection range. Thus, datasets recorded using PIR sensors are, in fact, a time series of sensor activation events, which contain very limited information that can be used to identify the corresponding individual. A single camera can capture rich information with different levels of granularities from the gross movements of subjects similar to that provided by simple motion detection sensors to richer information about posture, body motion, head and body orientation, fidgeting, and so on. In most cases, multiple PIR sensors and cameras are used in smart homes.

There have been many proposed approaches toward recognizing ADL in a home setting with PIR sensors and cameras, which can be broadly divided into two major categories: supervised and unsupervised approaches. In the supervised approaches, the task of recognizing ADL can be easily formatted into a classification problem where the model relies on labeled data for training the desired activities. Many popular machine learning algorithms such as support vector machine (SVM), naive Bayes classifier, decision tree, and neural network can be directly applied to activity recognition tasks.[22–27] Moreover, probabilistic graphical models, such as HMM, dynamic Bayesian network, and conditional random fields (CRFs), have been used to model the activity transition sequence for activity recognition purposes.[28–34]

PIR sensors are commonly used with supervised learning approaches. Ordóñez et al.[35] proposed hybrid discriminative models by combining an artificial neural network and an SVM with an HMM to recognize ADL parameters from PIR sensor streams. van Kasteren et al.[36] recognized

activities using an HMM and CRFs on a PIR sensor dataset of 28 days. In Ref. 37, they focused on modeling activities from PIR sensor data using hidden semi-Markov models (SMM) and semi-Markov CRFs.

Also, cameras have been used with supervised learning techniques. Chaaraoui et al.[38] proposed a vision-based monitoring system that relies on a multiview silhouette-based pose representation where key pose models are learned. Then dynamic time warping is used to recognize human actions. Ahmad and Lee[39] proposed a method for human action recognition from an arbitrary view image sequence by using optical flow and human body shape features. Finally, HMMs are trained to declare the action performed in the image sequence. Chung and Liu[40] applied a hierarchical context HMM for behavior understanding from video streams in a nursing center. Duong et al.[41] applied the switching hidden SMM to learn and recognize human behaviors and detect anomalies using multiple cameras. Natarajan and Nevatia[12] evaluated a coupled hidden SMM for activity recognition on simulated and laboratory data.

Even though the majority of the proposed activity recognition approaches are supervised methods, most of them share the same limitation that the accurate activity labels for PIR sensor datasets and cameras are very difficult to get. For almost all of the current smart home testbeds with PIR sensors and cameras, the data collection and data labeling are two separate processes for which the activity labeling for the collected data is extremely time consuming and laborious because it is usually based on direct video coding and manually labeling. Clearly, this limitation prevents the supervised approaches from being easily generalized to the real-world situation where activity labels are usually not available for a huge amount of sensor data.

Therefore, many unsupervised approaches have been proposed to handle the problem that activity labels are not available. Many of them are based on sequence mining approaches that use different sliding windows to find frequent patterns,[42–44] rule-based engines,[13,45] and topic models[46,47] to discover repeated activities from raw sensor event sequences. Castanedo et al.[47] tackled the problem of the large amount of sensor data by employing topic models to learn the latent structure and the dynamics of sensor network data in office environments. The authors used two datasets to analyze the data with the aim of learning and discovering what is happening in the monitored environment. Alwan et al.[13] explored the spatial–temporal relationships among PIR sensor events using a rule-based approach to infer the occurrence of activities with a high degree of confidence on 37 days of test data in a living laboratory. Theekakul et al.[45] proposed a rule-based framework using the mean of dynamic activities to infer the device orientation, from which the appropriate set of activity classification rules and threshold parameters can be selected. In Ref. 48, the authors monitored the behavioral patterns for a senior citizen living independently to perform analysis in the form of behavioral rules. An evolving fuzzy rule-based system has been proposed in Ref. 49 for modeling activities that evolve over time, according to the changes observed in the way an activity is performed from PIR sensor readings. The Millennium Home[50] detects deviations from normal daily activities by using rule-based techniques.

Despite the popularity of PIR sensors, they are known for having the following problems: (1) highly bursty output, which limits PIR systems to single-person scenarios; (2) self-triggering due to sudden changes in environmental conditions, such as temperature, ventilating, and air conditioning; (3) PIR sensors cannot sense immobile people.[17] With cameras, the detection of people standing is possible, because those persons tend to move parts of the upper body (head, shoulders, and hands), which could be easily detected by foreground detection algorithms.[51,52] For this reason, researchers found an alternative in using cameras to detect different ADL parameters and abnormal behavioral patterns. However, cameras are regarded with caution in terms of coping with user privacy concerns. In this case, postprocessing algorithms are required to solve privacy issues.[38]

There are other efforts to demonstrate the usage of visual sensors in simple scenarios. The authors in Ref. 53 constructed a camera sensor network for abnormal behavior detection in outdoor environments for short sequences (500 and 236 frames) with a video resolution of 320×240 pixels. Downes et al.[54] designed an integrated mote for wireless sensor networks where the cameras are a combination of medium resolution (CIF) and low resolution (30×30) pixels. They demonstrated the use of their wireless sensor network with a single-sensor node, which produces images of 30×30 pixels to count pedestrians passing a walkway. Rowe et al.[55] presented FireFly Mosaic, a wireless image sensor network system that has been deployed in an apartment for activity analysis with an image size of 352×288 pixels. Instead of using low-resolution cameras to analyze the occupancy map and to track people, Grünwedel et al.[56] resized images captured by high-resolution cameras.

The proposed low-resolution visual sensor network has shown promising results in the application of AAL. In Ref. 57, the authors proposed a novel measure to find similar patterns of behavior between each pair of days from the users' detected positions, based on heatmaps and Earth Movers Distance. Then an exemplar-based approach is used to identify sleeping, eating, sitting activities, and walking patterns. They used a dataset of 14 days. Xie et al.[21] analyzed the behavior patterns of an elderly person using statistical features extracted from the senior citizen tracks, such as the time of getting up and going to bed, the walking distance over a day, and the number of tracks detected in a specific area. Then the statistical features are clustered using a random sample consensus principle method to detect the behavior patterns. Eldib et al.[20] measured the socialization level of a senior citizen by detecting visits via HMM. In Ref. 58, a video-based approach has been proposed to detect sleep duration and quality among older adults. The authors analyzed sleep patterns and nightly bathroom visits indirectly to recognize sleep disorders.

Our approach of building an in-home monitoring system is different from the work in Refs. 38, 40, and 41. In the aforementioned papers, camera calibration is a prerequisite step to track the users' locations for the ADL parameters analysis. By contrast, we perform behavior analysis in an uncalibrated environment. In addition, there are more attempts to apply unsupervised approaches for activity recognition using PIR sensors than cameras.[59–62] Also, cameras with regular imaging resolutions often raise privacy concerns and increase the cost of the sensor network. We solve these
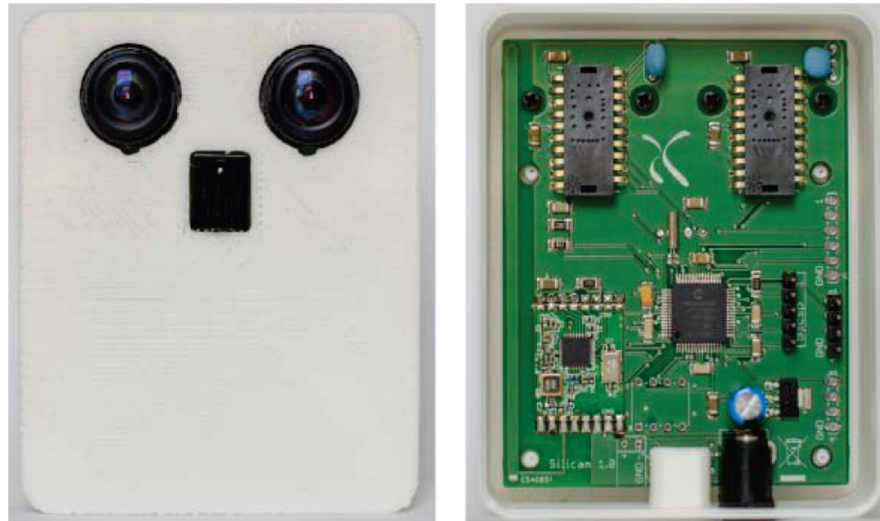
Fig. 1 The camera consists of stereo pair of image sensors controlled by a digital signal controller. Each image sensor delivers an image with a resolution of 30×30 pixels.

problems by using low-resolution cameras. Finally, in Refs. 13, 35, 36, 53, and 55, the authors performed behavior analysis on small datasets (several weeks) and the datasets were captured in lab environments. On the contrary, we perform long-term behavior analysis on 10 months of real-data recordings in a real-service flat setup.

3 Service Flat Setup

The in-home monitoring system deployed in the service flat is composed of 10 cameras,[18] as shown in Fig. 1. Each camera includes a stereo pair of visual sensors producing images of 30×30 pixels and a digital signal controller. The visual sensor images often suffer from artifacts due to read-out problems such as electrical interference, and it does not have built-in processing capabilities, such as lens shading correction resulting in a reduction of the image's brightness (vignetting). This can be solved by performing devignetting on the digital signal controller.

The cameras consist of two Agilent ADNS-3060 high-performance optical mouse sensors. These sensors are used in gaming applications. Camilli and Kleihorst[18] used this sensor with a small adaptation to produce videos of 30×30 pixels at 100 frames/s. The sensors connect over a serial peripheral interface bus directly to the internal memory of the digital signal processor, which performs the video processing. In our work, each microcontroller in each sensor performs preprocessing, including devignetting (correcting for lower brightness at the periphery of the image), automatic gain control, and noise reduction.

The results in this paper are obtained from a system setup in a service flat, covering an area of 8×4.4 m^2. Figure 2 displays the living space layout with camera positions.
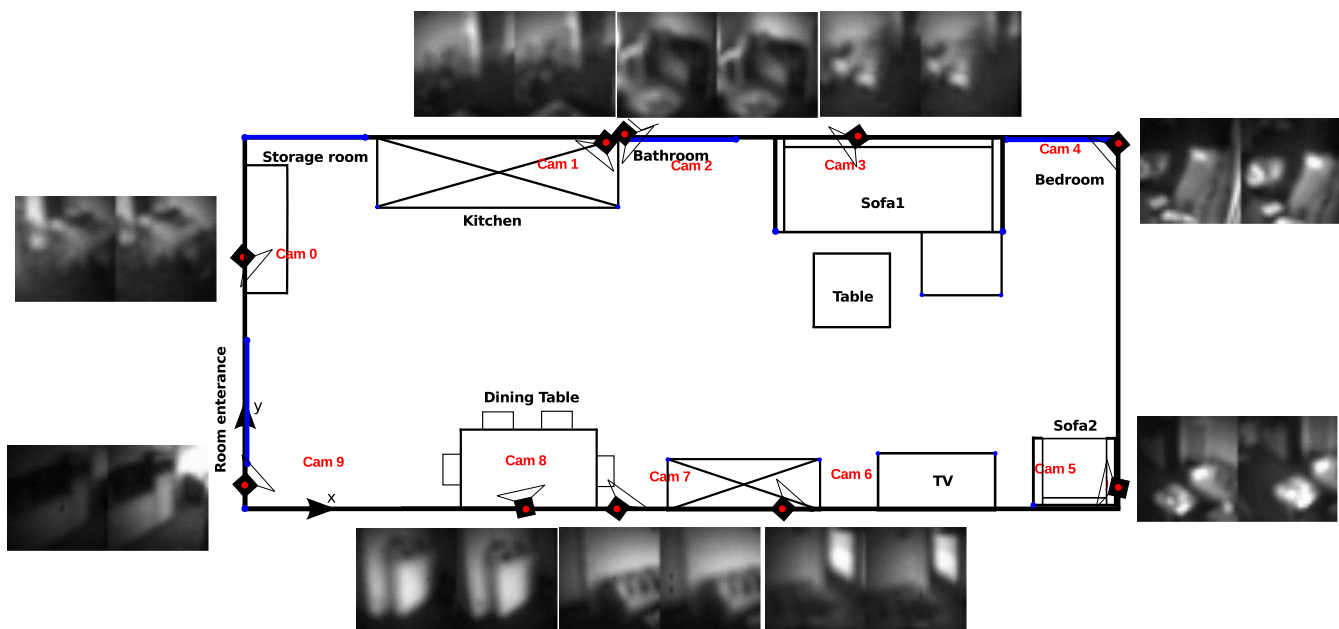

Fig. 2 Living space layout showing the configuration of 10 visual sensors (each containing stereo image sensor) covering an area of 8×4.4 m^2.

There are several challenges in the current setup:

- The collected visual sensor data includes 4 months of partial recordings (from 5 to 7 running visual sensors out of 10) and 6 months of full recordings. Our approach should make use of both types of recordings to have a continuous long-term dataset for behavior analysis.
- The visual sensors cannot show what the senior citizen is doing in the absence of sufficient lighting.
- The visual sensors are not installed in the bedroom or the bathroom for the preservation of privacy. This increases the difficulty of detecting accurate sleep durations and bathroom visits. Also, other sensing devices in other rooms are unavailable.

4 In-Home Monitoring System

Our proposed in-home monitoring systems include three processing layers: a low-level, a mid-level, and a high-level layer. In the low-level processing layer, the visual sensor features are extracted by computing the foreground pixels to track the motion level in each visual sensor. Then a simple feature vector is formed, containing the most active visual sensors at time instant t. In the mid-level processing layer, an HMM uses the feature vector as observation sequences to estimate the corresponding state sequences. The states are the different locations in the service flat. In the high-level processing layer, a rule-based approach for activity discovery utilizes the spatial context, such as location, and the temporal context, such as time duration, to infer the ADL parameters of the senior citizen. Figure 3 shows a diagram with the three processing layers of the proposed in-home monitoring system.

There are several constraints, which have led us to depict the current proposed system architecture. First, shape detection to extract the silhouette of the person for pose representation is a difficult task under low-resolution constraints, because of large variations in both pose and orientation,

poor and quickly changing lighting conditions, and the appearance of a person changes with body movements. Second, tracking algorithms require a calibrated visual sensor network and to have overlap between the cameras' FOVs. Both tasks are not practical to be performed in real-life environments, because the visual sensors can be moved accidentally. Therefore, recalibration and adjusting the cameras' FOV are needed.

4.1 Low-Level Processing Layer

In this layer, the visual sensor video capturing and preprocessing are done as in Ref. 9. We operate the visual sensor to produce images of 30×30 pixels and an image depth of 6 bit/pixel. In the preprocessing stage, a denoising step is applied by averaging the gray values of each pixel over time. The second preprocessing step is to produce a sharp image of the outside world by applying devignetting and also by correcting any pixel-dependent dark stream current in the visual sensors.

The images captured by the visual sensors suffer from noisiness and poor and quickly changing lighting conditions, which are quite prominent indoors. In a previous study,[9] several foreground/background algorithms have been tested to handle this effect. The correlation method has shown sufficient robustness to illumination changes. In this paper, we opted to use the correlation method, as shown in Fig. 4. The correlation method parameters have been tuned to produce the best visualization results and to work with the minimum lighting conditions. As a future work, we plan to study different parameter settings. Table 3 summarizes the tuned parameters.

Next, we propose a simple feature vector from which we will estimate the presence of the senior citizen location. Let $p_k(t)$ be the average number of foreground pixels in frames $t \ldots t + L - 1$ of camera k. If some of the $p_k(t)$ exceed a threshold T_p, then we consider this as a possible indication that a person is in the room. T_p is selected by computing the average number of the foreground pixels in the background image. Then we output a feature vector $x_t = (k, k')$, where k
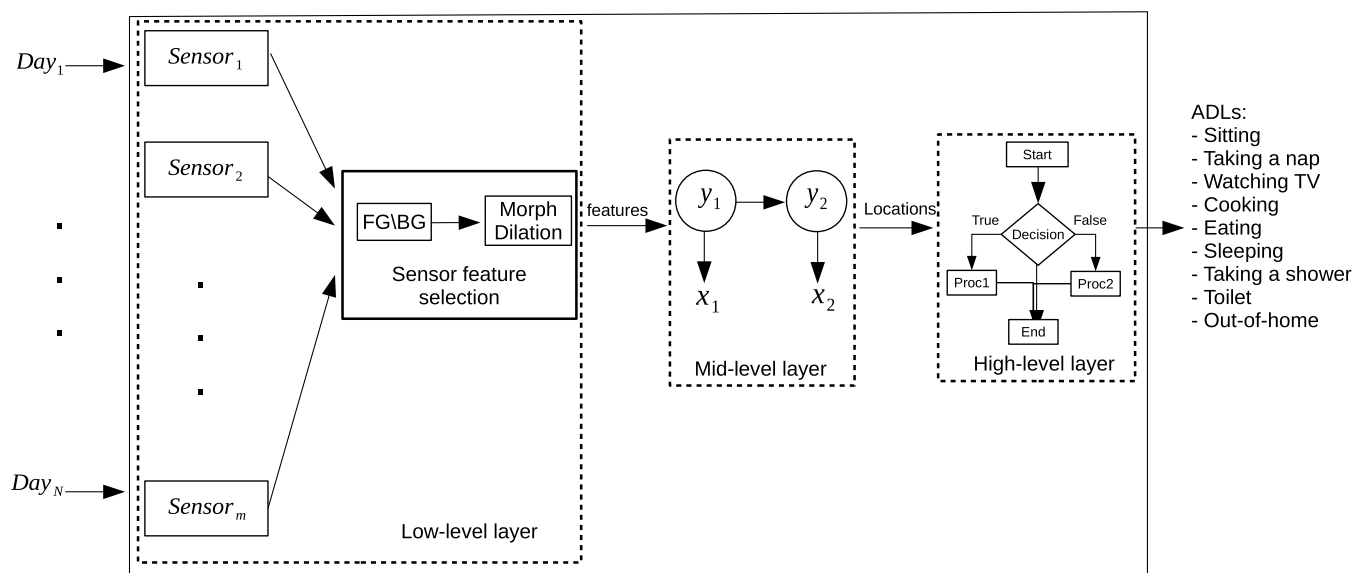

Fig. 3 System architecture of the proposed approach.
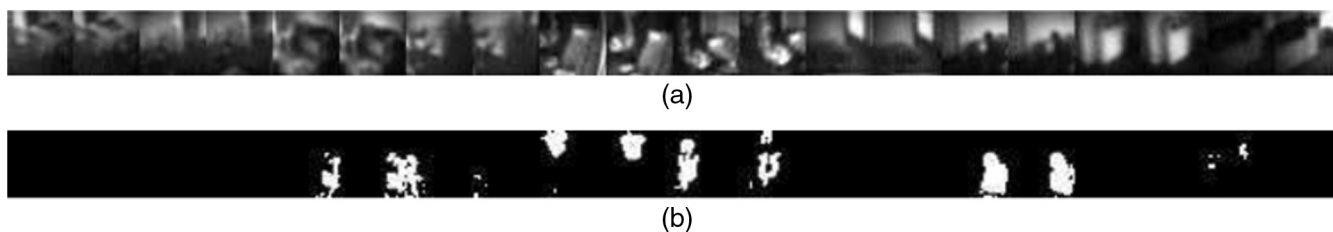
(a)

(b)

Fig. 4 (a) Original images and (b) foreground images produced by the correlation method.

Table 3 Tuned parameters of the correlation method.

Parameters	Values
Size of ω	3
ρ_{min}	0.985
α	0.005

and k' are the indices of the cameras producing the largest and second largest $p_k(t)$ at time instant t. The indices of the cameras in x_t are ordered from the largest to the smallest $p_k(t)$ value. We chose not to have more than two camera indices. In order to model the distribution of the observation vector exactly, we would need to consider all possible combinations of values in the vector dimensions. This would require n^n parameters per camera, where n is the number of features (camera index in this case), which easily results in a large number of parameters and requires accordingly large numbers of training elements.

Computing the feature vectors in all cameras causes the HMM model to suffer from a high computational complexity in the training stage. Another way is to use a supervised learning approach to train a model using all the cameras' foreground pixel percentages to estimate the senior citizen locations. For this purpose, we used the kNN classifier as a comparison to the HMM model. In the results section, we will show that the HMM approach outperforms the traditional kNN classifier when compared to ground truth.

In the following section, we will estimate senior citizen location based on a feature vector x_t, which contains all observations at a given time.

4.2 Mid-Level Processing Layer

An HMM is defined in terms of an observable measurement variable x_t and a hidden state variable y_t. These variables change with time t. In our case, the observable variable is the feature vector produced by the low-level processing layer. The hidden variable y_t is the estimated location of the senior citizen at time t. In this paper, the location is actually a discrete index, with each index value representing a possible location (e.g., a room or a part of a room).

Let $x_1, x_2, \ldots, x_T$ be the sequence observed. In the following, we will use the short-hand notation $X_{1:t}$ to denote the subsequence of observations with $t = 1, \ldots, T$. The corresponding sequence of hidden states is represented as $y_1, y_2, \ldots, y_T$, where y_t can assume one of Q possible states $1, \ldots, Q$.

Our HMM assumes that only two dependencies exist, represented by directed arrows in Fig. 5. First, the hidden variable y_t at time t statistically depends only on the previous hidden variable y_{t-1} (first order Markov assumption). Second, the observable variable x_t at time t depends only on the hidden variable y_t at the same time instant. We can, therefore, specify the HMM using three probability distributions:

- The probability of the initial states, $p(y_1)$ representing the probability that a location y occurs at the beginning of the state sequence.

- The probability of the state transition, $p(y_t|y_{t-1})$ representing the probability of switching from one state $y_{t-1} = i$ (e.g., kitchen) at time $t-1$ to another state $y_t = j$ (e.g., dining table) at the next time step, t. This represents the probability of transitions between locations.

- The probability of the observation, $p(x_t|y_t)$, indicating the probability that state y_t (e.g., Sofa 1) would
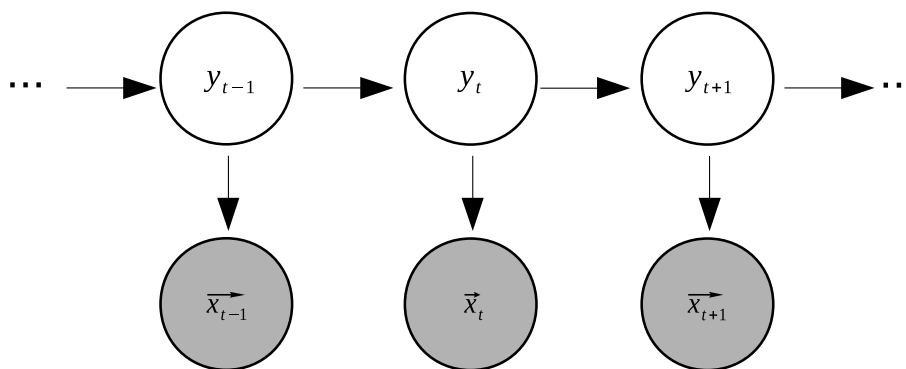

Fig. 5 A graphical representation of an HMM. The dark nodes represent observable variables, whereas the white nodes represent hidden variables.
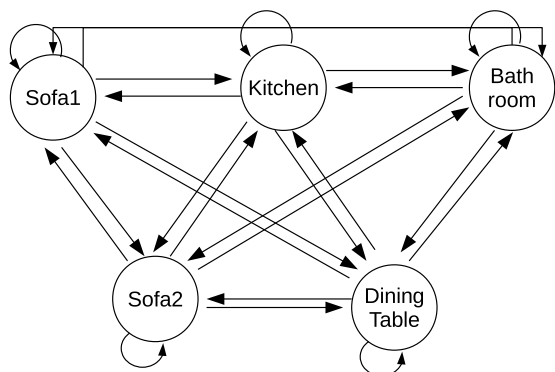
Fig. 6 A fully connected HMM configuration to recognize five distinct locations.

generate observation x_t. This represents the probability of a particular location, generating a specific associated visual sensor event.

Learning the parameters of these distributions corresponds to maximizing the joint probability of a sequence of states $Y_{1:t}, t = 1, \ldots, T$ and corresponding observations $X_{1:t}$. The joint probability of all observations and hidden states is

$$p(Y_{1:T}, X_{1:T}) = \prod_{t=1}^{T} p(x_t|y_t)p(y_t|y_{t-1}). \quad (1)$$

The inference problem consists of finding the single best state sequence (path) that maximizes $p(Y_{1:T}, X_{1:T})$. Although the number of possible paths grows exponentially with the length of the sequence, the best state sequence can be found efficiently using the Viterbi algorithm.[63] Using dynamic programming, we can discard a number of paths at each time step. This results in a computational complexity of $O(TQ^2)$ for the entire sequence. Our HMM is fully connected, as indicated in Fig. 6, where all the transitions are allowed. Finally, the HMM model is trained based on the Baum–Welch parameter estimation algorithm.[64]

4.3 High-Level Processing Layer

The commonly used approach to reason about human activities from sensor data is to identify ADL parameters that are sufficiently important and interesting to track and then model and detect occurrences of those ADLs. Modeling all of the human activities in a supervised-based approach faces a number of challenges and obstacles. First, in order to model and detect activities, a large amount of sensor data must be available. The sensor data should be labeled with the actual activities (the "ground truth" labels). In real-world in-home monitoring systems, such prelabeled data are very difficult to obtain. Second, the time that is spent on activities, which are easy to annotate (e.g., sleep times), is only a fraction of an individual's total behavioral routine. Therefore, modeling and tracking only preselected activities ignores the important insights that other activities can provide on routine behavior and activity context of the individual.

In this section, we propose a rule-based method to discover potential activity classes from unlabeled visual sensor data. We look at the duration and the location to find

activities. To achieve this, we utilize the extracted state sequence $\{y_1, y_2, \ldots, y_T\}$, that is most likely to be generated from the given observation sequence $\{x_1, x_2, \ldots, x_T\}$. This way, we interpret the meaning of the state sequence path.

Any human activity is associated with spatial and temporal contexts. Since the senior citizen locations are extracted from the previous layer, this represents the temporal context of a particular location such as being in the sofa or in the bathroom. The temporal context is the time interval between motions in a particular spatial context. The time interval between motions is defined as the time that has elapsed between two consecutive motions in a particular location.

When the senior citizen is located in the sofa, the most probable activities associated with the sofa are sitting or taking a nap. The temporal context can easily differentiate between both activities, the sitting activity contains more micromovements (the time interval between motions is low) while taking a nap activity contains less micromovements (the time interval between motions is high). If the senior citizen is inside the bathroom, the kind of the activity performed in the bathroom depends on the spent time, taking a shower requires a larger spent time than performing a toilet activity such as washing hands. We used the K-means clustering to determine the upper and the lower time interval between motion thresholds. For instance, based on the time interval between motions, there are two clusters centered around the sofa. In the first cluster, the time interval between motion tends to be high. This provides an indication of taking a nap activity. While in the second cluster, the time interval between motion tends to be low. This shows the senior citizen is active around the sofa. We follow the same approach for each location to define the appropriate thresholds. Finally, Fig. 7 shows the rules generated to discover activities related to bathroom and sofa locations.

On the other hand, eating and cooking activities are not only associated with spatial and temporal contexts, but also by the ratio of the spent time between the kitchen and the dining table. A person is considered performing a cooking activity, if the spent time in the kitchen area is higher than for the dining table. Similarly, the eating activity is detected when the spent time in the dining table is higher than for the kitchen. Sitting (e.g., using a laptop) is another activity that is associated with the dining table. The sitting activity occurs when the time interval between motions is high and with no spent time in the kitchen. The service flat exit door is close to the kitchen and the dining table (see Fig. 2), the out-of-home activity is found by detecting the last time an individual is seen at one of the two locations and after that he is not
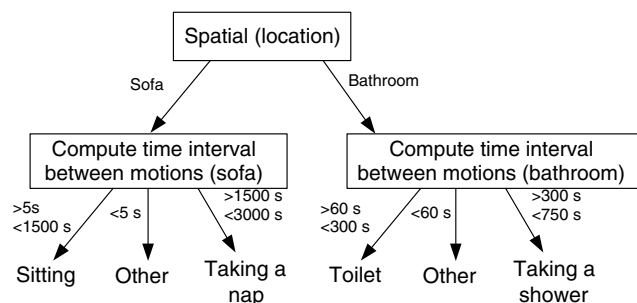

Fig. 7 A set of rules to identify bathroom and sofa activities as sitting, taking a nap, toilet, taking a shower, or other.
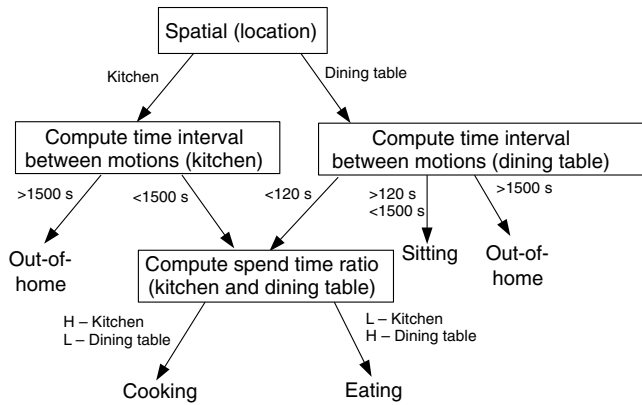
Fig. 8 A set of rules to identify kitchen and dining table activities as out-of-home, sitting, eating, or cooking. L and H stand for low and high micromovements, respectively.

seen for a long time. Similarly, the K-mean clustering is used to define the upper and the lower time intervals between motion thresholds. Figure 8 shows the rules generated to discover activities related to kitchen and dining table locations.

The rule-based approach can be used to discover repeating occurrences. Our rule-based approach is able to discover 13 activities. Often, the rule-based approach does not lead to results with 100% certainty. For example, if the time interval between motions in the kitchen is very short, this could typically mean that the senior citizen is cooking, but it is also possible that he is cleaning the kitchen. The rate of false-positive detection can be reduced by the use of additional appliance sensors.

5 Experiments

5.1 Dataset

For validating the performance of our proposed framework, we collected 10 months of real-life recordings using a network of 10 low-resolution visual sensors producing images of 30×30 pixels at a frame rate of 50 fps. Video capturing is time synchronized. Figure 9 shows an overview of the number of running visual sensors in the dataset. The minimum number of running visual sensors is 5 and the maximum

number is 10. The dataset includes 60% of 8 to 10 running visual sensors (162 days) and 40% of 5 to 7 running visual sensors (96 days). The number of running visual sensors has some impact on the results, as shown in Sec. 5.2.

The ground truth is collected from the diaries. In the diaries, the senior citizen wrote down some of his activities such as sleep time, wake up time, the start and the end time of each time being out-of-home. The diaries are verified by an informal caregiver (e.g., family member). The diaries are missing information about the ADL of the senior citizen, such as the amount of time the senior citizen spent taking a nap, sitting, cooking, eating, taking a shower, watching a TV, and so on. We performed a visual inspection of the videos in order to collect ground truth for some of the senior citizen activities. Also, the data interpreted has been demonstrated to some caregivers in the project meetings and workshops.[65] The caregivers were excited about what is possible from the visual sensor data analysis.

5.2 Location Analysis

We compare the performance of the proposed HMM described in Sec. 4.2 for estimating the senior citizen location with a kNN classifier against ground truth. In the kNN classifier, a data vector of size m' is constructed, where m' is the number of the visual sensors ($m' = 10$). The data vector holds 10 floating numbers, where each number represents the foreground pixel percentage of the visual sensor. For training the kNN model, 3 days of video recordings were annotated. For each second, we labeled the senior citizen location with the corresponding data vector. There are five classes representing the locations in the service flat (Sofa 1, Sofa 2, kitchen, dining table, and bathroom). Finally, the kNN classifier is trained with $k = 5$. Other classifiers have been tested, but kNN has found to give the best performance among them.

The comparison is done by computing the time the senior citizen spent in five locations in the service flat. The ground truth is collected by visually inspecting the video recordings and computing an approximation of the time the senior citizen spent in each location. For testing the performance of the HMM approach and the kNN classifier against ground truth, 10% of the dataset, which corresponds to 30 days, was
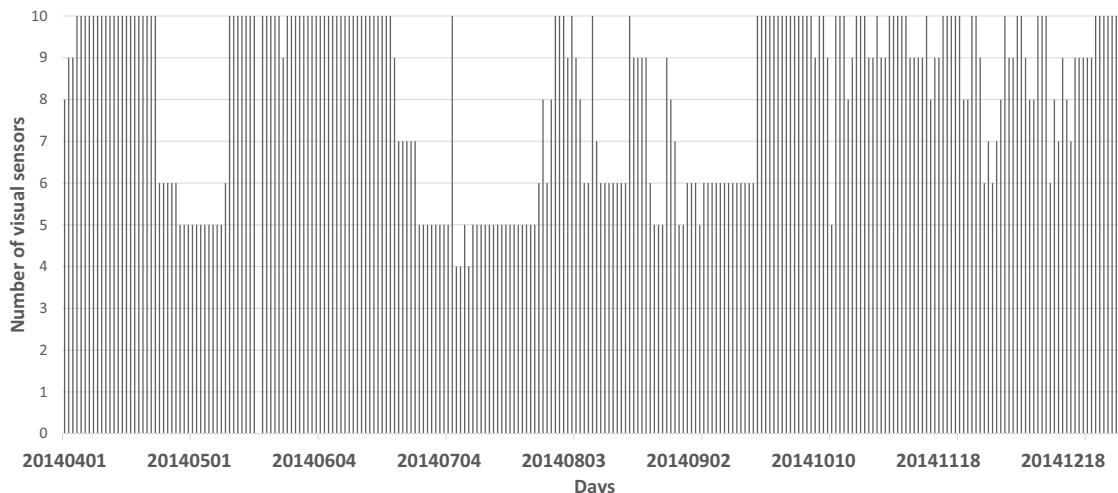

Fig. 9 Dataset analysis in terms of the number of running visual sensors.
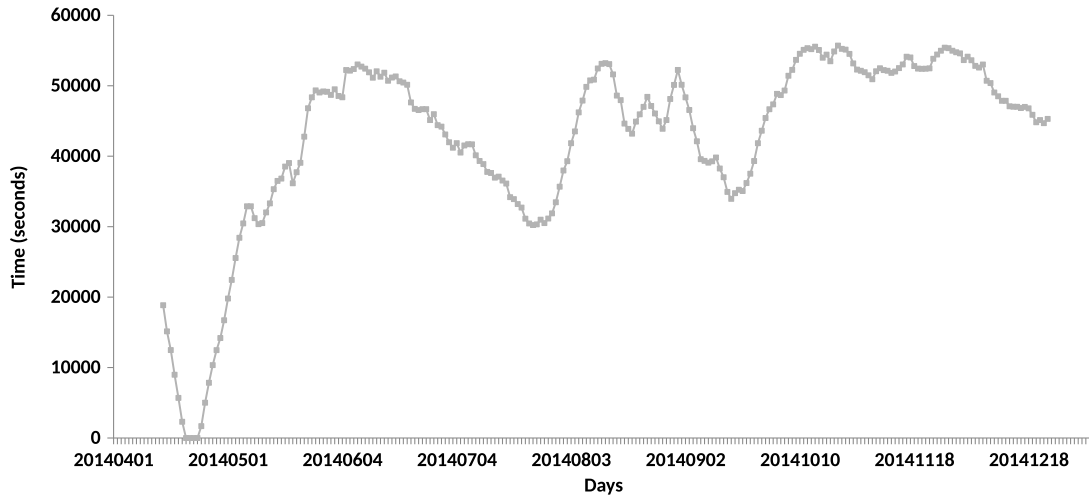
Fig. 10 The overall estimated time spent after applying the moving average over an interval of 14 days.

selected for the evaluation. We used the moving running average to smooth out short-term fluctuations in the time spent in all locations (average time spent in the kitchen, with the time spent in the Sofa 1, with the time spent in Sofa 2, with the time spent in the dining table, and with the time spent in the bathroom), in order to choose days with interesting results for the ground truth comparison.

Figure 10 shows the estimated time spent after applying the moving average filter over an interval of 14 days. For the ground truth comparison, we selected days with high and low peaks to verify the estimated time spent in each location with different numbers of running visual sensors. Figure 11 shows the comparison between the ground truth and the estimated time spent. In Fig. 11(a), the estimated times spent of HMM for Sofa 1, bathroom, and kitchen are 4.8, 0.73, and 1.35 h, which are better than the kNN approach (3.92, 0.98, and 2.65 h) when compared to the ground truth (5.33, 0.83, and 1.33 h). The kNN approach has better estimated time spent for Sofa 2 (2.77 h) than HMM (2.27 h) when compared to ground truth (2.66 h). Finally, both kNN and HMM approaches have similar estimated time spent for dining table. In Fig. 11(b), the estimated times spent of HMM for Sofa 1 and dining table are 8.60 and 0.87 h, which are significantly less than the estimated times spent of the kNN classifier (9.65 and 2.29 h) when compared to ground truth (7.5 and 1.03 h). The HMM and kNN have similar estimated times spent for Sofa 2, bathroom, and kitchen.

To further analyze the performance of our approach with kNN against ground truth, we use the mean absolute error (MAE):

$$\text{MAE} = \sum_{r=1}^{N} \frac{|v_r - v_r'|}{N}, \qquad (2)$$

where v_r is the estimated time spent for day r, v_r' is the actual time spent for day r, and N is the number of ground truth days. The relative absolute error (RAE) is computed to measure the error percentage:

$$\text{RAE} = \frac{\sum_{r=1}^{N} \frac{|v_r - v_r'|}{v_r'}}{N} \times 100. \qquad (3)$$

Finally, we measure the Spearman's rank correlation coefficient (ρ) to assess the relationship between the estimated time spent and the ground truth. The MAE, the Spearman's correlation coefficient, and the RAE results of the kNN and HMM are shown in Table 4 for the different locations. The overall MAE of HMM is 17.34 min, while the MAE of the kNN classifier is 29.34 min. Similarly, the RAE of HMM is 19.73%, while the RAE of the kNN classifier is 66.01%. The Spearman's coefficients show that the correlation between the estimated time spent of our approach and the ground truth is higher than that for the kNN classifier. Clearly, our approach using HMM outperforms the kNN classifier in accuracy with MAE, ρ and RAE.

The MAE of HMM for Sofa 1 and dining table (29.82 and 26.59 min) are considerably higher than for Sofa 2, kitchen, and bathroom (7.62, 13.83, and 8.83 min). The number of running visual sensors has an impact on the HMM performance, we found that there were more missing running visual sensors in the dataset around Sofa 1 and dining table than for Sofa 2, kitchen, and bathroom. The RAE and ρ of our approach for Sofa 1, Sofa 2, and kitchen are better than those for dining table and bathroom. We did a visual inspection of some of the ground truth videos, the visual inspection has shown that the senior citizen tends sometimes to sit on the dining table chair or go to the bathroom in very low lighting conditions. This results in difficulty for the visual sensors to detect the presence of the senior citizen.

The estimated time spent in each location by our approach is sufficiently reliable enough to perform behavior analysis on.

5.3 Activity of Daily Living Analysis

An activity is associated with location and duration contexts. Based on these two contexts, we identified 13 ADL parameters. The ADL parameters are cooking (kitchen), eating (dining table), sitting (Sofa 1, Sofa 2, and dining table), taking a nap (Sofa 1 and Sofa 2), and watching TV which is detected by computing the intensity values of the TV's region of interest. If the intensity precedes a threshold, then the TV is on. Otherwise, the TV is off (Sofa 1), taking a shower (bathroom), toilet (bathroom), being out-of-home, going to sleep and waking up. In order to estimate the
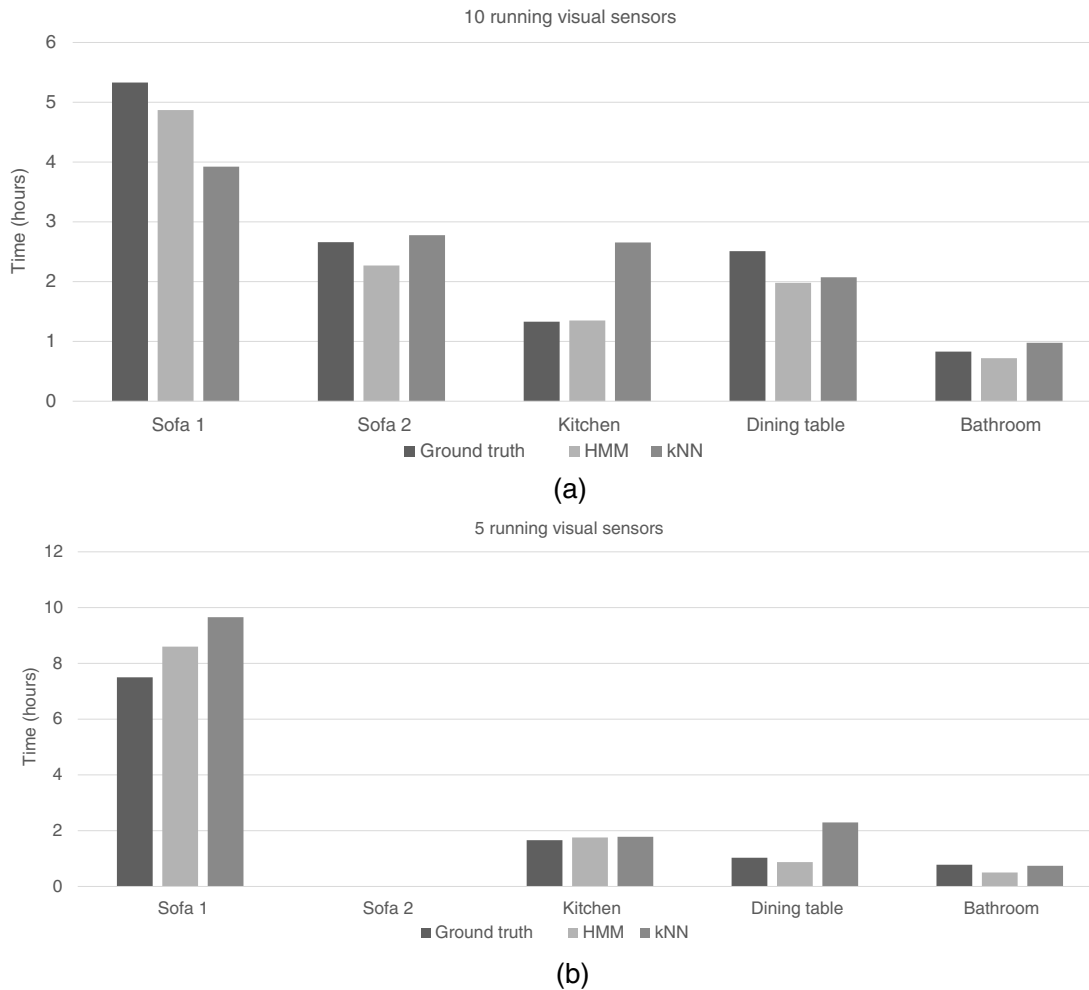
Fig. 11 A comparison between the estimated time spent of the HMM approach and kNN classifier against ground truth: (a) 10 running visual sensors and (b) 5 running visual sensors.

Table 4 Results for HMM and the kNN classifier. This table shows the MAE, the Spearman's rank correlation coefficient (ρ), and the RAE for spent times in Sofa 1, Sofa 2, kitchen, dining table, and bathroom.

Location	HMM			kNN		
	MAE (minute)	RAE (%)	ρ	MAE (min)	RAE (%)	ρ
Sofa 1	29.82	8.32	0.947	43.37	12.00	0.905
Sofa 2	7.62	21.73	0.963	11.93	150.23	0.906
Kitchen	13.83	18.54	0.942	34.82	47.17	0.665
Dining table	26.59	25.41	0.883	29.81	29.21	0.870
Bathroom	8.83	24.63	0.878	26.72	91.40	−0.069

accuracy of our activity discovery approach described in Sec. 4.3, we collected ground truth of 6 days by inspecting the videos visually to check some of the findings. In the ground truth, a set of criteria are defined to extract the ground truth activities. For example, the senior citizen is said to be cooking if he was using the oven, the cupboard, or the fridge. Table 5 describes the chosen criteria for each ground truth

activity. Finally, we compute an approximation of the time the senior citizen takes to perform an activity per day.

Table 6 shows the MAE and the RAE of the rule-based activity discovery approach against ground truth. The Spearman's rank correlation coefficient is not computed for this comparison because the sample size is too small to allow a reliable calculation. The overall MAE of the activity

Table 5 A set of chosen criteria for collecting ground truth activities.

Activity	Criteria
Sit	Reading, using a laptop, or relaxing
Take a nap	Lying on the sofa
Watch TV	TV is on with sitting activity
Cooking	Cupboard, oven, or fridge in use
Eating	Preparing the dining table
Take a shower	Spend more than 5 min in the bathroom
Toilet	Spend less than 5 min in the bathroom

Table 6 Results for the rule-based activity discovery approach. This table shows the MAE and the RAE for spent times of several activities.

Activity	MAE (min)	RAE (%)
Sit—Sofa 1	19.26	9.94
Take a nap—Sofa 1	6.32	11.73
Watch TV—Sofa 1	13.40	3.02
Sit—Sofa 2	8.74	27.94
Take a nap—Sofa 2	1.30	3.41
Eat—Dining table	12.54	29.41
Cook—Kitchen	7.64	27.39
Sit—Dining table	16.45	18.03
Take a shower—Bathroom	1.08	6.05
Toilet—Bathroom	7.08	45.72

discovery approach is 9.39 min, while the overall RAE is 18.27%. The activity discovery approach is reliable enough to find repeated activity patterns. But, some of the activities have more reliable accuracy than others. According to the MAE, sitting activity on Sofa 1 and dining table has the highest MAE, but their RAE is not the highest. Toilet (45.72%), cooking (27.39%), eating (29.41%), and sitting on Sofa 2 (27.94%) have the highest RAE values. This can be attributed to the very low-resolution of the cameras and the associated limitations in image processing and low lighting conditions.

Figure 12 shows the percentages of the different ADL parameters based on our activity discovery approach and the ground truth for a single day. In this particular day, the senior citizen spent around 50% of his time watching TV and sleeping. He spent more time at the dining table (11.89%) than at the sofa (7.58%). After checking the videos, we noticed that the senior citizen was using his laptop at the dining table. The senior citizen did not take any naps. However, he had regular eating, cooking, toilet, and taking a shower activities. Our analysis from the estimated results agrees with the ground truth. The caregiver is interested in detecting changes in the senior citizen behavior. This cannot be shown from day to day ADL parameters reports.

We generated a monthly ADL parameters report based on the rule-based activity discovery approach, so the caregiver can compare between the percentages of the ADL parameters to find any behavioral changes. The following results have been confirmed by checking the diaries and the videos. Figure 13 shows the ADL parameters of 4 months during the summer (May and June) and the winter (November and December). Sitting in Sofa 1, watching TV, and being out-of-home have noticeable behavioral changes. The senior citizen tended to sit more in May (21.02%) and in June (24.87%) than in November (14.45%) and in December (11.92%). He watched more TV in November (27.41%) and in December (22.12%) than in May (7.33%) and in June (7.50%). He was out-of-home for longer periods of time in December (19.81%). According to diaries, the senior citizen was hospitalized. Sleep duration almost remained constant with over 20%. The rest of activities did not show significant behavioral changes.

We performed more detailed analysis on activities with changes and near-constant activities such as sleeping, watching TV, and sitting on Sofa 1. The analysis also includes walking mobility patterns. Our analysis aims at detecting health deterioration or improvement after the hospitalization period. According to diaries, the senior citizen
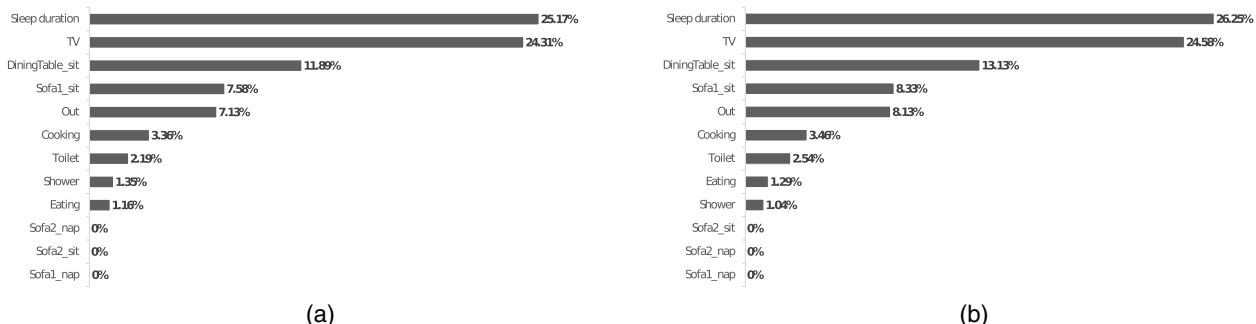

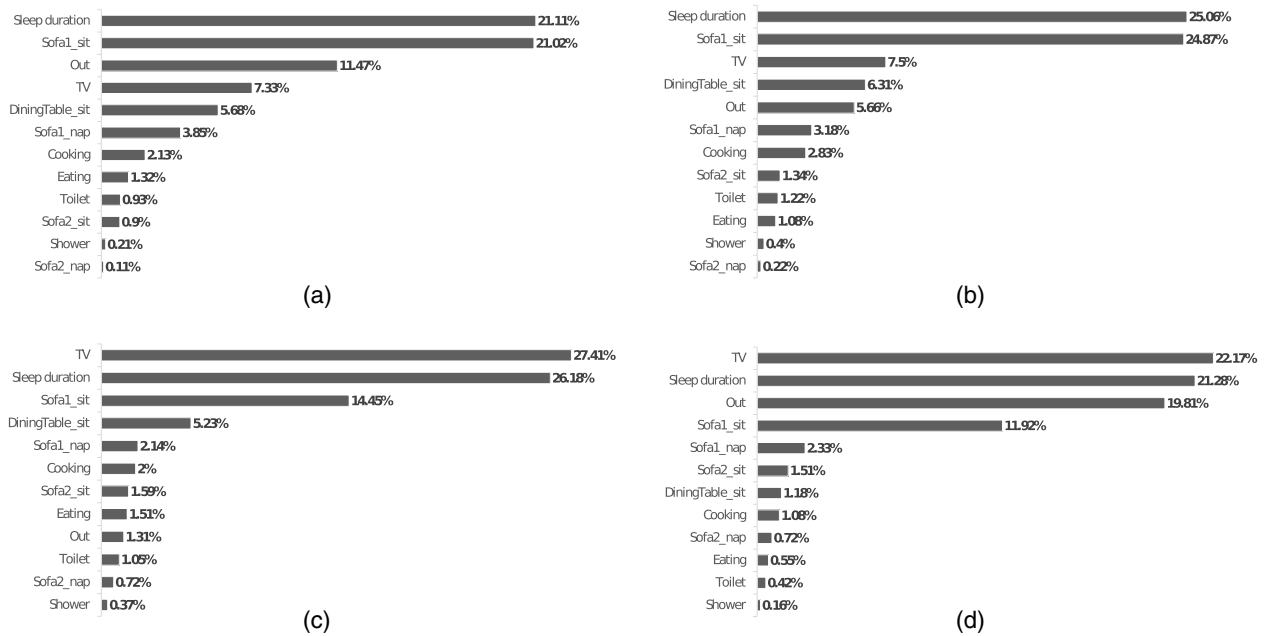**Fig. 12** ADL of a single day: (a) estimated and (b) ground truth.

Fig. 13 ADL of 4 months: (a) May, (b) June, (c) November, and (d) December.

was hospitalized in April, December, and January for 18, 8, and 9 days, respectively. Figure 14 shows activities of sitting, taking a nap and watching TV on Sofa 1 from April to December. After the hospitalization period in April, the senior citizen was in a recovery process proven by high sitting times during May and June. The sitting time decreased after June, indicating health improvement. The senior citizen preferred to watch more TV in the winter than in the summer. For taking a nap activity, the senior citizen took more naps in the summer than in the winter. This analysis has been discussed and confirmed with caregivers in project meetings.

5.3.1 Mobility pattern analysis

We analyze the number of trips and the average walking time per trip for each pair of locations. We did a visual inspection of the videos for 3 days to collect ground truth. We computed an approximation of the number of trips and the average walking time per trip. Table 7 shows the MAE and RAE between estimated results and ground truth. The overall MAE and RAE of the number of trips between locations

is 4.2 (trips) and 16.67%. The overall MAE and RAE of the average walking time are 2.51 (s) and 20.99%, respectively. Sofa 1 and bathroom have the highest MAE and RAE when compared to other locations. This is attributed to the very close distance between the bathroom and Sofa 1 (see Fig. 2). The estimated results are reliable enough to see a general trend of the walking mobility patterns of the senior citizen.

One way to detect health deterioration or improvement after the hospitalization period is by analyzing mobility patterns over longer periods. We computed the number of trips and the average walking time per trip between locations over three periods: April to June, July to September, and October to December. An informal caregiver (e.g., family member), who used to visit the senior citizen three to four times per week, confirmed our analysis. Table 8 shows the average walking time per trip between kitchen, Sofa 1, Sofa 2, and dining table locations. The senior citizen needed more time to perform a trip after the hospitalization period in April to June. This shows that the walking speed was slow. The walking time to perform trips between locations
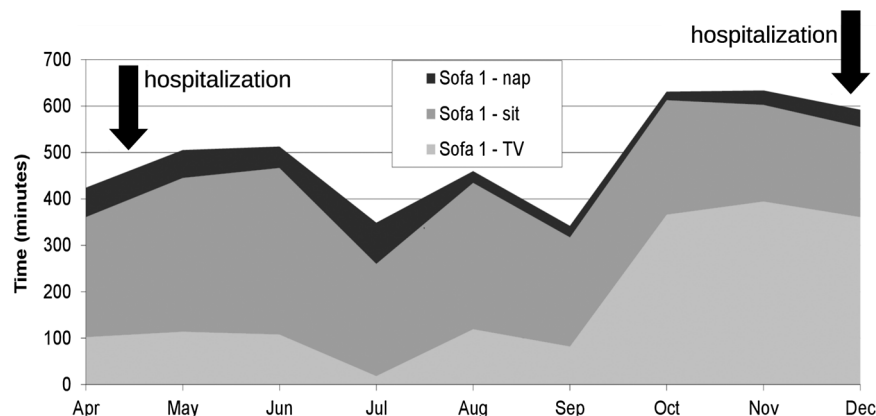

Fig. 14 The ADL parameters performed on Sofa 1 from April to December.

Table 7 Results for the number of trips and the average walking time per day for each pair of locations. This table shows the MAE and the RAE between estimated results and ground truth.

Location 1—Location 2	Number of trips		Average walking time	
	MAE (#trips)	RAE (%)	MAE (s)	RAE (%)
Sofa 1—Sofa 2	6	12.58	2.16	21.61
Sofa 1—Kitchen	7.33	29.74	3.95	39.55
Sofa 1—Dining table	5.66	17.04	1.45	13.29
Sofa 1—Bathroom	9	15.30	1.72	18.76
Sofa 2—Kitchen	1.33	33.33	1.10	2.16
Sofa 2—Dining table	0	0	0	0
Sofa 2—Bathroom	0	0	0	0
Kitchen—Dining table	8.33	12.51	2.55	19.51
Kitchen—Bathroom	2.66	25.15	8.04	59.33
Dining table—Bathroom	1.66	21.11	4.15	35.72

Table 8 The walking time per trip for three periods: April to June, July to September, and October to December.

Location 1—Location 2	April to June	July to September	October to December
Sofa 1—Sofa 2	52.85	20.43	12.74
Sofa 1—Kitchen	39.69	27.91	15.23
Sofa 1—Dining table	26.18	17.23	13.49
Kitchen—Dining table	30.52	11.64	13.11

subsequently improved in July to September. This indicates an improvement in the walking speed. Finally, the user recovered his normal walking speed in October to December because the walking time per trip was the lowest in this period. Also, Fig. 15 shows the average walking time per trip and per month to go from the kitchen to the dining table, from the bathroom to the dining table, and from Sofa 1 to the dining table. After the hospitalization, the average walking time to perform a trip was high from April to July. After July, the user needed less time to walk between locations (10 to 15 s). By asking the caregivers, they confirmed the walking time per trip of the senior citizen was improving from their observations.

Finally, Table 9 shows the number of trips to go from Sofa 1 to the dining table, from Sofa 1 to kitchen, and from kitchen to dining table. The number of trips is noticeably decreasing between kitchen to Sofa 1 and kitchen to dining table. We checked some of the video over the three periods. From the visual inspection, the senior citizen used to eat less at the dining table. Also, his cooking activity decreased as well. He preferred to eat his meals while watching TV. Also, Fig. 14 confirms an increasing TV activity in October to December. This shows a change in his behavior. The number of trips did not have a significant change between Sofa 1 and dining table. The visual analysis of the videos showed that the senior citizen used to visit the dining table either to do his daily exercises or to use his laptop.

5.3.2 Out-of-home analysis

We compare our approach of detecting the number of times the senior citizen has been out-of-home to ground truth. Out-of-home is defined as the number of times a person leaves his own home for some period per day. The ground truth is collected from the diaries. The ground truth covers 112 days (May, June, October, November, and December). The hospitalization period has not been taken in the analysis. Table 10 shows the confusion matrix for out-of-home activity. The precision is 87.5% and recall is 70.0%. The overall accuracy is 78.38%. The accuracy can be higher when other sensors are used, such as a door sensor.

Finally, the number of times the senior citizen has been out-of-home per month is shown in Fig. 16. In October to December period, the number of times the senior citizen has been out-of-home is less than the other months. This

Fig. 15 The average walking time per trip and per month for dining table.

Table 9 The number of trips per day for three periods: April to June, July to September, and October to December.

Location 1—Location 2	April to June	July to September	October to December
Kitchen—Sofa 1	36	46	17
Kitchen—Dining table	80	62	54
Sofa 1—Dining table	34	31	37

Table 10 Confusion matrix of out-of-home activity.

		Prediction	
		Yes	No
Truth	Yes	42	6
	No	18	45

Fig. 17 A comparison between sleep duration estimates and ground truth: (a) May and June; (b) October, November, and December.

indicates that the senior citizen stayed at home because it is cold and dark to be out-of-home.

5.3.3 Sleep duration analysis

The wake up time is detected when the senior citizen produces sufficient movement in the service flat in the morning. The movement should last for several minutes (e.g., more than γ min) to indicate the senior citizen has actually woken up. In case of waking up in the middle of the night to go to the bathroom, the senior citizen does not put any lights on until he reaches the bathroom. Then he turns the bathroom lights on. Once he finished visiting the bathroom in the middle of the night, he turns the bathroom lights off again and goes back to the bedroom. The wake up time in the middle of the night is not counted when this pattern occurs. This way the actual wake up time is not being confused with a wake-up call for a nightly bathroom visit. While the time of going to bed is detected by considering the last movement the senior citizen produces (e.g., turning off the TV at night).

We compare the sleep duration estimates against ground truth. The ground truth is collected from diaries in which senior citizen wrote down when he went to bed and woke up. Figure 17 compares the estimates and the ground truth of sleep duration for two different periods. The vertical error bars in Fig. 17 show the overestimates and the underestimates of sleep durations. About 20% of the cases are
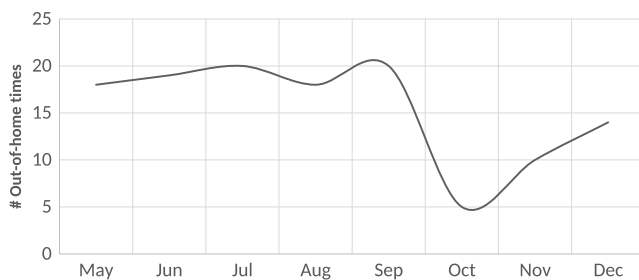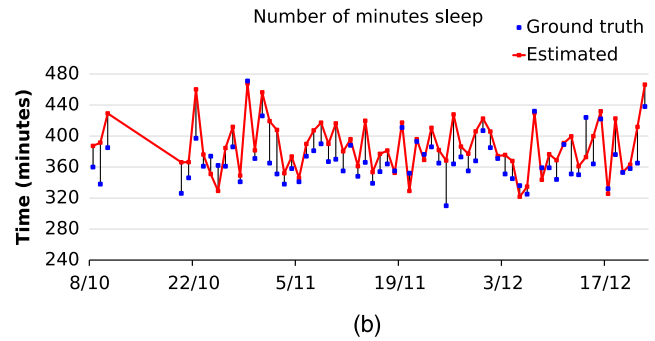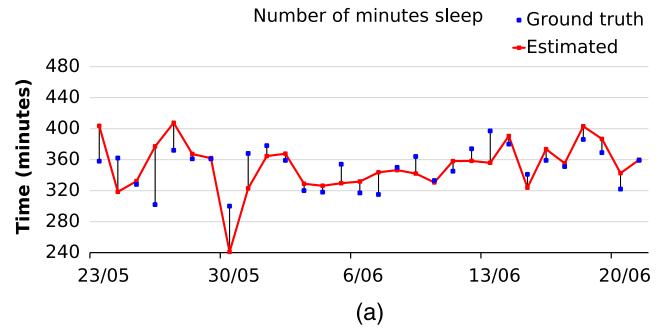
overestimated and 6% of the cases are underestimated by more than 30 min. From the visual inspection, the main cause for the overestimates is that the elderly does not turn the lights on after waking up. Finally, the MAE of the sleep duration estimates is 22.91 min. Despite waking up without turning the lights on, our approach of estimating the sleep duration provides promising results close to the ground truth. The accuracy can be increased by using other sensors inside the bed room, such as PIR sensors and thermopiles. On average, the user sleeps 6 h.

6 Conclusions

In this paper, we presented a network of low-resolution visual sensors (30×30 pixels) installed in a service flat of a senior citizen as an alternative to PIR sensors and high-resolution cameras. We proposed a framework for estimating the locations and performing behavior analysis under low-resolution constraints. Our framework is composed of three processing layers: in the low-level processing layer, the motion level in each visual sensor is computed to form a feature vector. In the mid-level processing layer, an HMM is employed to estimate the senior citizen locations without calibration. Finally, an approach for activity discovery is proposed to identify 13 ADL parameters based on spatial and temporal contexts.

We collected 10 months of real-life video recordings. First, we compared our approach of estimating the senior citizen locations based on HMM with a kNN classifier against ground truth for 30 days. The comparison to ground truth has shown that the HMM outperforms the traditional kNN classifier. Second, we evaluated the activity discovery approach against ground truth of 6 days. The results showed reliable extraction of the senior citizen activities. Then we analyzed some of the ADL parameters of the senior citizen during different months in the summer and in the winter. The

Fig. 16 The number of times being out-of-home per month.

results showed several behavioral changes in watching TV, taking a nap, and sitting on Sofa 1. Third, we computed the walking mobility patterns based on the number of trips per day and the average walking time per trip to show health improvement after hospitalization periods. Also, we compared the walking mobility patterns results against ground truth of 3 days. The estimated results were reliable enough to show a general trend of the senior citizen walking mobility patterns. Finally, our approach of detecting out-of-home activity has achieved a precision of 87.5% and a recall of 70.0%. Fourth, we compared the sleep duration estimates against 5 months of ground truth. The sleep duration estimates achieved MAE of 22.91 min.

The use of cameras to monitor residents in assisted living facilities or nursing homes is not new. Traditionally, these systems rely on high-resolution cameras to make sure all relevant information is captured. This approach is not only expensive (deployment and maintenance-wise), but also it requires high bandwidths to transport and process the high-resolution video feeds, and it is often challenged from a privacy perspective, too. Moreover, existing systems are generally "not self-aware." They are not capable of monitoring (let alone interpreting) behavior or the evolution of disorders over time. We took a different approach by replacing high-resolution cameras with low-cost low-resolution visual sensors to capture what is happening in a given premises in a way that is privacy compliant and using our proposed architecture to translate the input from low-resolution video feeds into valuable information about the resident's condition. The ADL parameters were sufficient enough to show interesting facts about the senior citizen's health condition (recovering from a hospitalization period by analyzing mobility patterns). Also, the results presented in this paper were usable by the caregivers and the senior citizen. The data were interpreted in real time by visualizing the ADL parameters on a smart display. So, the caregivers can monitor the ADL parameters on a daily basis and detect any abnormal behavior.

References

1. J. op de Beke, "*Confronting demographic change: a new solidarity between the generations*," in Communication from the Commission (2005).
2. Center for Cardiovascular Education, Inc., *Common Diseases with Aging*, Parlay International, California, www.caregiverresourcecenter.com/disease.pdf (10March2016).
3. P. Bartosz, *Projecting Future Health Care Expenditure at European Level: Drivers, Methodology and Main Results*, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission (2010) https://ideas.repec.org/p/euf/ecopap/0417.html.
4. B. Rechel et al., "Ageing in the European Union," *Lancet* **381**(9874), 1312–1322 (2013).
5. European Commission (DG ECFIN) and the Economic Policy Commitee (AWG), "The 2015 ageing report," European Commission (2015).
6. M. Murphy and P. Martikainen, "Use of hospital and long-term institutional care services in relation to proximity to death among older people in Finland," *Soc. Sci. Med.* **88**(0), 39–47 (2013).
7. R. Schulz and P. R. Sherwood, "Physical and mental health effects of family caregiving," *Am. J. Nurs.* **108**, 23–27 (2008).
8. G. Doblhammer and U. Ziegler, "Future elderly living conditions in Europe: demographic insights," in *Gender, Health and Ageing*, G. M. Backes, V. Lasch, and K. Reimann, Eds., pp. 267–292, VS Verlag fr Sozialwissenschaften, Wiesbaden (2006).
9. B. B. Nyan et al., "Human mobility monitoring in very low resolution visual sensor network," *Sensors* **14**(11), 20800–20824 (2014).
10. R. kleihorst and and P. Verhoeve, "Little sister," http://www.iminds.be/en/projects/2014/03/04/littlesister (10March2016).
11. A. Ben et al., "Social networks for older adults to promote an active life (SONOPA),"http://www.sonopa.eu/ (10March2016).
12. P. Natarajan and R. Nevatia, "Coupled hidden semi Markov models for activity recognition," in *IEEE Workshop on Motion and Video Computing*, pp. 10–10 (2007).
13. M. Alwan et al., "Validation of rule-based inference of selected independent activities of daily living," *Telemed. e-Health* **11**(5), 594–599 (2005).
14. D. Hodgins et al., "Healthy aims: developing new medical implants and diagnostic equipment," *IEEE Pervasive Comput.* **7**, 14–21 (2008).
15. U. Anliker et al., "Amon: a wearable multiparameter medical monitoring and alert system," *IEEE Trans. Inf. Technol. Biomed.* **8**, 415–427 (2004).
16. Y. Masuda et al., "An unconstrained monitoring system for home rehabilitation," *IEEE Eng. Med. Biol. Mag.* **24**(4), 43–47 (2005).
17. T. Teixeira, G. Dublon, and A. Savvides, "A survey of human-sensing: methods for detecting presence, count, location, track, and identity," ENALAB Technical Report (2010).
18. M. Camilli and R. Kleihorst, "Demo: mouse sensor networks, the smart camera," in *Fifth ACM/IEEE Int. Conf. on Distributed Smart Cameras*, pp. 1–3 (2011).
19. M. Ziefle, C. Rocker, and A. Holzinger, "Medical technology in smart homes: exploring the user's perspective on privacy, intimacy and trust," in *IEEE 35th Annual Computer Software and Applications Conf. Workshops*, pp. 410–415 (2011).
20. M. Eldib et al., "Detection of visitors in elderly care using a low-resolution visual sensor network," in *Proc. 9th ACM Int. Conf. on Distributed Smart Cameras*, Vol. 6, ACM (2015).
21. X. Xie et al., "PhD forum: analyzing behaviors patterns of the elderly from low-precision trajectories," in *Proc. of the Int. Conf. on Distributed Smart Cameras*, pp. 47:1–47:2, ACM, New York, New York (2014).
22. M. C. Mozer et al., "The neural network house: an overview," in *Current Trends in Connectionism*, pp. 371–380 (1995).
23. E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Proc. Pervasive Computing: Second Int. Conf.*, Linz/Vienna, Austria, pp. 158–175, Springer Berlin Heidelberg, Berlin, Heidelberg (2004).
24. T. S. Barger, D. E. Brown, and M. Alwan, "Health-status monitoring through analysis of behavioral patterns," *IEEE Trans. Syst. Man Cybern. Part A* **35**(1), 22–27 (2005).
25. B. Logan et al., *A Long-Term Evaluation of Sensing Modalities for Activity Recognition*, Springer, Berlin, Heidelberg (2007).
26. L. Atallah et al., "Real-time activity classification using ambient and wearable sensors," *IEEE Trans. Inf. Technol. Biomed.* **13**(6), 1031–1039 (2009).
27. M. Fahim et al., "Activity recognition based on SVM kernel fusion in smart home," in *Computer Science and its Applications*, pp. 283–290, Springer, Dordrecht (2012).
28. W. Pentney et al., "Sensor-based understanding of daily life via large-scale use of common sense," in *Proc. American Association for Artificial Intelligence*, Vol. 1, pp. 4–9 (2006).
29. S. Wang et al., "Common sense based joint training of human activity recognizers," in *Int. Joint Conf. on Artificial Intelligence*, Vol. 7, pp. 2237–2242 (2007).
30. D. L. Vail, M. M. Veloso, and J. D. Lafferty, "Conditional random fields for activity recognition," in *Proc. of the 6th Int. Joint Conf. on Autonomous Agents and Multiagent Systems*, Vol. 235, ACM (2007).
31. D. H. Hu and Q. Yang, "Cigar: concurrent and interleaving goal and activity recognition," in *Proc. American Association for Artificial Intelligence*, Vol. 8, pp. 1363–1368 (2008).
32. T. van Kasteren et al., "Accurate activity recognition in a home setting," in *Proc. of the 10th Int. Conf. on Ubiquitous Computing*, pp. 1–9, ACM (2008).
33. L. Chen et al., "Sensor-based activity recognition," *IEEE Trans. Syst. Man Cybern. Part C* **42**(6), 790–808 (2012).
34. D. J. Cook et al., "Casas: a smart home in a box," *Computer* **46**(7), 62–69 (2013).
35. F. J. Ordóñez, P. de Toledo, and A. Sanchis, "Activity recognition using hybrid generative/discriminative models on home environments using binary sensors," *Sensors* **13**(5), 5460–5477 (2013).

36. T. van Kasteren et al., "Accurate activity recognition in a home setting," in *Proc. of the 10th Int. Conf. on Ubiquitous Computing*, pp. 1–9, ACM, New York, New York (2008).

37. T. L. M. van Kasteren, G. Englebienne, and B. J. A. Kröse, "Activity recognition using semi-Markov models on real world smart home datasets," *J. Ambient Intell. Smart Environ.* **2**, 311–325 (2010).

38. A. A. Chaaraoui et al., "A vision-based system for intelligent monitoring: human behaviour analysis and privacy by context," *Sensors* **14**(5), 8895–8925 (2014).

39. M. Ahmad and S. W. Lee, "HMM-based human action recognition using multiview image sequences," in *18th Int. Conf. on Pattern Recognition*, Vol. 1, pp. 263–266 (2006).

40. P. C. Chung and C. D. Liu, "A daily behavior enabled hidden Markov model for human behavior understanding," *Pattern Recognit.* **41**(5), 1572–1580 (2008).

41. T. Duong et al., "Activity recognition and abnormality detection with the switching hidden semi-Markov model," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 838–845 (2005).

42. H. Mannila, H. Toivonen, and A. I. Verkamo, "Discovering frequent episodes in sequences extended abstract," in *Proc. the First Conf. on Knowledge Discovery and Data Mining*, pp. 210–215 (1995).

43. T. Gu et al., "An unsupervised approach to activity recognition and segmentation based on object-use fingerprints," *Data Knowl. Eng.* **69**(6), 533–544 (2010).

44. P. Rashidi and D. Cook, "Mining sensor streams for discovering human activity patterns over time," in *IEEE 10th Int. Conf. on Data Mining*, pp. 431–440 (2010).

45. P. Theekakul et al., "A rule-based approach to activity recognition," *Lect. Notes Comput. Sci.* **6746**, 204–215 (2011).

46. T. Huynh, M. Fritz, and B. Schiele, "Discovery of activity patterns using topic models," in *Proc. of the 10th Int. Conf. on Ubiquitous Computing*, pp. 10–19, ACM, New York, New York (2008).

47. F. Castanedo et al., "Learning routines over long-term sensor data using topic models," *Expert Systems* **31**(4), 365–377 (2014).

48. S. Dalal et al., "A rule-based approach to the analysis of elders' activity data: detection of health and possible emergency conditions," in *AAAI Fall 2005 Symp.* (2005).

49. F. J. Ordóñez et al., "Online activity recognition using evolving classifiers," *Expert Syst. Appl.* **40**(4), 1248–1255 (2013).

50. M. Perry et al., "Multimodal and ubiquitous computing systems: supporting independent-living older users," *IEEE Trans. Inf. Technol. Biomed.* **8**, 258–270 (2004).

51. M. Eldib et al., "A low resolution multi-camera system for person tracking," in *IEEE Int. Conf. on Image Processing*, pp. 378–382 (2014).

52. F. Cardinaux et al., "Video based technology for ambient assisted living: a review of the literature," *J. Ambient Intell. Smart Environ.* **3**, 253–269 (2011).

53. Y. Wang, D. Wang, and F. Chen, "Abnormal behavior detection using trajectory analysis in camera sensor networks," *Int. J. Distrib. Sens. Networks* **2014**, 839045 (2014).

54. I. Downes, L. B. Rad, and H. Aghajan, "Development of a mote for wireless image sensor networks," *Proc. of Cognitive Systems with Interactive Sensors* (2006).

55. A. Rowe, D. Goel, and R. Rajkumar, "Firefly mosaic: a vision-enabled wireless sensor networking system," in *28th IEEE Int. Real-Time Systems Symp.*, pp. 459–468 (2007).

56. S. Grünwedel et al., "PhD forum: multi-view occupancy maps using a network of low resolution visual sensors," in *2011 Fifth ACM/IEEE Int. Conf. on Distributed Smart Cameras*, Vol. 2, IEEE (2011).

57. M. Eldib et al., "Behavior analysis for aging-in-place using similarity heatmaps," in *Proc. 8th ACM/IEEE Int. Conf. on Distributed Smart Cameras*, Vol. 6, ACM/IEEE (2014).

58. M. Eldib et al., "Sleep analysis for elderly care using a low-resolution visual sensor network," *Lect. Notes Comput. Sci.* **9277**, 26–38 (2015).

59. Y. Wang and G. Mori, "Hidden part models for human action recognition: probabilistic versus max margin," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 1310–1323 (2011).

60. Q. Le et al., "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3361–3368 (2011).

61. M. Raptis, I. Kokkinos, and S. Soatto, "Discovering discriminative action parts from mid-level video representations," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1242–1249 (2012).

62. T. Guha and R. Ward, "Learning sparse representations for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 1576–1588 (2012).

63. L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE* **77**, 257–286 (1989).

64. P. Baggenstoss, "A modified Baum-Welch algorithm for hidden Markov models with multiple observation spaces," *IEEE Trans. Speech Audio Process.* **9**, 411–416 (2001).

65. iMinds, "AAL Forum 2015," http://www.aal-europe.eu/forum-2015/ (10 March 2016).

Mohamed Eldib received his BSc and MSc degrees in 2008 and 2011, respectively, in information technology from Cairo University, Egypt. Since 2013, he has been working toward his PhD degree at Ghent University in Belgium. His interests include smart environments, AI and machine learning applications in health care, and human factors in pervasive computing applications.

Francis Deboeverie received the Master of Science degree in electronics and ICT engineering technology from the University of Ghent in 2007. In 2014, he obtained his PhD in engineering from Ghent University, Belgium. He is currently a postdoctoral researcher at IPITELIN—iMinds group at Ghent University. His research interests include image interpretation with polynomial feature models for real-time vision systems.

Wilfried Philips received his diploma degree in electrical engineering from the University of Ghent, Belgium. Since October 1989, he has been working at the Department of Electronics and Information Systems of Ghent University, as a research assistant for the Flemish Fund for Scientific Research (FWO). In 1993, he obtained his PhD in electrical engineering from Ghent University, Belgium. His main research interests are image and video processing of multimedia data and restoration.

Hamid Aghajan obtained his PhD in electrical engineering from Stanford University in 1995. Currently, he is the director of Stanford's Wireless Sensor Networks Lab and Ambient Intelligence Research (A I R) Lab, which he established in 2003 and 2009, respectively. Focus of research in his group is on methods and applications of ambient intelligence with an emphasis on behavior modeling based on activity monitoring in smart homes.