Detection of Visitors in Elderly Care using a Low-resolution Visual Sensor Network

Mohamed Eldib[1], Francis Deboeverie[1], Dirk Van Haerenborgh[1],
Wilfried Philips[1], and Hamid Aghajan[1,2]

[1] Image Processing and Interpretation, Gent University/iMinds, Gent 9000, Belgium
[2] Ambient Intelligence Research Lab, Stanford, CA 94305, USA

ABSTRACT

Loneliness is a common condition associated with aging and comes with extreme health consequences including decline in physical and mental health, increased mortality and poor living conditions. Detecting and assisting lonely persons is therefore important-especially in the home environment. The current studies analyse the Activities of Daily Living (ADL) usually with the focus on persons living alone, e.g., to detect health deterioration. However, this type of data analysis relies on the assumption of a single person being analysed, and the ADL data analysis becomes less reliable without assessing socialization in seniors for health state assessment and intervention. In this paper, we propose a network of cheap low-resolution visual sensors for the detection of visitors. The visitor analysis starts by visual feature extraction based on foreground/background detection and morphological operations to track the motion patterns in each visual sensor. Then, we utilize the features of the visual sensors to build a Hidden Markov Model (HMM) for the actual detection. Finally, a rule-based classifier is used to compute the number and the duration of visits. We evaluate our framework on a real-life dataset of ten months. The results show a promising visit detection performance when compared to ground truth.

Keywords

Ambient assisted living (AAL), low-resolution visual sensors, visitor detection, Hidden Markov Models (HMM)

1. INTRODUCTION

It is estimated that the proportion of Europeans aged 65 years and older will increase by a percentage of 79%, namely from 84.6 million to 151 million. The population of Europeans older than 80 years is estimated to increase by a percentage of 181%, namely from 21.8 to 61.4 million [1]. The majority of these older adults want to age autonomously in their trusted home environment [10]. Shifting from institutional care to care at home is also in the interest of

policy makers, who are confronted with scarce budgets and lacking health care professionals [14]. However, age-related physical, cognitive and perceptual decline makes ageing in place challenging. Even the still healthy and active elders might need some form of assistance in the near future [9]. Informal caregivers, such as family members, often provide assistance, but caring for a family member can be stressful and overwhelming [26].

Loneliness and social isolation are probably the most prevalent psychosocial problems connected to ageing. Several factors contributing to loneliness and social isolation in later life such as: decline in physical and mental health; change of the social environment due to relocation; retirement or loss of a partner; the demand to care for a partner in poor health; the lack of transportation options and poor living conditions in general [23, 7]. To meet these challenges the concept of Ambient Assisted Living was introduced (AAL). The vision of AAL is to increase the several health benefits associated with socialization, accurate detection whether a person has visitors to their home, which is considered one of the essential aspects of elders' behaviors.

Detecting visitors is beneficial for the carers to construct a social logbook. For instance, each time one of the carers who visit an elder, his entrance is logged by the visual sensor placed at the entrance of the house. All carers receive periodically by email a report on earlier visits of caregivers when they came, and how intense was the visit (how long, how many visitors). The information is also archived in a shared agenda which will aid them to coordinate their help.

Without automatically detecting the visitors, carers such as family would have to rely on asking the elder about visitors during the week, and about their activity levels. The answers would be subjective and not very accurate. This is a well-known phenomenon even for people with top mental capabilities, let alone for possibly forgetful elders. The social logbook helps to give a measure of the elder's socialization and reassure the family.

We have been monitoring an elder user living in a service flat in Belgium for one year using a network of 10 low-resolution visual sensors [8]. Our earlier research focused on real-time tracking of people [17]. Currently we are focusing on recognising and quantifying visits on this long term dataset, e.g., by monitoring day by day changes in socialization levels and by detecting loneliness at the earliest phases. The visual sensors have a very low spatial resolution (30×30 pixels) which is a key feature to ensure a cheap monitoring solution: the low data rates allow cheap data processing and transmission. Despite the low resolution, these cam-

eras offer richer data analysis than e.g., PIR sensors which detect presence only. The focus of this paper is on using low-resolution visual sensors for objectively and unobtrusively detecting visitors, so that the carers can evaluate the socialization level of the elderly.

The main contributions of this paper are: (1) a framework for detecting visitors, formal and informal (family members and friends) caregivers. (2) The introduction and evaluation of visit detection using a network of low-resolution visual sensors which, to the best of our knowledge, has not been examined before. (3) The evaluation of this method on a real-life dataset covering ten months by comparing the results to ground truth.

The paper is organized as follows. In section 2 , we present related work. Section 3 gives an overview of the data collection using a low-resolution visual sensor network. Section 4 explains our visit detection framework, including the different layers. Section 5 describes our experimental results. Finally, Section 6 offers some conclusions.

2. RELATED WORK

The use of different sensors such as Passive Infrared Motion Sensor (PIR) [16, 2], visual sensors [17, 12] and Kinect sensors [15] became popular among researchers to analyze the Activities of Daily Living (ADL). PIR sensors detect motion and do not provide detailed information on motion patterns. Their use is mostly restricted to single-person scenarios [22]. However, there have been few attempts to detect visits using PIR sensor data. In [3], the authors presented an unsupervised method for modeling visits as an abnormal activity. They used a Markov modulated multidimensional non-homogeneous Poisson process (M3P2) to model weekly and daily variations. Petersen et al. [18] used a Support Vector Machine (SVM) to identify periods where visitors are present in the home. The model features were the dwell time, the number of sensor firings and the number of transitions between major living spaces (living room, dining room, kitchen and bathroom). They evaluated their approach on six weeks of sensor data in a living lab. In [16], an HMM is used for visitor detection using a binary sensor in a living lab. In [2], a comparison between a Naive Bayes classifier and an HMM for visitor detection is conducted on an office equipped with binary sensors for two months.

The use of multi-camera systems for visitor detection is limited due to the nature of the acquired high-resolution images and the privacy concerns of the users [24, 13]. Postprocessing on the images is required to address the privacy issue. In [20], the authors used high-resolution video sensor network to distinguish between different people in a multi-person setting, e.g., the resident vs. a visitor. They extracted the silhouettes of the persons as a means of preserving the privacy of the residents. Kuutti et al. [21] designed a home control system to identify visitors and confirm the person's identification using wired image cameras and RFID sensors.

Most of the ADL research uses data from living labs for relatively short time periods. As a result, visitor detection methods from literature are also only analyzed over short periods. Visual sensors have shown promising results in person tracking [11, 17] and behavior analysis [25, 12]. They are designed to offer an acceptable sensing mechanism for addressing the privacy concerns of the users, while the kinect sensor requires post-processing algorithms to solve the pri-

vacy concern problem [6]. In this paper, we utilize a network of low-resolution visual sensors [5] to detect visits in a long-term real-life dataset.

3. VISUAL SENSOR DATA

We have continuously collected data for more than ten months in a service flat of an elder living alone in Belgium with daily visits from caregivers and regular visits from family members. The service flat is equipped with a network of ten visual sensors covering an area of $8 \times 4.4 \ m^2$. Each visual sensor has a pair of image sensors (30×30 pixels resolution sensors used in computer mice). An overview of the location of the visual sensors in the service flat of one resident is shown in Figure 1. The elder is living his routine life and is not told to modify his behaviour in any way. The voluntary older person participating in the installation is 83 years old with diabetes and decreased mobility due to a little paralysis. However, the resident has a very clear mind. The visits were happening naturally without instructing the family members or the caregivers on the way performing the visits.

4. VISITOR DETECTION FRAMEWORK

Our visitor detection framework is composed of three processing layers. In the low-level layer, the visual sensor features are extracted by running a series of image processing operations such as foreground/background subtraction and morphological dilation to track the motion patterns in each visual sensor and to select the active visual sensors. Next, an HMM is used in the mid-level layer for visitor detection. Finally, a rule-based classifier to compute the number and the duration of each visit. Figure 2 shows the framework architecture.

4.1 Low-level processing layer

In this section we propose a simple feature vector from which we will estimate the presence of multiple people. Let $p_k(t)$ be the average number of foreground pixels in frames $t \ldots t+L-1$ of camera k. If all of the $p_k(t)$ exceed a threshold T_p, then we consider this as a possible indication that more than one person is in the room. This step is important to ensure the cameras have no overlapping field of view. We then output a feature vector $\vec{x}_t = (k, k')$ where k and k' are the indices of the cameras producing the largest and second largest $p_k(t)$ at time instant t. We selected not to have more than two camera indices for the following: First, the large dimension of the feature vector leads to high computational complexity of the training and detection in the HMM model. Second, there is at most one to two cameras allocated for each location in the service flat (e.g., camera 5 and camera 7 are allocated for the bathroom).

In the following section we will use this feature vector to detect the actual presence of visitors using an HMM.

4.2 Mid-level processing layer

We use an Hidden Markov Model (HMM) to tackle the problem of detecting visitors. An HMM is defined in terms of an observable variable $\vec{x}_t$ and a hidden variable y_t at each discrete time instant t. The observation variable is the feature vector $\vec{x}$ which holds the indices of the cameras from the low-level layer, and the hidden variable is the detection of visitors. Let $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$ be the sequence ob-
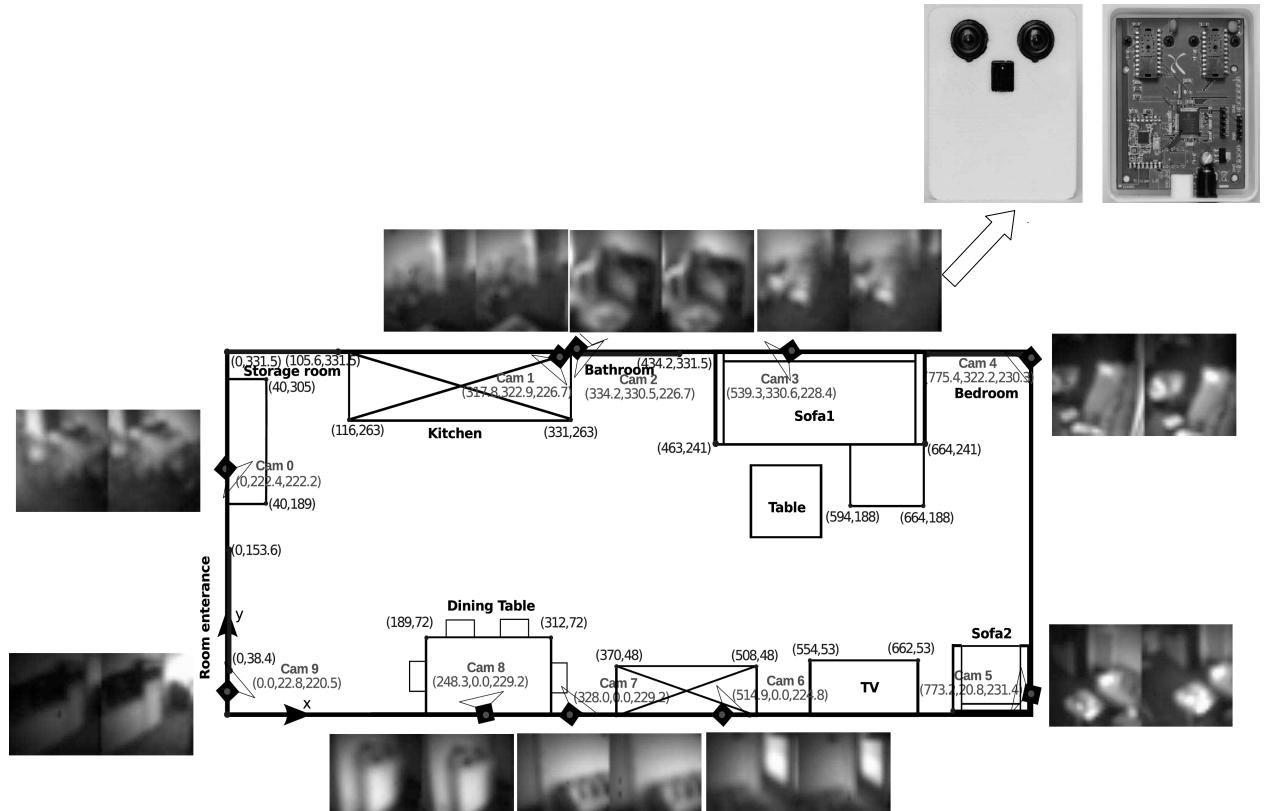
Figure 1: A layout of the service flat in Belgium equipped with a network of 10 visual sensors covering an area of 8×4.4 m^2.
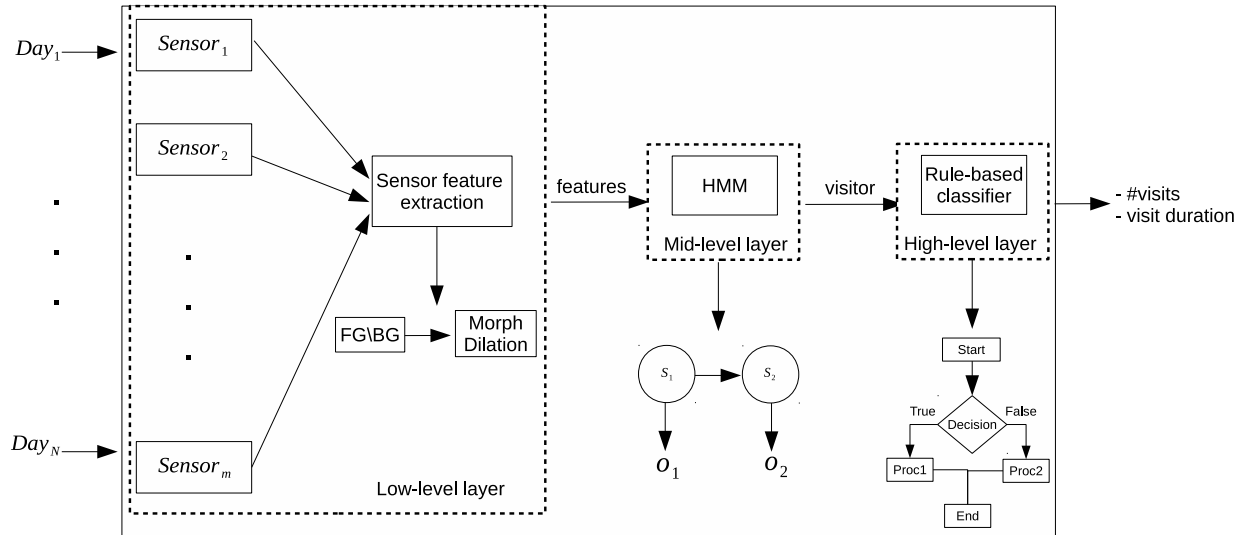

Figure 2: System architecture of the proposed visit detection framework.

served. The corresponding sequence of hidden states is represented as $Y = \{y_1, y_2, \ldots, y_T\}$ where for Q possible states $y_t \in \{1...Q\}$. The possible states are visitor and no visitor. An HMM is denoted as $\lambda = (A, B, \pi)$, where π is the initial distribution, A is the transition distribution and B is the observation distribution. Let $A = \{a_{ij}\}$ be the state transition from state $y_{t-1} = i$ to state $y_t = j$ and $B = \{b_j(\vec{x}_t)\}$ represent the probability of producing $\vec{x}_t$ at state $y_t = j$. Learning the parameters of these distributions corresponds to maximizing the joint probability of a sequence of states and observations. The three distributions described earlier can be factorized as follows:

$$p(Y_{1:T}, X_{1:T}) = \prod_{t=1}^{T} p(\vec{x}_t \mid y_t) p(y_t \mid y_{t-1}) \qquad (1)$$

Inferring which state sequences best explains a new sequence of observations can be found efficiently using the Viterbi algorithm [19]. Figure 3 shows the proposed HMM for the detection of visitors. HMM is trained based on the Baum-Welch parameter estimation algorithm [4].
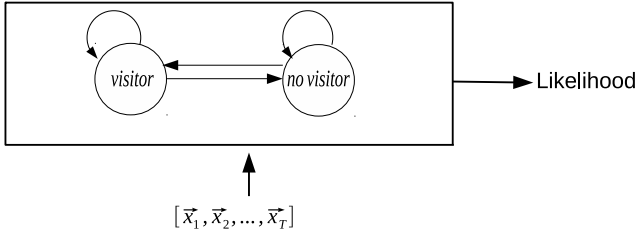
Figure 3: A Hidden Markov Model used to detect the presence of visitors.

4.3 High-level processing layer

In this section, we utilize the extracted state sequence $\{y_1, y_2, \ldots, y_T\}$, that is the most likely to have generated from the given observation sequence $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$ to build a rule-based approach to compute the number and the duration of visits. The inference rules have the generic form: IF A AND/OR B THEN C. For instance, the number of visits are found, when a visitor is detected and no other visits took place within a certain time range (e.g. one hour), then update the visitor count. As another example, the visits duration are found, when a visitor is detected and the visitor continuously confirms his/her presence within a certain time range, then update the time of being with a visitor. Visual inspection of the recordings has been performed to define the rules with the most appropriate thresholds to extract the number and the duration of visits.

5. EXPERIMENTS

We collected 10 months of real-life recordings using 10 low-resolution visual sensors producing an image of 30 x 30 pixels at a frame rate of 50 fps. The captured videos suffer from a problem, which is quite prominent indoors environment: intensity fluctuations due to fluorescent lights. In Europe, these lights flicker at a frequency of about 100 Hz. If the visual sensors are operated at a frame rate not harmonically

Month	#Ground truth	#Estimated ground truth	#Visual sensors
April	14	10	7
May	43	10	7
June	35	21	9
October	44	30	9
November	61	54	9
December	33	19	7
January 2015	0	0	9

Table 1: Comparison between the number of the ground truth and the estimated ground truth visits.

related to this flicker frequency, the captured videos will display unwanted intensity variations. As a result, we operate the camera at a frame rate of 50 fps to reduce this effect [17]. The collected visual sensor data includes partial (from 5 to 8 running visual sensors) and full recordings. We use both types of recordings to have a continuous data for visit analysis. Video capturing is synchronized. Each day of data corresponds to a 24 hour period starting from midnight.

There are two type of visits: (1) daily visits from the caregivers that could last for a few minutes (2) regular non-daily visits from family members that could last for several hours. The ground truth consist of diaries where the family members and the caregivers wrote down the start and the end time of each visit. According to the diaries, the concurrent visits do not happen too often. So, our approach of using low-resolution visual sensors assumes the user receives a single visit at a time.

Table 1 shows the comparison between the number of visits for the ground truth and the estimated ground truth per month. In May, the number of visits for the estimated ground truth (10) is the lowest compared to the ground truth (43), while April, November and January 2015 have the best estimated ground truth results with 10, 54 and 0 visits. June, October and December have the second best estimated results with 21, 30 and 19 visits, respectively.

Table 2 shows the visits duration of the ground truth and the estimated ground truth. April and January 2015 have the best estimated visit duration of 50.25 and 0 minutes compared to the ground truth, because the elderly was hospitalized in these months. We use the Mean Absolute Error (MAE) to evaluate the estimation of visit duration performance. The MAE is defined as the average of the absolute errors between the estimated ground truth visits duration and the ground truth visits duration:

$$MAE = \sum_{k=1}^{N} \frac{|\widehat{v_k} - v_k|}{N} \qquad (2)$$

where $\widehat{v_k}$ is the ground truth visit duration, v_k is the estimated ground truth visit duration and N is the total number of months. In table 2, May, October and November have the second best absolute errors of 165, 147 and 275 minutes. The MAE of the estimated ground truth visits duration is 195.25 minutes per month. This corresponds to MAE of 6.51 minutes per day. There are some obstacles to detect visitors with the applied sensors which result in some wrong detection: (1) some visits are not detected, because these visits last only for a few minutes which make them hard to detect, (2) the number of the active visual sensors have an impact on the detection of the visitor (3) often the rule-based approach does not lead to results with 100% certainty.

Month	Ground truth	Estimated ground truth	Absolute error
April	70	50	19.75
May	215	50	165
June	3960	3528	432
October	2020	1873	147
November	2190	1915	275
December	1195	867	328
January 2015	0	0	0

Table 2: Comparison between the visits duration of the ground truth and the estimated ground truth (minutes).
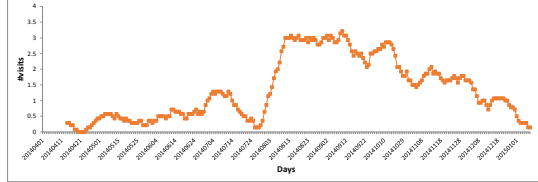
Figure 4: The running moving average of person visits.

In the second set of experiments, we use the moving running average to analyse our data points and to smooth out short-term fluctuations. Given a sequence of $(m)_{i=1}^{L}$, an n-moving average is a new sequence of $(u)_{i=1}^{L-n+1}$ defined from m_i by taking the mean of subsequence of n terms:

$$u_i = \frac{1}{n} \sum_{j=1}^{i+n-1} m_j \qquad (3)$$

Figure 4 shows the visit trends after applying the moving average. From April to July, there was a low visiting activity pattern because the elder was not present at home and he was hospitalized (April and May). In this period, the visits were performed most of the time by caregivers and only few visits were performed by family members. This indicates a low socialization level. From August to December, we can see a high visiting activity pattern, where the peak is in August and September. This suggests that the family members were less busy and had more time to visit the elder. The visits decrease until they reach the lowest duration in the end of December and January 2015, due to a new hospitalization during Christmas time.

Finally, Figure 5 presents the moving average of the visits duration. The highest peak is between late June until mid July, we have visually inspected the recordings in this period and we have found that the elder had long visits from a family member. Generally, the visits duration are longer after July. This confirms the occurrence of more family member visits in these months. Figure 6 shows the time durations when the elder was alone and with visitor(s). Also, we can clearly notice the periods of hospitalization. The time of being with a visitor is computed using a rule-based approach as explained in Section 6. While, the time of being alone is computed by simply subtracting the hours of the day from the time of being with a visitor.

6. CONCLUSION AND FUTURE WORK

We are interested to evaluate the socialization level of the
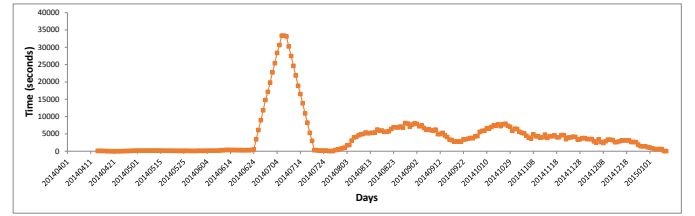
Figure 5: The running moving average of visits duration.
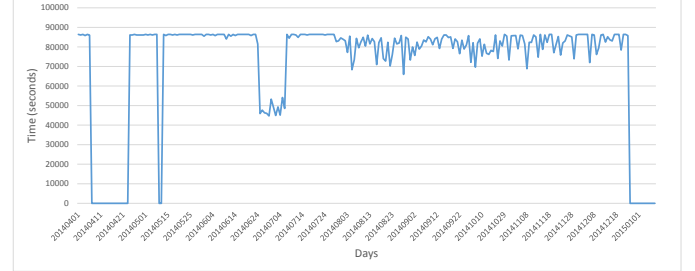
Figure 6: The time spent alone and with visitor(s).

elderly. To be certain that, the elder is not alone for long time which may affect his physical and mental health condition. In this context, we proposed a framework for detecting the presence of visitors using low-resolution visual sensors (30×30 pixels). Firstly, we extracted the visual sensor features by computing the foreground pixels to track the motion pattern in each visual sensor. Then, we used a Hidden Markov Model (HMM) for the actual detection. Finally, we have introduced a rule-based classifier to infer the number of visits and the visits duration. Our approach to sensing aimed at evaluating the feasibility of using low-resolution visual sensors for visitor detection. We have evaluated our framework on ten months of real-life recordings of an elder user. The results show promising performance compared to ground truth and the detailed analysis provided insights on the mental and physical health condition of the elderly. There is a need for further improvement, especially in detecting the short visits from the caregivers that last for only a few minutes. One of the improvement strategies is to combine the visitor detection through low-resolution visual sensors with other modalities such as the main door ring to log all the events. Another improvement strategy is to train a 3-state HMM or classifier (i.e. Support Vector Machine) in order to predict/detect among the 3 states: (1) no visitor, (2) caregiver visitor and (3) family visitor.

Acknowledgment

7. REFERENCES

[1] Second biennial report on social services of general

interest. European Commission, 2010.

[2] A. N. Aicha, G. Englebienne, and B. Kröse. How busy is my supervisor?: Detecting the visits in the office of my supervisor using a sensor network. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '12, pages 12:1–12:7, New York, NY, USA, 2012. ACM.

[3] A. N. Aicha, G. Englebienne, and B. Kröse. Modeling visit behaviour in smart homes using unsupervised learning. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 1193–1200, New York, NY, USA, 2014. ACM.

[4] P. Baggenstoss. A modified baum-welch algorithm for hidden markov models with multiple observation spaces. *Speech and Audio Processing, IEEE Transactions on*, 9(4):411–416, May 2001.

[5] M. Camilli and R. Kleihorst. Demo: Mouse sensor networks, the smart camera. In *Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on*, pages 1–3, Aug 2011.

[6] A. A. Chaaraoui, J. R. Padilla-LÃₛpez, F. J. FerrÃ₎ndez-Pastor, M. Nieto-Hidalgo, and F. FlÃₛrez-Revuelta. A vision-based system for intelligent monitoring: Human behaviour analysis and privacy by context. *Sensors*, 14(5):8895–8925, 2014.

[7] J. de Jong Gierveld and C. Tesch-RÃűmer. Loneliness in old age in eastern and western european societies: theoretical perspectives. *European Journal of Ageing*, 9(4):285–295, 2012.

[8] F. Deboeverie, J. Hanca, R. Kleihorst, A. Munteanu, and W. Philips. A low-cost visual sensor network for elderly care. *SPIE Newsroom*, December 2014.

[9] G. Demiris, M. J. Rantz, M. A. Aud, K. D. Marek, H. W. Tyrer, M. Skubic, and A. A. Hussam. Older adults' attitudes towards and perceptions of âĂŸsmart homeâĂŹ technologies: a pilot study. *Informatics for Health and Social Care*, 29(2):87–94, 2004.

[10] J. K. Eckert, L. A. Morgan, and N. Swamy. Preferences for receipt of care among community-dwelling adults. *Journal of Aging & Social Policy*, 16(2):49–65, 2004. PMID: 15148044.

[11] M. Eldib, N. B. Bo, F. Deboeverie, J. Nino, J. Guan, S. Van de Velde, H. Steendam, H. Aghajan, and W. Philips. A low resolution multi-camera system for person tracking. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 378–382, Oct 2014.

[12] M. Eldib, B. B. Nyan, F. Deboeverie, X. Xie, H. Aghajan, and W. Philips. Behavior analysis for aging-in-place using similarity heatmaps. In *8th ACM/IEEE international conference on Distributed Smart Cameras, Proceedings*, page 6. ACM/IEEE, 2014.

[13] J. Fernandez-Berni, R. Carmona-Galan, R. del Rio, R. Kleihorst, W. Philips, and A. Rodriguez-Vazquez. Focal-plane sensing-processing: A power-efficient approach for the implementation of privacy-aware networked visual sensors. *Sensors*, 14(8):15203, 2014.

[14] T. Heart and E. Kalderon. Older adults: Are they ready to adopt health-related ict? *International Journal of Medical Informatics*, 82(11):e209 – e231, 2013.

[15] A. Jalal, S. Kamal, and D. Kim. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors*, 14(7):11735–11759, 2014.

[16] A. Nait Aicha, G. Englebienne, and B. Kröse. How lonely is your grandma?: Detecting the visits to assisted living elderly from wireless sensor network data. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, UbiComp '13 Adjunct, pages 1285–1294, New York, NY, USA, 2013. ACM.

[17] B. B. Nyan, F. Deboeverie, M. Eldib, J. Guan, X. Xie, J. Niño Castañeda, D. Van Haerenborgh, M. Slembrouck, S. Van de Velde, H. Steendam, P. Veelaert, R. Kleihorst, H. Aghajan, and W. Philips. Human mobility monitoring in very low resolution visual sensor network. *Sensors*, 14(11):20800–20824, 2014.

[18] J. Petersen, N. Larimer, J. Kaye, M. Pavel, and T. Hayes. Svm to detect the presence of visitors in a smart home environment. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 5850–5853, Aug 2012.

[19] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, Feb 1989.

[20] M. Skubic, G. Alexander, M. Popescu, M. Rantz, and J. Keller. A smart home application to eldercare: Current status and lessons learned. *Technol. Health Care*, 17(3):183–201, Aug. 2009.

[21] C. Suh and Y.-B. Ko. Design and implementation of intelligent home control systems based on active sensor networks. *Consumer Electronics, IEEE Transactions on*, 54(3):1177–1184, August 2008.

[22] T. Teixeira, G. Dublon, and A. Savvides. Physical and mental health effects of family caregiving. *ENALAB technical report*, 2010.

[23] J. Wherton and D. Prendergast. The building bridges project: Involving older adults in the design of a communication technology to support peer-to-peer social engagement. In A. Holzinger and K. Miesenberger, editors, *HCI and Usability for e-Inclusion*, volume 5889 of *Lecture Notes in Computer Science*, pages 111–134. Springer Berlin Heidelberg, 2009.

[24] T. Winkler and B. Rinner. Security and privacy protection in visual sensor networks: A survey. *ACM Comput. Surv.*, 47(1):2:1–2:42, May 2014.

[25] X. Xie, F. Deboeverie, M. Eldib, W. Philips, and H. Aghajan. Phd forum: Analyzing behaviors patterns of the elderly from low-precision trajectories. In *Proceedings of the International Conference on Distributed Smart Cameras*, ICDSC '14, pages 47:1–47:2, New York, NY, USA, 2014. ACM.

[26] S. H. Zarit, K. E. Reever, and J. Bach-Peterson. Relatives of the impaired elderly: Correlates of feelings of burden. *The Gerontologist*, 20(6):649–655, 1980.