Behavior Analysis for Aging-in-Place using Similarity Heatmaps

Mohamed Eldib[1], Nyan Bo Bo[1], Francis Deboeverie[1],
Xingzhe Xie[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

The demand for healthcare services for an increasing population of older adults is faced with the shortage of skilled caregivers and a constant increase in healthcare costs. In addition, the strong preference of the elderly to live independently has been driving much research on "ambient-assisted living" (AAL) systems to support aging-in-place. In this paper, we propose to employ a low-resolution image sensor network for behavior analysis of a home occupant. A network of 10 low-resolution cameras (30×30 pixels) is installed in a service flat of an elderly, based on which the user's mobility tracks are extracted using a maximum likelihood tracker. We propose a novel measure to find similar patterns of behavior between each pair of days from the user's detected positions, based on heatmaps and Earth mover's distance (EMD). Then, we use an exemplar-based approach to identify sleeping, eating, and sitting activities, and walking patterns of the elderly user for two weeks of real-life recordings. The proposed system achieves an overall accuracy of about 94%.

Keywords

Aging-in-Place, Behaviour Analysis, Low-Resolution Cameras

1. INTRODUCTION

Thanks to the advances in medicinal healthcare in the past few decades people live longer and healthier lives than before. According to the UN world population prospects [17], by 2020 more than 25% of Europe's population will be over 60. During the same period, the number of people over 80 will double. The institutions that take care of the elderly citizens will run into operational and financial problems in the approaching decade. Given the fact that a high ratio of older adults prefer to live independently in their homes, and given the costs of nursing home care [1], it is clear that technology can play an important role in solving these problems

by supporting the elderly, helping the informal and formal caregivers, and by creating a safe environment for elderly people and reducing the workload of the caregivers.

In this paper, we propose a low-resolution image sensor network for the purpose of behavior analysis of an elderly user living at home. We outline the main contributions of this paper as: (1) a low-resolution multi-camera system installed in a service flat of an elderly for real-life data collection; (2) a novel measure to find similar performed activities between each pair of days based on heatmaps and Earth mover's distance (EMD) [8]; (3) a method to extract similar activities such as sleeping, eating, sitting, in-kitchen work, and walking around from two weeks of recordings using an exemplar-based approach [2]; and (4) the introduction and evaluation of a novel sensing system, i.e. a network of low-resolution cameras, for addressing the problem of behavior analysis, which to the best of our knowledge, has not been examined before. Given the nature of the acquired low-resolution images, such a network may prove to offer an acceptable sensing mechanism for addressing the privacy concerns of the users.

The paper is organized as follows. In section 2, we present related work. Section 3 provides an overview of the installed low-resolution image sensor network in the service flat. Section 4 presents our methods to define and treat similarity heatmaps and the use of an exemplar-based approach for behavior analysis. Section 5 describes our experimental results. Finally, Section 6 offers some conclusions.

2. RELATED WORK

There has been a plethora of work to develop "ambient-assisted living" (AAL) systems which are capable of improving, preventing and curing the wellness and health conditions of the elderly. The major body of work in monitoring Activities of Daily Living (ADLs) falls into two categories: a) using body-worn sensors to monitor a person's activities[13, 3, 15], and b) implementing a behavior monitoring system based on sensors in a smart home [20, 19, 18]. In body-worn sensors, various wearable and e-textile sensors could be attached to the subject of interest to monitor vital signs such heart, respiration, blood pressure and blood glucose, as well as to detect falls. But, wearable sensors face a few disadvantages such as battery life, high cost, being forgotten to be worn by the user, and the requirement to be worn on specific body parts to provide reliable measurements.

In the second category of works on AAL, smart homes are equipped with various kinds of sensors such as video cam-

eras, Passive Infrared Motion (PIR) sensors , microphones, and magnetic switches to support aging-in-place for the elderly. CASAS [20] is an on-going smart home project aiming at monitoring a list of daily activities to evaluate completeness of daily tasks in order to detect signs of dementia. The Millennium Home [18] detects any deviations from normal daily activities by using rule-based techniques. Age In Place [19] aims at early detection of urinary tract infections for the elderly. Nguyen applied Hierarchical Hidden Markov Model (HMM) to learn and detect a set of high-level complex behaviors with shared structure from the user's movement trajectory [16]. In [6], Brdiczka learned human behavior models in a smart home environment. Different situations, such as "introduction" and "presentation", are inferred from multiple observations using HMM: individual activities are recognized from video recordings using 3D tracking system, and speech activities from audio recording.

In [9, 12, 22], the researchers tried to identify abnormal behaviors from normal ones using various models. In [9], Cuddihy trained an alert line for the inactivity level based on motion and door sensors, to detect abnormal long activity of the seniors. In order to detect the unusual fall from the normal daily activities, Foroughi [12] extracted a human body ellipse, projection histograms of the segmented silhouette and the human head position as features, to show the temporal change of the human shape initiated by human activities. In [22], Naohiko proposed a method to learn human behavior patterns from the trajectories. Multi-Dimensional Scaling (MDS) was applied to project the trajectories onto a low-dimensional space, and a K-means method was used to cluster the projected locations into different patterns. Then, an HMM was employed to model the time-series features of the trajectories in each cluster, and subsequently, abnormal trajectories were detected based on the probabilistic distance between HMMs. The authors in [23] provided a replacement of low-resolution cameras by using an infrared ceiling-mounted sensor network for behavior analysis and fall detection.

Our approach for behavior monitoring in a smart living space is different from the work in [12, 22, 16], in that we opted to use low-resolution cameras, and not cameras with regular imaging resolutions which often raise privacy concerns in users. Our aim has also been to reduce the installation cost of the data collection system in a smart home to around 100€ per room, while delivering good performance in behavior analysis, matching the performance of other sensing modalities.

Figure 1: Low-resolution image sensor [7].

3. SERVICE FLAT SETUP

The sensor network deployed in the service flat is composed of 10 camera systems, each system being a stereo pair of image sensors with a resolution of 30×30 pixels. In an earlier work [10], we used low-resolution cameras of [7] for person tracking, and reported promising tracking performance. We use the same low-resolution cameras in the current study. Figure 1 shows the low-resolution camera. The sensor network is installed in a living space in the service flat, covering an area of 8×4.4 m^2. Figure 2 displays the living space layout with camera positions. The 10 low-resolution cameras are connected to two computers (5 cameras per computer). The cameras produce an output of 80 gigabytes for each day. Biweekly visits are arranged to collect the data. The elderly occupant is at the age of 80 with hearing impairment and walking abnormalities.

Each camera in the room was calibrated separately. We follow the same approach as in [10] based on correlation method [4] and maximum likelihood tracker [5], due to their good performance with low-resolution cameras of [7]. The overall system architecture can be seen in Figure 3.

3.1 Track Extraction

After each camera captures a new frame, the maximum likelihood tracker [5] analyzes the frame to separate moving objects from the static background using correlation method [4]. This produces a number of blobs, some of which correspond to noise or non-interesting moving objects such as chairs. Each blob is then checked if it is overlapping with the bounding box of the tracked persons in the previous frame. Only non-overlapping blobs are matched across all camera views using homography and well-matched blobs are initialized as a new person for tracking. Next, in each camera view the likelihood that a person is in a particular position in the room is calculated using the known position at the previous frame as prior knowledge. The fusion center computes the joint-likelihood based on likelihood computed by each camera, and estimates the most likely new position of the person. Finally, jointly estimated positions are sent back to all camera views as a prior for likelihood computation in the next frame.

The parameters of correlation were tuned to achieve optimal detection. The obtained results were good enough for tracking as shown in Figure 4.

4. SIMILARITY HEATMAPS

The user's detected ground positions are used to generate similarity heatmaps for activity analysis like sleeping, eating, and sitting activities, and walking patterns. A heatmap is defined as a map indicating the regions where most of a person's positions occur, or most of their time is spent at.

A heatmap is created as follows: (1) each day is divided into an equal number of equal time intervals. Choosing a proper time interval is important to the activity analysis process. An interval which is too small may not include enough time to perform an activity, while an interval which is too large may include time to perform multiple activities. The choice of the time interval is explained in the experiments section; (2) for each time interval, the ground position occurrences are accumulated in a heatmap. The heatmap is a representation of the living space coordinates. Every time a person's position occurs, the counter for that ground position is increased by 1 in the heatmap. As an example,
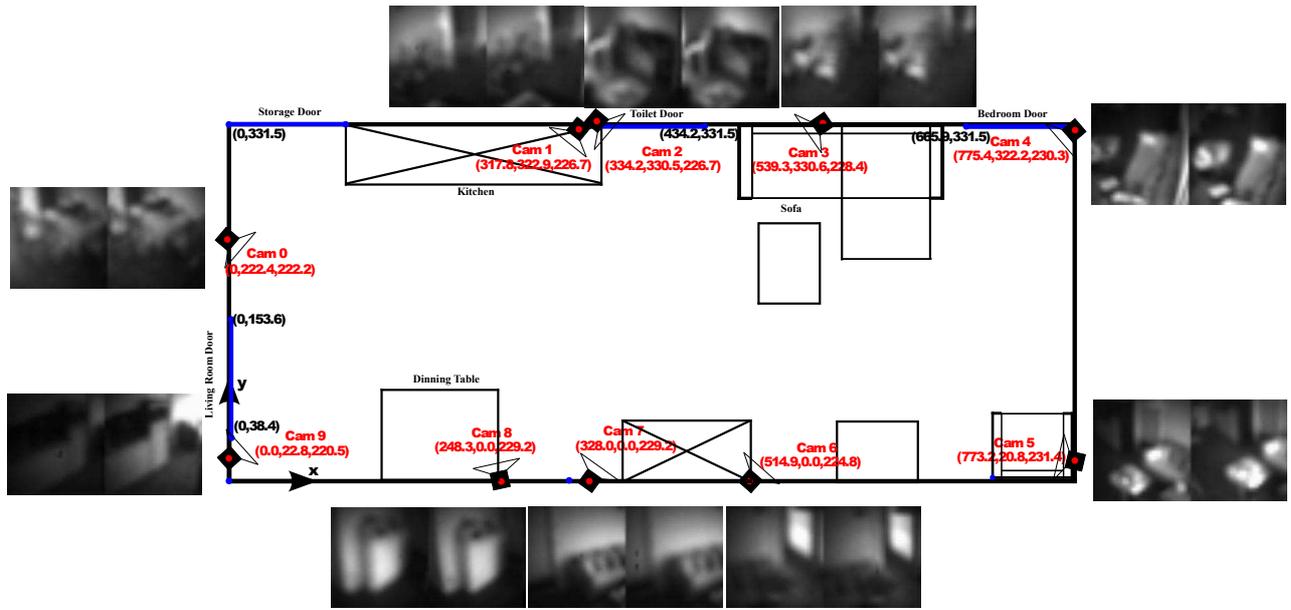
Figure 2: Layout of the living space setup (8×4.4 m^2) with camera positions.
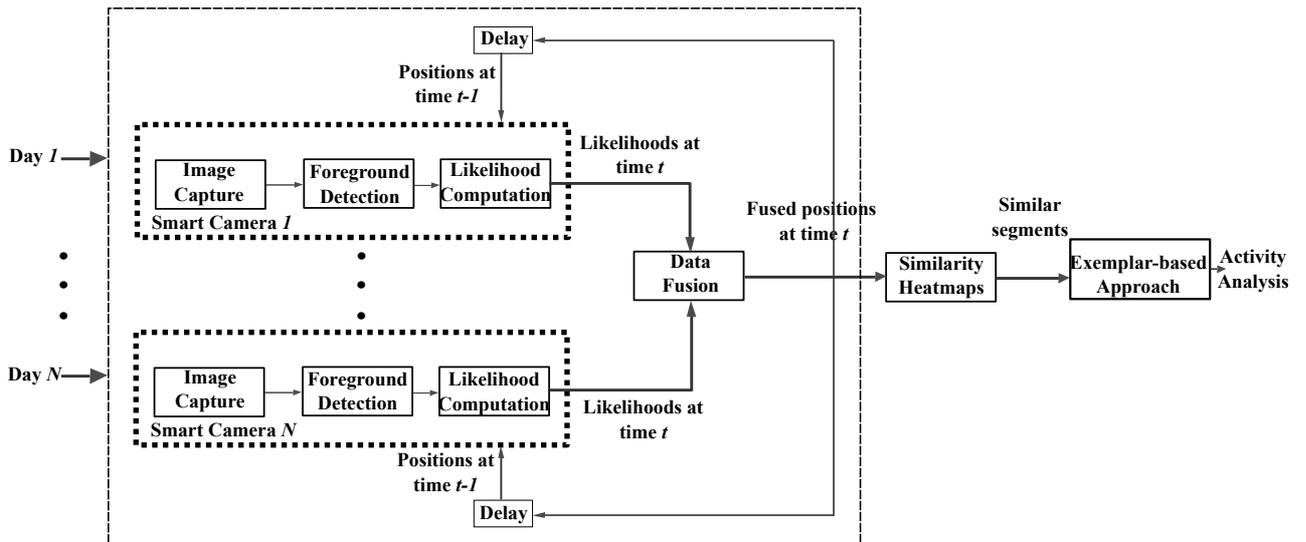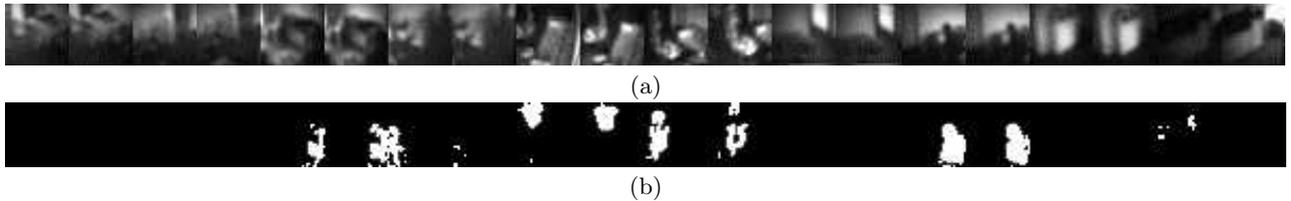

Figure 3: Processing architecture.

(a)

(b)

Figure 4: (a) Original image, (b) foreground detection by the correlation method.
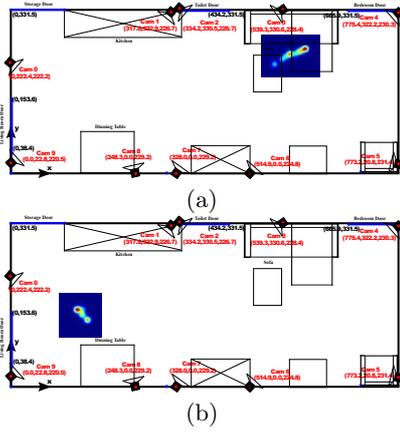
Figure 5: Heatmaps for (a) sitting, (b) eating.

Figure 6: Heatmaps for (a) walking around, (b) in-kitchen.

let's say we have two heatmaps of the user working in the kitchen. In one, the user stays in one location most of the time, and in the other he walks around. We want to select a distance metric which when applied to these two heatmaps results in a lower cost than when the heatmaps related to the user's presence in two different areas of the home. The earth mover's distance (EMD) serves this purpose [8].

The similarity between the time intervals for each pair of days i and j is found using EMD. The EMD is a measure of the distance between two probability distributions which are represented by signatures, or the minimal amount of work that must be performed to transform one distribution into the other by moving "distribution mass" [21]. The signatures are sets of weighted features that capture the distributions. In this work, the signature contains the location occurrence at each point followed by the location coordinates. Then, the EMD is used to compute the Manhattan distance between the signatures:

$$E_{i,j} = \text{EMD}(S_i^c, S_j^c) \text{ for } c \in i \,\&\, j, \qquad (1)$$

where S_i^c and S_j^c are the signatures for time interval c of days i and j. Finally, the most similar time intervals between days i and j are found by taking the minimum value for each row in $E_{i,j}$.

4.1 Exemplar-based Approach

Activity analysis can be treated as a single-layered approach, when each activity is considered as a particular class of image sequences. Alternatively, activity analysis can be treated as a hierarchical approach, when a model (e.g. layered hidden Markov models) is constructed to represent and recognize high-level human activities such as fighting, kicking and punching activities. The use of exemplar-based modelling has shown promising results in activity analysis [24, 11] and human-object interaction [14]. In our experiments, we used an exemplar-based approach, which is a subcategory of single-layer approaches. In this approach, the relation between classes of activity and their corresponding training samples are mapped directly using sequence of images [2]. We redefine the exemplar-based approach such that a heatmap is labelled with activity classes.

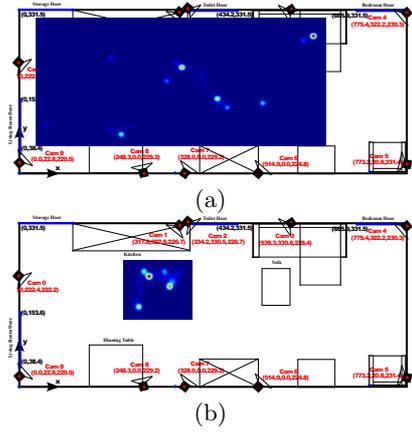Our exemplar-based approach is composed of two stages: (1) training stage, and (2) testing stage. In the training stage, each day is divided into number of equal time intervals, where each time interval has a corresponding heatmap. The heatmap is manually labelled with an activity class. There is a class of 5 activities (sleeping, eating, sitting, walking around and in-kitchen). The output for each day is a text file with time intervals and their corresponding activities. Figures 5 and 6 show heatmaps for sitting, eating, walking around, and in-kitchen activity examples. Sleeping is identified when there is no activity detected during a number of time intervals.

In the testing stage, the unknown activities for the time intervals of day j (test sample) are found by extracting the most similar time intervals of day i (training sample) using $E_{i,j}$. Then, the activities that correspond to the most similar time intervals of day i are considered the predicted activities for the time intervals of day j.

In our experiments, the number of the training samples were 4 days, so the activity class that has the highest number of occurrences in a time interval c in the 4 training samples, is considered the predicted activity for the time interval of the test sample.

5. EXPERIMENTS

We collected two weeks of real life recordings using 10 low-resolution cameras of 30×30 pixels, with a frame rate of 50 fps. The videos were first synchronized. Each day of data corresponds to a 24 hour period starting from midnight, and the chosen time interval was 15 minutes, resulting in 96 time intervals per day. This time interval is not considered large since the elderly occupant was suffering from walking abnormalities, and it took him some time to perform an activity.

5.1 Results

Our method was able to identify sleeping, eating, sitting, walking around, and in-kitchen activities. The details of ADL occurrences can be found in Table 1. The sitting occurrence number is high due to walking abnormalities of the elderly user, and the elderly user prefers to sit more than to walk. We used 4 days as training samples and 10 days as test samples. The class of activities for each time interval of the test sample are found by using the training samples as follows: (1) the time intervals of the test samples similar to

Activity class	Number of occurrences
Sleep	10
Eat	32
Walk around	51
Sit	71
kitchen	27

Table 1: ADLs for test samples.

		Prediction			
	Sleep	Eat	Walk	Sit	Kitchen
Sleep	10	0	0	0	0
Eat	0	30	0	2	0
Walk	0	5	46	0	0
Sit	0	1	4	63	3
Kitchen	0	2	9	0	16

(row label on left: Truth)

Table 2: Confusion matrix of ADLs for test samples.

the time intervals of the training sample are extracted according to similarity heatmaps, and (2) the class of activities for the test samples are found by using the exemplar-based approach.

Table 2 shows the confusion matrix for 10 days of ADLs. As one can see, promising results are obtained. The similarity heatmap measure is able to predict all sleeping activities during the 10 days correctly. Eating, sitting and walking around activities have the second best predicted values (30, 63, and 46 respectively), while in-kitchen activities have the least accurate predicted value (16). The overall accuracy of our proposed behavior analysis is 94.6%. Table 3 shows comparative results for each ADL. The achieved performance is promising, and that is due to the similarity heatmap's ability to find similar behaviors between pairs of days. In summary, sleeping, walking around and sitting have the best overall performance because these activities are characterized by their own similarity heatmaps, whereas eating and in-kitchen activities are performing at a lower level because the reported ground positions are often close to each other (dinning table and the kitchen area are close to each other).

6. CONCLUSIONS

In this paper, we reported results of behavior analysis for an eldely user based on an installed network of low-resolution image sensors in the user's service flat. Our approach to sensing aimed at evaluating the feasibility of using low- resolution cameras for activity analysis. We used a maximum likelihood tracker combined with a correlation-based method as foreground detection for extracting tracks and ground po-

Activity class	Percision	Recall	Accuracy
Sleep	100%	100%	100%
Eat	78.95%	93.75%	94.82%
Walk around	77.97%	90.2%	90.58%
Sit	96.92%	88.73%	94.76%
kitchen	84.21%	59.25%	92.78%

Table 3: Precision, recall and accuracy of ADLs for test samples.

sitions of the occupant across the living space. Then, we proposed a similarity heatmap measure as a way of finding similar activities between time segments in pairs of days. Next, we employed an exemplar-based approach to analyze sleeping, eating, sitting, walking around and in-kitchen activities. We evaluated our method on two weeks of real-life recordings from the elderly user. The results show promising performance of our proposed system with a 94.6% overall accuracy.

8. REFERENCES

[1] Centers for disease control and prevention, the state of aging and health in america 2007. *N. A. on an Aging Society*, Available: http://www.raytron.co.jp.

[2] J. Aggarwal and M. Ryoo. Human activity analysis: A review. *ACM Comput. Surv.*, 43(3):16:1–16:43, Apr. 2011.

[3] U. Anliker, J. Ward, P. Lukowicz, G. Troster, F. Dolveck, M. Baer, F. Keita, E. Schenker, F. Catarsi, L. Coluccini, A. Belardinelli, D. Shklarski, M. Alon, E. Hirt, R. Schmid, and M. Vuskovic. Amon: a wearable multiparameter medical monitoring and alert system. *Information Technology in Biomedicine, IEEE Transactions on*, 8(4):415–427, Dec 2004.

[4] N. B. Bo, S. Gruenwedel, P. Van Hese, J. Niño-Castañeda, D. Van Haerenborgh, D. Van Cauwelaert, P. Veelaert, and W. Philips. Phd forum: Illumination-robust foreground detection for multi-camera occupancy mapping. In *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*, pages 1–2, Oct 2012.

[5] N. Bo Bo, P. Van Hese, S. Gruenwedel, J. Guan, J. Niño-Castañeda, D. Van Haerenborgh, D. Van Cauwelaert, P. Veelaert, and W. Philips. Robust multi-camera people tracking using maximum likelihood estimation. In J. Blanc-Talon, A. Kasinski, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, volume 8192 of *Lecture Notes in Computer Science*, pages 584–595. Springer International Publishing, 2013.

[6] O. Brdiczka, M. Langet, J. Maisonnasse, and J. L. Crowley. Detecting human behavior models from multimodal observation in a smart home. *Automation Science and Engineering, IEEE Transactions on*, 6(4):588–597, 2009.

[7] 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.

[8] S.-H. Cha and S. N. Srihari. On measuring the distance between histograms. *Pattern Recognition*, 35(6):1355 – 1370, 2002.

[9] P. Cuddihy, J. Weisenberg, C. Graichen, and M. Ganesh. Algorithm to automatically detect

abnormally long periods of inactivity in a home. In *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*, pages 89–94. ACM, 2007.

[10] M. Eldib, N. B. Bo, F. Deboeverie, J. Nino, J. Guan, S. V. de Velde, H. Steendam, H. Aghajan, and W. Philips. A low resolution multi-camera system for person tracking. In *Image Processing (ICIP), 2014 21th IEEE International Conference on*, 2014.

[11] A. Elgammal, V. Shet, Y. Yacoob, and L. S. Davis. Learning dynamics for exemplar-based gesture recognition. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR'03, pages 571–578, Washington, DC, USA, 2003. IEEE Computer Society.

[12] H. Foroughi, B. S. Aski, and H. Pourreza. Intelligent video surveillance for monitoring fall detection of elderly in home environments. In *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, pages 219–224. IEEE, 2008.

[13] D. Hodgins, A. Bertsch, N. Post, M. Frischholz, B. Volckaerts, J. Spensley, J. M. Wasikiewicz, H. Higgins, F. Von Stetten, and L. Kenney. Healthy aims: Developing new medical implants and diagnostic equipment. *Pervasive Computing, IEEE*, 7(1):14–21, Jan 2008.

[14] J.-F. Hu, W.-S. Zheng, J. Lai, S. Gong, and T. Xiang. Recognising human-object interaction via exemplar based modelling. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3144–3151, Dec 2013.

[15] Y. Masuda, M. Sekimoto, M. Nambu, Y. Higashi, T. Fujimoto, K. Chihara, and T. Tamura. An unconstrained monitoring system for home rehabilitation. *Engineering in Medicine and Biology Magazine, IEEE*, 24(4):43–47, July 2005.

[16] N. T. Nguyen, D. Q. Phung, S. Venkatesh, and H. Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden markov model. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 955–960. IEEE, 2005.

[17] J. op de Beke. Confronting demographic change: a new solidarity between the generations. *COMMUNICATION FROM THE COMMISSION*, 2005.

[18] M. Perry, A. Dowdall, L. Lines, and K. Hone. Multimodal and ubiquitous computing systems: supporting independent-living older users. *Information Technology in Biomedicine, IEEE Transactions on*, 8(3):258–270, Sept 2004.

[19] M. Rantz, M. Skubic, R. Koopman, L. Phillips, G. Alexander, S. Miller, and R. Guevara. Using sensor networks to detect urinary tract infections in older adults. In *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, pages 142–149, June 2011.

[20] P. Rashidi and D. Cook. Keeping the resident in the loop: Adapting the smart home to the user. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(5):949–959, Sept 2009.

[21] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision*, 40(2):99–121, Nov. 2000.

[22] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato, and Y. Fujino. Learning motion patterns and anomaly detection by human trajectory analysis. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 498–503. IEEE, 2007.

[23] S. Tao, M. Kudo, and H. Nonaka. Privacy-preserved behavior analysis and fall detection by an infrared ceiling sensor network. *Sensors*, 12(12):16920–16936, 2012.

[24] D. Weinland, E. Boyer, and R. Ronfard. Action recognition from arbitrary views using 3d exemplars. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–7, Oct 2007.