Pose Estimation and Movement Detection for Mobility Assessment of Elderly People in an Ambient Assisted Living Application

Julia Richter(✉), Christian Wiede, and Gangolf Hirtz

Department of Electrical Engineering and Information Technology,
Technische Universität Chemnitz, Reichenhainer Str. 70, 09126 Chemnitz, Germany
{julia.richter,christian.wiede,g.hirtz}@etit.tu-chemnitz.de

Abstract. In European countries, the increasing number of elderly with dementia causes serious problems for the society, especially with regard to the caring sector. As technical support systems can be of assistance to caregivers and patients, a mobility assessment system for demented people is presented. The grade of mobility is measured by means of the person's pose and movements in a monitored area. For this purpose, pose estimation and movement detection algorithms have been developed. These algorithms process 3-D data, which are provided by an optical stereo sensor installed in a living environment. The experiments demonstrated that the algorithms work robustly. In connection with a human machine interface, the system facilitates a mobilisation as well as a more valid assessment of the patient's medical condition than it is presently the case. Moreover, recent advances with regard to action recognition as well as an outlook about necessary developments are presented.

Keywords: Pose estimation · Stereo vision · Image understanding · Video analysis · 3-D image processing · Machine learning · Support vector machine · Ambient assisted living

1 Introduction

The increasing life expectancy is an important achievement of modern medicine. Over the coming years, the number of elderly people will continually rise and with it the number of demented people [3]. Due to this development, care facilities will encounter challenges in maintaining the quality of human care.

People in an early state of dementia should remain in their familiar household as long as possible in order to mitigate these problems. The encouragement of their cognitive, social and physical functions will also help to keep their quality of life at high level. Next to activation, assessing the need of care in regular intervals is another task medical experts are facing. Since the health status of a person is examined only irregularly at present, the result is highly dependent on the form on the inspection day and might be further influenced by the fact that patients can prepare for the inspection. Additionally, many patients put particular concern

A. Fred et al. (Eds.): ICPRAM 2015, LNCS 9493, pp. 172–184, 2015.
DOI: 10.1007/978-3-319-27677-9_11

on personal hygiene on that day and when questioned about their physical and psychological comfort, they usually feel embarrassed and avoid talking about their problems. The medical findings are therefore not always reliable.

In this paper, only persons living alone at home without the care of a partner are considered. The focus lies on the physical capabilities of the demented person – and particularly his or her mobility. This parameter was measured by the detection of the general pose (i. e. standing, sitting and lying) and of the person's movements in the living environment. To this end, a single, wide angle stereo camera was mounted at the ceiling. The information gathered about the general pose and the movements were recorded over a certain period of time. If long periods of inactivity were detected, the demented person was encouraged to do some exercises or to go for a walk. The communication was realized via a human machine interface, i. e. a tablet or a monitor, on which the messages appeared, optionally in combination with an acoustic signal. Furthermore, statistics were calculated from the recorded data. At a later time, such statistics could be analysed by medical personnel to notice considerable changes in a patient's mobility and to draw reliable conclusions about the need of care.

2 Related Work

Various works address the subject of supporting elderly people in their home environment. The assistance concepts are closely related to the topic of AAL (Ambient Assisted Living). Their unobtrusive integration into the living environment is one of the most important requirement for AAL systems.

Clement et al. detected ADLs (Activities of Daily Living) with the help of 'Smartmeters', which measure the energy consumption of household devices [5]. A Semi-Markov model was trained in order to construct behaviour profiles of persons and to draw conclusions about their state of health. Kalfhues et al. analysed a person's behaviour by means of several sensors integrated in a flat, e. g. motion detectors, contact sensors and pressure sensors [8]. Link et al. employed optical stereo sensors to discern emergencies, i. e. falls and predefined emergency gestures [10]. Chronological sequences of the height of the body centre and the angle between the main body axis and the floor were analysed. Belbachier et al., who also applied stereo sensors to detect falls [2], used a neural network-based approach to classify the fall event. The major advantage of optical sensors is their easy integration into a flat. A considerable amount of additional information can be obtained by applying image processing algorithms, especially in connection with RGB-D sensors, which deliver red, blue and green channel images as well as depth information. Therefore, we decided to use a stereo camera in our study. Although other sensors that provide RGB-D data, such as the Kinect, could also be installed in a flat, they show features that have proved to be disadvantageous with regard to the application field of AAL: Firstly, if the Kinect is mounted at the ceiling, the range and the field of view do not cover the complete room. It would be necessary to integrate several Kinect sensors at different places in a flat, which is hardly applicable. Secondly, the resolution is not sufficient enough for the recognition of objects that are far away from the sensor. When, thirdly, several Kinects are installed for better coverage of the room, they are apt to influence each other, due to their active technique for determining depth information. Consequently, although the Kinect is highly performant for a variety of applications, we considered this sensor as unsuitable for AAL purposes.

The approaches listed above either address ADL detection or emergency scenarios. In the context of assessing the health status of persons, several former projects have focused especially on the analysis of mobility. Scanaill et al. employed body-worn sensors for mobility telemonitoring [13]. However, this type of sensor unsuitable for demented persons, as this group tends to forget to put them on or puts them off intentionally. In the work of Steen et al., another way of measuring mobility was presented [14]. In first field tests, several participants' flats were equipped with laser scanners, motion detectors and contact sensors. By means of these sensors, the persons could be localised within their flats. Apart from this, the traversing time between the sensors as well as walking speeds were computed. These field tests gave evidence that the evaluation of sensor data allows conclusions about mobility.

In addition to a person's location and the movements, we think that the pose, i. e. standing, sitting and lying, provides also an indication of a person's mobility. We therefore introduce a pose estimation algorithm, which detects the pose of a person within the area observed by a single stereo camera.

There is a variety of pose estimation algorithms that use optical sensors. They differ, for example, with respect to such parameters as camera type (mono, stereo), inclusion of temporal information and utilisation of explicit human models. Ning et al. discerned the human pose using a single monocular image [11]. By modifying a bag-of-words approach, they were able to increase the discriminative power of features. They also introduced a selective and invariant local descriptor, which does not require background subtraction. The poses walking, boxing and jogging could be classified after supervised learning. Agarwal et al. determined the pose from monocular silhouettes by regression [1] and thus needed neither a body model nor labelled body parts. Along with spatial configurations of body parts, Ferrari et al. additionally considered the temporal information in their study [6]. Haritaoglu et al. employed an overhead stereo camera in order to recognize the 'pick' movement of customers while shopping [7]. In this study, a three dimensional silhouette was computed by back-projecting image points to their corresponding world points by the use of depth information and calibration parameters. The persons' localizations were found at regions with significant peaks in the occupancy map. The pose is determined by calculating shape features instead of using an explicit model. Other approaches applied the Kinect sensor. Their results proved that the Kinect, when suitable for the particular application, leads to results of high quality. Ye et al. estimated the pose from a single depth map of the Kinect [15]. They then compared this map with mesh models from a database. In a first step, a similar pose was searched by point cloud alignment using principal component analysis and nearest neighbour search. In a second step, the found pose was refined. Missing information of occluded parts could be replaced by data from the corresponding mesh model. As a result, skeleton joints comparable to the Kinect skeleton output could be determined.

Another study addressed the design of a scale and viewing angle robust feature vector, which describes a person's head-to-shoulder signature [9]: Points between head and shoulder are first assigned to vertical slices. The points within each slice are then projected to a virtual overhead view and the feature vector is eventually composed of the slices' spans. The authors aim at detecting persons in a 3-D point cloud. However, this approach can also be adapted and utilized for pose estimation.

3 Mobility Assessment

This section describes the algorithms for movement detection and pose estimation. First of all, the person has to be detected and localized within the monitored area. Therefore, the stereo camera is extrinsically calibrated with respect to a defined world coordinate system. The 2-D position is measured in relation to the origin of this coordinate system. On the basis of this position, the person is classified as 'moving' if the position changes considerably between two successive frames in a video sequence. The pose estimation requires three steps. Firstly, 3-D points belonging to the person are extracted from the back-projected point cloud. Secondly, discriminative feature vectors, which allow a reliable classification, are designed. Finally, a suitable machine learning technique is selected and a model is trained with feature vectors generated from training examples.

3.1 Person Localisation

The person localisation is performed on the back-projected 3-D point cloud obtained from the stereo camera [12]. Hypotheses of possible foreground regions are generated in a first step, so a mixture of Gaussian algorithm is applied to the world z-map, which represents the z component, i. e. the height, of the corresponding world point for every pixel.

The mixture model is calculated for every pixel in the map and updated for every new frame according to the new pixel value. The model was described by [16] and is expressed as follows:

$$p(x^{(t)}|\chi_{\mathrm{T}}; BG + FG) \sim \sum_{m=1}^{M} \hat{\pi}_{\mathrm{m}}^{(t)} \cdot N(x^{(t)}; \hat{\mu}_{\mathrm{m}}^{(t)}, \hat{\sigma}_{\mathrm{m}}^{2(t)}) \tag{1}$$

$p(x^{(t)}|\chi_{\mathrm{T}}; BG+FG)$ is the probability density function for the value x of a pixel in the z-map for frame t with the history χ_{T}. This density function models both the background BG and the foreground FG. M denotes the number of Gaussian distributions N. Each distribution is characterised by its mean value $\hat{\mu}_{\mathrm{m}}^{(t)}$ and its variance $\hat{\sigma}_{\mathrm{m}}^{2(t)}$. $\hat{\pi}_{\mathrm{m}}^{(t)}$ denotes the influence of every single distribution on the mixture model.

In a second step, the points within the foreground mask are projected on a virtual overhead plan view. The final determination of the persons' positions is executed on this view. The detected person is characterised by a centre point $\overrightarrow{p} = (x, y, z)$, the expansion in each direction – $expansion_{\mathrm{x}}$ and $expansion_{\mathrm{y}}$ – and an orientation α related to the world coordinate system. An example of detected persons is illustrated in Fig. 1.

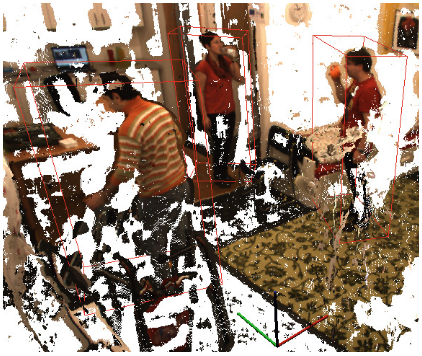


Fig. 1. Example point cloud with detected persons [12]. Detected persons are visualised via red cuboids defined by a 3-D centre point and expansions in each direction. White areas indicate regions, where 3-D world points cannot be calculated due to the lack of depth information (Color figure online).

3.2 Movement Detection

For movement detection, only vectors $\overrightarrow{p}_{\mathrm{xy}}$ containing the x and y component of the 3-D centre point $\overrightarrow{p}$ are processed.

The distance $distance_{\mathrm{frame}}$ that a person moves between two frames depends on the $frame\,rate$ and can be estimated with:

$$distance_{\mathrm{frame}} = v_{\mathrm{movement}} \cdot t_{\mathrm{frame}} = \frac{v_{\mathrm{movement}}}{frame\,rate}. \tag{2}$$

Provided a person is walking with a speed v_{movement} of at least 0.5 m/s and the frame rate is about 5 FPS, the $distance_{\mathrm{frame}}$ is estimated at 100 mm. We consider a person to be moving when a threshold distance of more than X m is covered. Therefore, we utilize a sliding window containing the vectors $\overrightarrow{p}_{\mathrm{xy}}^{(t-i)}$ with $i = \{0, ..., 4\}$. Each $distance_{\mathrm{j}}$ crossed between two successive frames is calculated according to Eq. 3 with $j = \{0, ..., 3\}$. It is the Euclidean norm between the person's position in the frame $t - j$ and the position in the previous frame $(t - j - 1)$.

$$distance_{\mathrm{j}} = \left\| \overrightarrow{p}_{\mathrm{xy}}^{(t-j)} - \overrightarrow{p}_{\mathrm{xy}}^{(t-j-1)} \right\| . \tag{3}$$

Afterwards, the distances are summed up to the final $distance$ between the five frames of the sliding window:

$$distance = \sum_{j=0}^{3} distance_{\mathrm{j}}. \tag{4}$$

The distance between two frames is only added to the sum if its value exceeds $distance_{\mathrm{frame}}$. Furthermore, the threshold X mentioned above for this sum is estimated according to the product of the estimated distance between two frames $distance_{\mathrm{frame}}$ and the number of distances $nDist$ within the window:

$$\begin{aligned} X &= distance_{\text{frame}} \cdot nDist \\ &= 100\,\frac{\text{mm}}{\text{frame}} \cdot 4\,\text{frames} = 400\,\text{mm}. \end{aligned} \tag{5}$$

Moreover, the decision about movement or non-movement is realised via a finite state machine consisting of the two states '*movement*' and '*non-movement*'. At the transitions, the *distance* is compared with two different thresholds T_{high} and T_{low} that are slightly lower/higher than the estimated threshold distance X (hysteresis):

$$\begin{aligned} T_{\text{high}} &= 500\,\text{mm}, \\ T_{\text{low}} &= 300\,\text{mm}. \end{aligned} \tag{6}$$

The hysteresis suppresses oscillations near the estimated threshold value. Finally, the value of $movement^{(t)}$ is recorded over time, so that it can be analysed later. Generally, these threshold values can be adjusted when conditions in terms of velocity and frame rate are altering.

3.3 Pose Estimation

Point Cloud Extraction. The presented pose estimation algorithm processes 3-D world points belonging to the person. Every point of the point cloud has therefore to be classified as person or non-person. For that purpose, both the previously calculated cuboid and the foreground mask are used for classification. The algorithm is outlined in the following pseudo code, which is performed for every detected person. The geometric context is illustrated in Fig. 2.

```
R = sqrt(expansion.x^2 + expansion.y^2);
for all points:
if (foregound
    && z < 2*expansion.z
    && expansion.x < R
    && expansion.y < R )
{
    (xT,yT) = CoordinateTransformation(x, y);
    if (!( xT < expansion.x
        && yT < expansion.y ))
    {
        deletePoint(x,y);
    }
}
else
{
    deletePoint(x,y);
}
```

Points are removed from the cloud if they belong neither to the foreground nor to the interior of the cuboid. In order to reduce processing power, it is first checked whether a point $(x_{\mathrm{pc}}, y_{\mathrm{pc}})$ is within the person's radius R. If this is the case, the point is transformed from the world coordinate system $(x_{\mathrm{w}}, y_{\mathrm{w}})$ to the person's coordinate system $(x_{\mathrm{p}}, y_{\mathrm{p}})$, which enables a direct comparison of the point coordinate with the corresponding expansion $expansion_{\mathrm{x}}$ and $expansion_{\mathrm{y}}$. The person's coordinate system is defined by its origin, namely the 2-D centre point $\overrightarrow{p}_{\mathrm{xy}}$, and its rotation angle α.

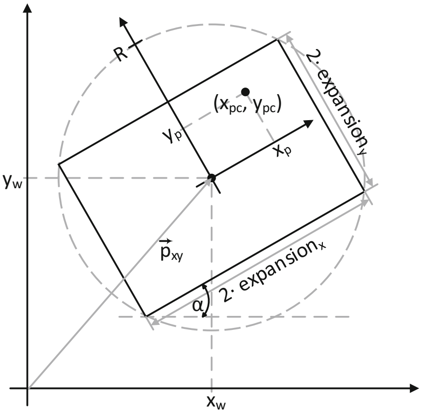


Fig. 2. Classifying points from the point cloud as person or non-person by means of coordinate transformation. If a point in the point cloud lies inside the circle defined by radius R, this very point is transformed from the world coordinate system to the person's coordinate system. Provided that the point has been classified as foreground, it belongs to the person if its x and y component fall below the corresponding expansion.

The remaining points are denoted as the person's point cloud $points_{\mathrm{person}}$. Figure 3 shows the extracted point clouds of three persons.

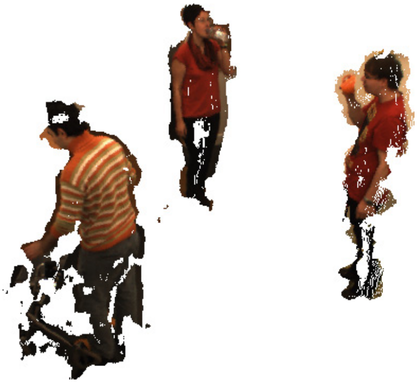


Fig. 3. Point clouds of three persons.

Feature Vector Generation. The determination of a person's pose is based on the points extracted in the previous step. In order to train a machine, a discriminative feature vector has to be designed first. For that purpose, the point cloud is divided into 20 vertical bins of 110 mm height each, which start at a z value of −100 mm. During the extrinsic calibration, the origin of the world coordinate system is set on the floor plane of the room. The plane formed by the x and the y axis runs parallel to the floor while the z axis is directed at the ceiling. Therefore, the floor is defined by a z value around zero. According to their z component, all points are assigned to one of these bins. Consequently, each bin contains the number of points that fall within a certain z range. All bins together form a feature vector. In a final step, the feature vector is normalized by dividing every item by the total number of points n. The process of feature vector generation is visualised in Fig. 4.

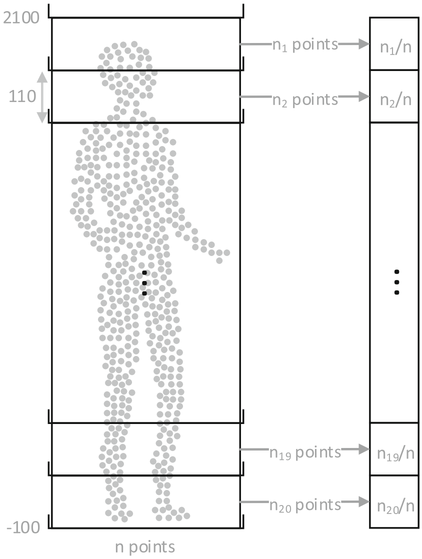
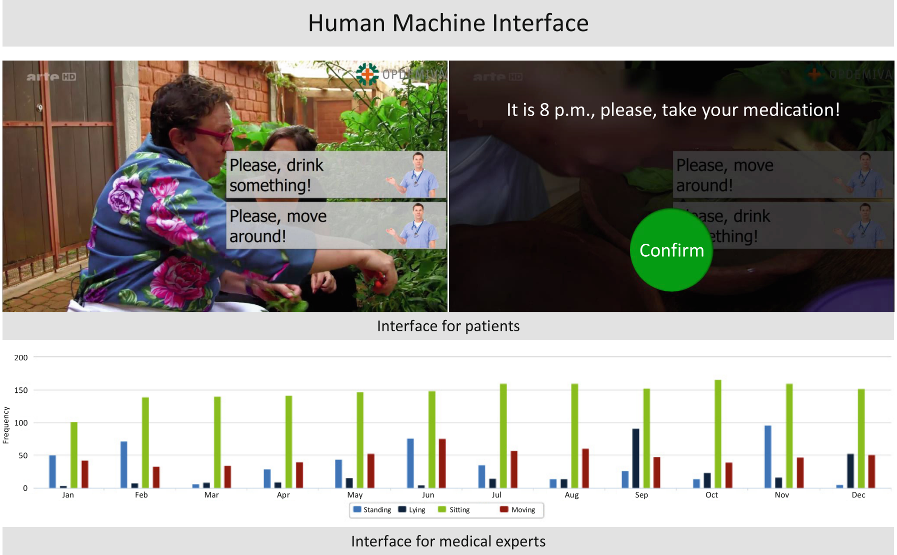


Fig. 4. Feature vector generation from point cloud. All numbers in mm.

Training. After the feature vector generation, a machine was trained in a supervised manner, i. e. with labelled training samples . Video sequences with three different persons (P3, P4 and P7) were recorded for this purpose in a laboratory flat and manually labelled (about 3000 images). Furthermore, a linear Support Vector Machine (SVM) was chosen. The SVM is a *discriminative*, *maximum margin* classifier. The term '*discriminative*' means that the variable to be predicted, i. e. the posterior probability, is modelled whereas '*maximum margin*' refers to the fact that an optimization problem is formulated: A separating hyperplane has to be determined, so that the margin between two adjacent classes is maximized. The outer vectors of the classes form the support vectors. These are the vectors with the minimum distance to the separating hyperplane. We decided to



Human Machine Interface

Please, drink something!

Please, move around!

It is 8 p.m., please, take your medication!

Please, move around!

…ase, drink …thing!

Confirm

Interface for patients

Interface for medical experts

Fig. 5. Design for all: Interface for both patients and for medical personnel. The images at the top show the interface for the patient. Reminders appear time-controlled. The patient can remove the messages either by touching the display (touch screen) or by performing the action, e. g. when movement is detected by the sensor. The image at the bottom shows a graph that presents pose and movements over a month. Other intervals can also be selected.

use this type of classifier, because it ranks among the classifiers with the best performance if the amount of training data is limited [4].

3.4 Human Machine Interface

The medical staff can view the statistically prepared mobility data via a web interface. Additionally, if no movement is detected over a certain period of time, which can be specified beforehand, a reminder appears on a tablet as well as on a touch display. This touch display might be a TV set, so that the person is activated while watching TV, for example. In that way, the person can be immediately addressed in an unobtrusive way. Examples of such scenarios are illustrated in Fig. 5.

4 Experimental Results

In order to determine the performance of the trained pose classifier, we recorded several test sequences. A total number of 2958 samples was classified during the test.

The first test case consisted of realistic scenarios in the laboratory flat with two elderly volunteers (P1 and P2). In the second test case, we attached high importance to the fact that the test sequences had been recorded in a completely different environment compared to the scene where the training sequences have been recorded. We installed, therefore, a test set-up with a stereo camera similar to the one in the laboratory flat. The sequences were recorded with four persons (P3 - P6), of whom two had already participated in the training sequences (P3 and P4).

Table 1 shows the results for the elderly persons in the laboratory flat while Tables 2 and 3 indicate the classification results for both types of test persons in the special test set-up. The letters L and C in the table headings stand for classified pose and labelled pose respectively. All numbers are percentages.

The experiments show that the classification results are of high quality. These first tests also revealed that the algorithm does work reliably in different surroundings and with different persons. The misclassification rate for 'Lying' in Table 1 is obviously very high compared to the other scenarios. This is, however, caused by the sparse and noisy point cloud at the place, where the person was

Table 1. Classification results for persons P1 and P2.

L C	Standing	Sitting	Lying
Standing	100	0	6.5
Sitting	0	100	0
Lying	0	0	93.5

Table 2. Classification results for persons P3 and P4.

L C	Standing	Sitting	Lying
Standing	97.6	0	0
Sitting	0	100	0
Lying	2.4	0	100

Table 3. Classification results for persons P5 and P6.

L C	Standing	Sitting	Lying
Standing	100	0	0
Sitting	0	100	1
Lying	0	0	99

lying at this time. The location was relatively far away from the stereo sensor, so that the stereo matching algorithm reached its limits.

For the purpose of movement evaluation, we recorded and labelled a video sequence, in which persons were either walking through the room or standing somewhere at the spot. We could thus compare the labels with the output of the algorithm (moving/non-moving) and calculate the true-positive rate TPR and the false-positive rate FPR were calculated. $mov_{\text{detected}|\text{neg}}$ denotes the number of frames where movement was detected although the label was non-movement, $mov_{\text{detected}|\text{pos}}$ the number of frames where movement was detected and the label was movement, $mov_{\text{neg,labelled}}$ the number of frames labelled as non-movement, $mov_{\text{pos,labelled}}$ the number of frames labelled as movement.

$$TPR = \frac{mov_{\text{detected}|\text{pos}}}{mov_{\text{pos,labelled}}} = \frac{288}{298} \approx 96.6\,\% \tag{7}$$

$$FPR = \frac{mov_{\text{detected}|\text{neg}}}{mov_{\text{neg,labelled}}} = \frac{5}{193} \approx 2.6\,\% \tag{8}$$

These values show that significant movements between different positions in the monitored area are detected by the algorithm.

5 Action Recognition

Latest developments of our system aim at monitoring and analysing activities important for the need of care of demented persons. Such activities are related to nourishment, social contacts and personal hygiene. On the basis of the registered activities, assistance, such as reminders or activation messages, can be provided for patients. Caring personnel could benefit from the generated information by involving it in their care planning. New advances in our project show that activities, such as drinking, can be reasonably well detected by means of machine learning techniques. Figure 6 shows a feature vector example that has been used for training a machine in order to recognise drinking from a bottle.

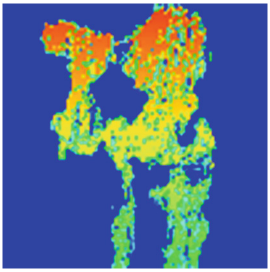


Fig. 6. Feature vector for recognising drinking.

Further activities can be recognised by combining the person's position and pose with context information, for example the knowledge about the location of furniture or of different utensils in the flat. When a person is localised in the bed and at the same time is detected to be lying for a longer time, then it is assumed that the person is sleeping. In addition to sleeping in the bed, actions such as resting in an armchair, taking a shower and washing hands while standing in front of a basin are several examples of activities the system is capable to detect at the moment.

6 Conclusions

In this paper, we presented an approach to measure significant indicators for mobility, i. e. a person's pose and movement. The most significant finding to emerge from this study is that the proposed machine learning technique works reliably in different environments and with different persons. In combination with movement detection (e. g. crossing a room), conclusions about a person's mobility can be drawn. In that way, long-term diagnostics involving mobility observations can lead to more reliable diagnoses of the health status, which will result in a better assessment of the need of care. Moreover, activation and mobilization by means of a HMI can support the demented persons in preserving their functional abilities.

7 Future Work

Further work needs to be done to increase the accuracy of the action registration and to extend the range of detectable activities.

An essential aspect of our future studies will be the conduction of field tests in cooperation with our medical partners. The application of the system in the field over a longer period of time will provide data for a long-term statistical data analysis and for system validation. Since the focus of the presented approach lies on the patient, the HMI has to be attuned to the special needs of demented people, which shall result in a patient-oriented assistance and assessment system.

With regard to the demographic developments, the quality of care for demented people has to be ensured. The proposed approach can contribute to a more valid assessment and to the preservation of the patient's quality of life. Not only would this be of high benefit for our caring sector, but it could also increase the quality of life of demented persons and their relatives.

Acknowledgements. This project was funded by the European Fund for Regional Development (EFRE).

References

1. Agarwal, A., Triggs, B.: Recovering 3D human pose from monocular images. IEEE Trans. Pattern Anal. Mach. Intell. **28**(1), 44–58 (2006)

2. Belbachir, A. N., Litzenberger, M., Schraml, S., Hofstatter, M., Bauer, D., Schon, P., Humenberger, M., Sulzbachner, C., Lunden, T., Merne, M.: CARE: a dynamic stereo vision sensor system for fall detection. In: 2012 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 731–734. IEEE (2012)
3. Berlin Institut für Bevölkerung und Entwicklung: Demenz-Report. http://www.berlin-institut.org/fileadmin/user_upload/Demenz/Demenz_online.pdf (2011). Accessed 07 July 2014
4. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Sebastopol (2008)
5. Clement, J., Ploennigs, J., Kabitzsch, K.: Erkennung verschachtelter ADLs durch Smartmeter. Lebensqualität im Wandel von Demografie und Technik (2013)
6. Ferrari, V., Marin-Jimenez, M., Zisserman, A.: Progressive search space reduction for human pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
7. Haritaoglu, I., Beymer, D., Flickner, M.: Ghost 3d: detecting body posture and parts using stereo. In: Proceedings Workshop on Motion and Video Computing, pp. 175–180. IEEE (2002)
8. Kalfhues, A.J., Hübschen, M., Löhrke, E., Nunner, G., Perszewski, H., Schulze, J.E., Stevens, T.: JUTTA–JUsT-in-Time Assistance: Betreuung und Pflege nach Bedarf. In: Shire, K.A., Leimeister, J.M. (eds.) Technologiegestützte Dienstleistungsinnovation in der Gesundheitswirtschaft, pp. 325–349. Springer, Heidelberg (2012)
9. Kirchner, N., Alempijevic, A., Virgona, A.: Head-to-shoulder signature for person recognition. In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 1226–1231. IEEE (2012)
10. Link, N., Steiner, B., Pflüger, M., Kroll, J., Egeler, R.: safe@ home-Erste Erfahrungen aus dem Praxiseinsatz zur Notfallerkennung mit optischen Sensoren. Lebensqualität im Wandel von Demografie und Technik (2013)
11. Ning, H., Xu, W., Gong, Y., Huang, T.: Discriminative learning of visual words for 3D human pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
12. Richter, J., Findeisen, M., Hirtz, G.: Assessment and care system based on people detection for elderly suffering from dementia. In: IEEE Fourth International Conference on Consumer Electronics, ICCEBerlin 2014, pp. 59–63. IEEE (2014)
13. Scanaill, C.N., Carew, S., Barralon, P., Noury, N., Lyons, D., Lyons, G.M.: A review of approaches to mobility telemonitoring of the elderly in their living environment. Anna. Biomed. Eng. **34**(4), 547–563 (2006). http://link.springer.com/article/10.1007/s10439-005-9068-2
14. Steen, E. E., Frenken, T., Frenken, M., Hein, A.: Functional Assessment in Elderlies Homes: Early Results from a Field Trial. Lebensqualität im Wandel von Demografie und Technik (2013)
15. Ye, M., Wang, X., Yang, R., Ren, L., Pollefeys, M.: Accurate 3d pose estimation from a single depth image. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 731–738. IEEE (2011)
16. Zivkovic, Z.: Improved adaptive gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2, pp. 28–31. IEEE (2004)