Action Recognition in Bed using BAMs for Assisted Living and Elderly Care

Manuel Martinez and Lukas Rybok and Rainer Stiefelhagen
Institute of Anthropomatics and Robotics,
Karlsruhe Institute of Technology
Karlsruhe, Germany
{name.surname}@kit.edu

Abstract

There is a large interest on performing elderly care monitoring using Computer Vision. It has the potential to provide a better scene understanding than current sensing approaches at an affordable price, but there are still considerable practical challenges that have limited its deployment. The BAM descriptor is a privacy-conscious, calibration-free representation of a single-person bed obtained from a depth camera, and thus is very practical for uninterrupted monitoring. It has been used to recognize static and time-invariant phenomena such as sleeping position and agitation with great success. In this work, we explore BAM-based feature representations for higher level scene understanding. To this end, we created a database of 17 actions typical for elderly care which we use to evaluate our approach demonstrating promising results. We hope that this level of high scene understanding would allow the prediction of accidents in elderly care before they happen, instead of triggering an alarm after they happen.

1 Introduction

Most developed countries are currently experiencing strong population ageing, due to increased life expectancy and declining birth rates. Current elderly care mechanisms are being revised to take this phenomena into consideration.

Nursing homes require big personnel crews to provide adequate care 24 hours a day all year-round. With elderly numbers growing, there will be more demand for care providers, when working-age people will become more scarce. The ratio of non-workers to workers is projected to rise from 37% in 2007 to 72% in 2060 [16]. The current situation is not sustainable without dropping significantly the quality of life of our senior citizens.

Several programs have been devised to offer more human alternatives to overcrowded nursing homes. The *Aging in Place* initiative supports the idea of allowing elderly people to age in their current residences. Technological solutions have been devised to enable those who prefer living on their own, being part of the field of Assisted Living.

There is a particular interest focus in technologies able to monitor sleep. In nursery homes, the nurses make periodic rounds during the night to check if the residents are sleeping peacefully or had any problem or accident. Nursing alarms have been devised for the task, but require expensive installations and the large number of false positives makes nurses prone to ignore alarms (alarm fatigue [4]).

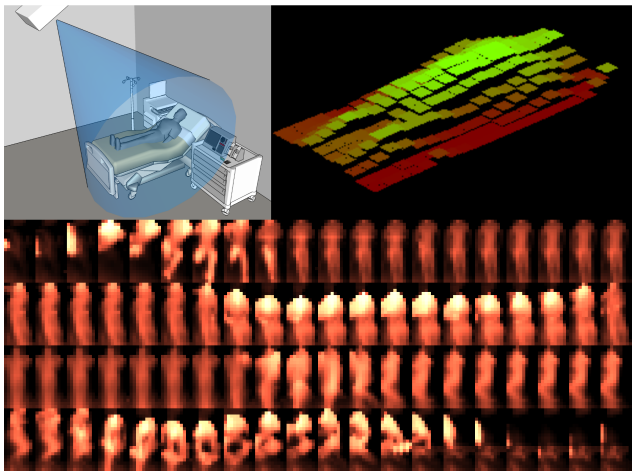

Figure 1: The Bed Aligned Map (BAM) descriptor is obtained non-obtrusively from a camera system mounted above the bed (top left). The bed is automatically detected and flattened, and its area divided in 10 x 20 cells (top right). BAMS offer a synthetic representation that protects the privacy of the user while providing enough information to recognize common actions such as: getting inside the bed (1st row), drinking water (2nd row), change sleeping position (3rd row), and leaving the bed (4th row).

Getting a better knowledge of the behavior of sleeping patients, would provide an inherent advantage over the basic alarm setup. Not only would it increase the accuracy of emergence signals, but also might help to predict accidents by detecting changes in the behavior of the patient.

There is an increasing interest in using computer vision systems as sensors to monitor activity in assisted living homes. Cameras have the advantage that a single system can be used for a wide variety of purposes, from detecting accidents [7], breathing patterns [1, 10, 22], awakeness [13], bed occupancy [11, 12, 15], agitation [5, 6, 11, 17], sleeping position [11], facial stress [3, 9], etc. Furthermore, the live feed provided by the camera can be transmitted to a professional, if needed, to asses the severity of the alarm.

Action recognition approaches [21] offer a high level of abstraction by measuring the patient behavior. Most approaches work are based on motion patterns with space time interest points [8, 19], and recently, if 3d data is available, the reconstructed body pose is also used [14, 18]. However, image processing algorithms are significantly complex, and due to the uncontrolled environment that is a nursery home, less robust.

The BAM [11] was designed to offer a robust, calibration free, illumination independent, and privacy-conscious representation of a bed in intensive care.

In this work, we explore the viability of the BAM representation to perform action recognition in elderly care scenarios. Our results show promising performance, and open the door to future developments such as automated behavioral reporting, and intelligent alarm systems where accidents in elderly care can be predicted instead of detected.

2 BAM Action Dataset

In order to evaluate the BAM feature representation for action recognition, we extend the *"BAM!"* dataset provided by Martinez *et al.* [11]. This dataset captures several volunteers being asked to perform a series of actions around a bed (*e.g.* getting inside the bed, leaving the bed, change sleeping positions, interact with a nurse, manipulate objects, etc.). Although the dataset was captured using a depth camera above the bed, the data is stored using the BAM descriptor. BAM stands for Bed Aligned Map, and is constructed by locating the bed within the depth image and flatten it to remove the variability produced by articulated beds. Then the bed is divided in a 20x10 grid and the average depth on each cell is stored (see Fig. 1).

Being a low resolution descriptor, it is not possible to identify who is appearing in the images (patients or visitors), and controversial topics such as nakedness are also not an issue. Therefore, not being affected by most privacy problems, BAM allows uninterrupted and unattended patient monitoring.

The dataset provides BAMs at 10 frames per second for 21 persons. Actions are not labeled, but the timestamps of the instructions given to the participants are also included.

2.1 Action Dataset Description

From the original *"BAM!"* dataset we have created an action dataset to evaluate action recognition [1] Due to privacy reasons, the original image data is not available.

To build our dataset, we analyzed all the instructions that the volunteers received and selected the ones that could correspond to a recognizable action or activity. For those selected instructions, we segmented the 10 seconds following the delivery of the instruction. As we have 10 BAMs per second, each segment is 100 BAMs long. We clustered the instructions that belong to basically the same action. We ended up with 17 actions (see Table 1).

Early during our research, we observed that there is a clear hierarchical overlap between actions. For example, changing bed positions or leaving the bed implies some sort of agitation, and at the higher level, interacting with a nurse usually implied a change in the resting position. To deal with this overlap of actions, we split them in three different problem sets: low level agitation (Table 2), bed movement(Table 3), and high level actions (Table 4). Each of the problems are evaluated separately.

[1]The dataset is available at:
https://cvhci.anthropomatik.kit.edu/~manel/sphere

Instruction	Action
Get into bed (eyes open)	Get Into Bed
Get into lateral right (open)	Supine To Right
Get into lateral left (open)	Right To Left
Leave the bed (eyes open)	Leave Bed
Nurse arrives and speak	Nurse Arrives
Nurse manipulates infusion lines	Nurse Manip.
Nurse leaves	Nurse Leaves
Agitate a little (supine)	Agitate Low
Agitate slightly more (supine)	Agitate Med
Agitate a lot (supine)	Agitate High
Agitate a little (fetal)	Agitate Low
Agitate slightly more (fetal)	Agitate Med
Agitate a lot (fetal)	Agitate High
Perform repetitive movements	Repetitive Mov.
Open and close fists several times	Repetitive Mov.
Shout aggressively	Shouting
Remove bed cover and cover again	Bed Cover Manip.
Manipulate the infusion lines	Inf. Lines Manip.
Touch mouth with your hand	Touch Mouth
Get the cup and drink water	Drink Water
Get a dangerous item and manipulate it	Manipulate Object
Get into bed (eyes closed)	Get Into Bed
Get into lateral right(closed)	Supine To Right
Get into lateral left (closed)	Right To Left
Leave the bed (eyes closed)	Leave Bed

Table 1: On the left, there are the instructions given to the volunteers, and on the right side there are the corresponding actions we associate to each instruction. Several instructions trigger an equivalent action.

Action	Sequences
Agitate Low	42
Agitate Med	42
Agitate High	42

Table 2: Low Level Actions.

Action	Sequences
Get Into Bed	42
Supine To Right	42
Right To Left	42
Leave Bed	42

Table 3: Mid Level Actions.

Action	Sequences
Nurse Arrives	21
Nurse Manipulates	21
Nurse Leaves	21
Repetitive Movements	42
Shouting	21
Bed Cover Manipulation	21
Infusion Lines Manipulation	21
Touch Mouth	21
Drink Water	21
ManipulateObject	21

Table 4: High Level Actions.

3 Activity Recognition

We encode the appearance of the scene, by directly using BAMs that are provided with the dataset used for the evaluation [11]. In order to also capture motion, we further calculate the difference between BAMs in subsequent frames (further denoted as dBAM). Whole action sequences are encoded as a bag-of-words (BoW), which has been shown to yield state-of-the-art results in many action recognition tasks (*e.g.* [19]). To this end, we first learn a 1000-word codebook via k-means clustering. Then we either apply Vector Quantization (VQ) or Locality-constrained Linear Coding (LLC) [20] with sum-pooling to obtain a BoW representation of each sequence. Since, it has been shown that power normalization can increase the discriminative power of a feature vector [2], we first normalize the features to unit length and then raise each element of the feature vector to the power of 0.3. Finally, we standardize the features to zero-mean and unit-variance before using a linear multi-class SVM for action recognition. In our experiments we observed that features based on BAM and dBAM appear to have complementary properties and thus we also fuse BAM and dBAM features by concatenating their BoW encoding.

Feature	Encoding	high	med	low
BAM	VQ	52.4	92.1	50.0
BAM	LLC	62.8	93.4	**57.1**
dBAM	VQ	63.6	79.4	50.8
dBAM	LLC	66.2	86.1	48.4
dBAM	VQ	61.9	94.5	50.0
BAM+dBAM	VQ	62.0	94.5	50.0
BAM+dBAM	LLC	**67.5**	**97.0**	55.6

Table 5: Overall classification accuracy.

4 Experimental Results

The results on the low level actions are poor but expected (see Fig. 2). Our action recognition framework works better on very specific and non overlapping actions, and it had difficulties with the large variety of movements involved.

However the results are reasonably close to the ones published in [11], which makes us think that with more training data our results could be better than their ad-hoc solution.

The mid-level actions are clear and distinctive, and therefore we achieved an accuracy of 97.0% (see Fig. 3).

The results on the high level actions show a split. Reasonably unique and well defined actions, such as nurse interactions, manipulating of the bed cover, and the infusion lines, achieve great results. However, there is not enough training data to model accurately weakly described actions such as "repetitive movements". Finally, the BAM representation is probably not strong enough to distinguish clearly between grabbing an object and manipulating it, and grabbing a cup and drinking water from it, and therefore there is significant confusion between those two classes (see Fig. 4).

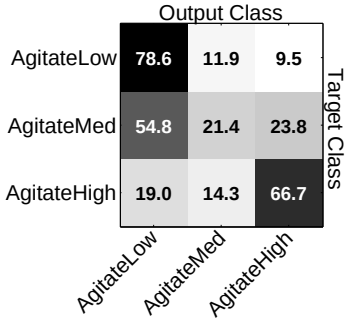

Figure 2: Confusion matrix for three different agitation levels for LLC encoded BAM+dBAM (total accuracy of 55.6%). It is clear that the action recognition framework works better for clearly defined actions.

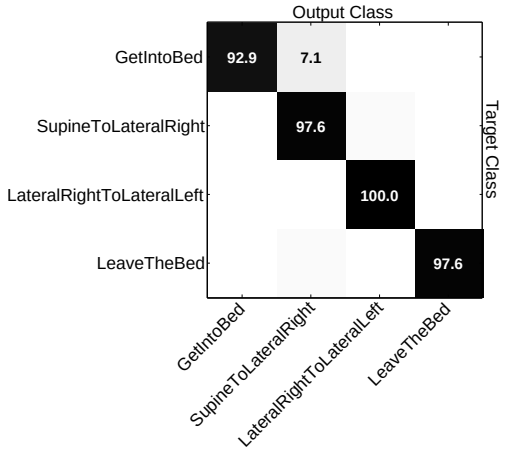

Figure 3: Confusion matrix of the bed related actions for LLC encoded BAM+dBAM. Those actions are clearly defined, and easily recognized with an accuracy of 97.0%.

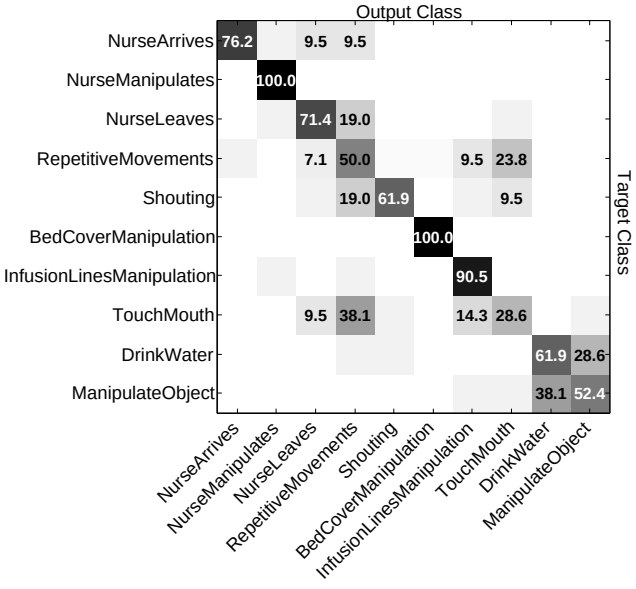

Figure 4: Confusion matrix of the higher level actions for LLC encoded BAM+dBAM (overall accuracy of 67.5%). We can observe a split behavior where some actions can be detected reliably, while others perform very poorly. It is noteworthy that the Drink Water and Manipulate Object actions, which are very similar, are mostly only confused between themselves.

5 Conclusions and Future Work

To the best of our knowledge, we have shown the first study on action recognition in a bed using computer vision. Using the privacy-conscious BAM descriptor, we can provide uninterrupted monitoring of patients with a modest computing platform, easing the path of integrating the system into a real product.

Our suggested framework obtain robust results while recognizing changes in the sleeping position with over 95% recognition rate. Our system perform poorly for basic actions (*i.e.* agitation), as it is a very subjective topic that every subject interpreted in a different way, and thus the variance was too big for the reduced number of training samples available.

On the more interesting high level actions, (*e.g.* manipulating bed covers, drink water, etc.) we got promising results, specially on very specific actions.

We plan to explore three strategies for improving the presented results. First, if more data is available, Convolutional Neural Networks might be trained in order to derive optimal descriptors from the data. Second, by developing a body-part detection on BAM in order to provide high-level information to the model. Third, using a hierarchical model for action recognition and thus splitting actions into subcomponents, to avoid similar actions to compete against each other.

Finally, we plan to evaluate action detection on non-segmented sequences, so that we can create automated activity reports of sleeping patients. We hope that those reports can be used to improve the quality of sleep of our senior citizens, and help preventing illnesses and possible accidents.

Acknowledgements: This work is supported by the German Federal Ministry of Education and Research (BMBF) within the SPHERE project.

References

[1] H. Aoki, Y. Takemura, K. Mimura, and M. Nakajima, *Development of non-restrictive sensing system for sleeping person using fiber grating vision sensor*, Micromechatronics and Human Science, 2001.

[2] Relja Arandjelovic and Andrew Zisserman, *Three things everyone should know to improve object retrieval*, CVPR, 2012.

[3] P. Becouze, C.E. Hann, J.G. Chase, and G.M. Shaw, *Measuring facial grimacing for quantifying patient agitation in critical care*, Computer Methods and Programs in Biomedicine, 2007.

[4] Linda Bell, *Monitor alarm fatigue*, American Journal of Critical Care **19** (2010), no. 1, 38–38.

[5] J. Geoffrey Chase, F. Agogue, C. Starfinger, Z.H. Lam, G.M. Shaw, A.D. Rudge, and H. Sirisena, *Quantifying agitation in sedated icu patients using digital imaging*, Computer Methods and Programs in Biomedicine, 2004.

[6] J Geoffrey Chase, Franck Agogue, Christina Starfinger, ZhuHui Lam, Geoffrey M Shaw, Andrew D Rudge, and Harsha Sirisena, *Quantifying agitation in sedated icu patients using digital imaging*, Computer methods and programs in biomedicine, 2004.

[7] P. Kittipanya-Ngam, O.S. Guat, and E.H. Lung, *Computer vision applications for patients monitoring system*, Information Fusion (FUSION), 2012.

[8] Ivan Laptev, Marcin Marszaek, Cordelia Schmid, and Benjamin Rozenfeld, *Learning realistic human actions from movies*, CVPR, 2008.

[9] M.N. Mansor, S. Yaacob, R. Nagarajan, L.S. Che, M. Hariharan, and M. Ezanuddin, *Detection of facial changes for icu patients using knn classifier*, Intelligent and Advanced Systems (ICIAS), 2010.

[10] M. Martinez and R. Stiefelhagen, *Breath rate monitoring during sleep using near-ir imagery and pca*, International Conference on Pattern Recognition (ICPR), 2012.

[11] Manuel Martinez, Boris Schauerte, and Rainer Stiefelhagen, *bam! depth-based body analysis in critical care*, Computer Analysis of Images and Patterns, Springer, 2013, pp. 465–472.

[12] Manuel Martinez and Rainer Stiefelhagen, *Automated multi-camera system for long term behavioral monitoring in intensive care units*, MVA, 2013.

[13] M. Naufal Bin Mansor, S. Yaacob, R. Nagarajan, and M. Hariharan, *Patient monitoring in icu under unstructured lighting condition*, Industrial Electronics & Applications (ISIEA), 2010.

[14] Eshed Ohn-Bar and Mohan M. Trivedi, *Joint Angles Similarities and HOG2 for Action Recognition*, CVPR Workshops, 2013.

[15] Optex, *Secnurse*, `optex.nl/caredetection/`, [Online; accessed 20-December-2014].

[16] Bernd Rechel, Emily Grundy, Jean-Marie Robine, et al., *Ageing in the european union*, The Lancet **381** (2013), no. 9874, 1312–1322.

[17] Miguel Reyes, Jordi Vitria, Petia Radeva, and Sergio Escalera, *Real-time activity monitoring of inpatients*, MICCAT, 2010.

[18] Raviteja Vemulapalli, Felipe Arrate, and Rama Chellappa, *Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group*, CVPR, 2014.

[19] Heng Wang, Alexander Klaser, Cordelia Schmid, and Cheng-lin Liu, *Action Recognition by Dense Trajectories*, CVPR, 2011.

[20] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong, *Locality-constrained Linear Coding for Image Classification*, CVPR, 2010.

[21] Daniel Weinland, Remi Ronfard, and Edmond Boyer, *A Survey of Vision-Based Methods for Action Representation , Segmentation and Recognition*, CVIU **115** (2011), no. 2, 224–241.

[22] Meng-Chieh Yu, Huan Wu, Jia-Ling Liou, Ming-Sui Lee, and Yi-Ping Hung, *Multiparameter sleep monitoring using a depth camera*, Biomedical Engineering Systems and Technologies, Springer, 2013, pp. 311–325.