Motion2Vector: Unsupervised Learning in Human Activity Recognition Using Wrist-Sensing Data

Lu Bai*
l.bai@ulster.ac.uk
Ulster University
Belfast, UK

Chris Yeung
Shearwater Systems Ltd
Canterbury, UK
chris.yeung@shearwatersystems.com

Christos Efstratiou
University of Kent
Canterbury, Kent
c.efstratiou@kent.ac.uk

Moyra Chikomo
University of Kent
Canterbury, Kent
mc810@kent.ac.uk

ABSTRACT

With the increasing popularity of consumer wearable devices augmented with sensing capabilities (smart bands, smart watches), there is a significant focus on extracting meaningful information about human behaviour through large scale real-world wearable sensor data. The focus of this work is to develop techniques to detect human activities, utilising a large dataset of wearable data where no ground truth has been produced on the actual activities performed. We propose a deep learning variational auto encoder activity recognition model - Motion2Vector. The model is trained using large amounts of unlabelled human activity data to learn a representation of a time period of activity data. The learned activity representations can be mapped into an embedded activity space and grouped with regards to the nature of the activity type. In order to evaluate the proposed model, we have applied our method on public dataset - The Heterogeneity Human Activity Recognition (HHAR) dataset. The results showed that our method can achieve improved result over the HHAR dataset. In addition, we have collected our own lab-based activity dataset. Our experimental results show that our system achieves good accuracy in detecting such activities, and has the potential to provide additional insights in understanding the real-world activity in the situations where there is no ground truth available.

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09…$15.00
https://doi.org/10.1145/3341162.3349335

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing**.

KEYWORDS

human activity recognition, deep learning, variational auto encoder

ACM Reference Format:
Lu Bai, Chris Yeung, Christos Efstratiou, and Moyra Chikomo. 2019. Motion2Vector: Unsupervised Learning in Human Activity Recognition Using Wrist-Sensing Data. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3341162.3349335

1 INTRODUCTION

With the increasing popularity of consumer wearable devices augmented with sensing capabilities (smart bands, smart watches), there is a significant focus in extracting meaningful information about human behaviour through large scale real-world wearable sensor data [13]. Human activity recognition (HAR) through wearable devices is recently considered a vital tool for future healthcare applications, especially in support for elderly people and patients with certain long-term conditions [2–4]. HAR can help patients with continuous care and rehabilitation needs at home and provides clinicians with additional insight of the patients' performance.

Traditional approaches in developing HAR systems, rely on the collection of properly labeled training datasets, where the actual activities of the users are captured accurately either through controlled experiments, or through self reports [1, 12, 14]. These approaches however suffer from scalability issues as the process of collecting ground truth through self-report cannot be employed on a large scale and over

long periods of time. In order to progress with the wider application of HAR systems there is a need for developing techniques were human activity classifiers can be developed with minimum need for accurate ground truth from the participants.

In this work we aim to explore the development of an unsupervised model to infer what people do from wearable sensor data. Our focus is on the use of large scale wearable datasets, which do not contain any prior labeling of the activities that users perform. Interpreting and using such data imposes significant challenges in developing appropriate human activity recognition techniques, without the need for accurately labelling data to produce a training dataset.

Specifically, we focus on the analysis of a large dataset of activities captured through wearable wrist-bands (Microsoft Band 2) as part of the Innovate UK funded project "Epilepsy Networks". The dataset was produced by a cohort of 37 patients suffering from epilepsy, using wearable wrist-bands during their daily lives, for a long period of time (approximately 6500 days in total, with an average of 112 days per participant). The dataset consists of raw accelerometer, gyroscope readings along with physiological data (i.e. heart rate) as captured by the wearable device. During the deployment the participants were not required to submit any ground truth about their daily activities.

Our aim is to develop a technique to detect the daily activities of participants through this unlabeled dataset. To do this we propose an approach based on an auto-encoder deep learning model, called Motion2Vector, which is developed through unsupervised training on this large dataset. The purpose of the model is to convert a time period of activity data into a movement vector embedding within a multidimensional space, as a representation of a certain activity type. That representation helps us group similar activities together within the embedded space. The technique can help identify when and for how long similar activities take place, but the actual meaning/context for such activities requires limited knowledge of ground truth. We evaluate the approach through public dataset - HHAR and our own lab-based activity dataset. The core contribution of this work includes:

- We propose a variational autoencoder (VAE) deep learning technique to train a model using a large wrist-band dataset of real-world activities, without labelling. The proposed model enables movement embedding utilising the raw input from wearable sensor data.
- We deploy our data collection system and collect the datasets in lab-based session in order to evaluate our trained model. The lab-based session is accurately labelled by the researcher who is carrying out the experiment.
- We validate our trained model on both public datasets and our collected datasets.

2 RELATED WORK

Significant work in HAR focuses on the use of inertial sensor to detect human activities [10, 16, 18]. This is motivated primarily by the wide availability of such sensors on consumer devices such as smartphones and smart watches. Most of these systems rely heavily on well labeled training datasets, generated mostly through controlled lab experiments. The development and training of appropriate machine learning classifiers typically involves the extraction of hand-crafted features from the raw data before being used to train an appropriate classifier. Manual extraction of features from raw data typically requires appropriate domain knowledge on the type of activity that should be detected, or would require extensive human observation.

With the increased popularity in deep learning models in machine learning, there has been also a shift in applying deep learning techniques in HAR [13]. In [17], a CNN model is used to automate the feature extraction under the supervision of output labels. [5], a LSTM-RNN model is created in order to explore the time dependencies of the human motion data. Supervised deep learning models can save development time in creating the appropriate features, however, it still requires accurately labelled datasets to train the model. Semi-supervised models aim to integrate the supervised learning and unsupervised learning. They generally rely on small amounts of labelled data [6, 8]. However, in situations where the collection of ground truth is not realistic or possible, semi-supervised models are still not feasible solutions.

When considering scenarios where sensor data has been collected already, but without any prior ground truth, exploring unsupervised learning techniques is the only viable approach. In unsupervised learning there is no requirement for labelled data. The purpose is to find hidden patterns within the data, and identify groups of similar activities [11]. Evaluating an unsupervised model is challenging when no labeled data is available. In our work, we exploit a small set of labeled activity dataset as part of the evaluation of the produced model.

3 DEEP LEARNING ARCHITECTURE

Our objective is to develop a HAR system that is trained on a large unlabeled dataset of sensor data captured by smart wrist bands. The purpose of the HAR system is to be able to group similar activities together. In order to achieve this we first need to train a model to encode raw inertial sensor data into a vector that represents the movement characteristics captured by the sensor data.

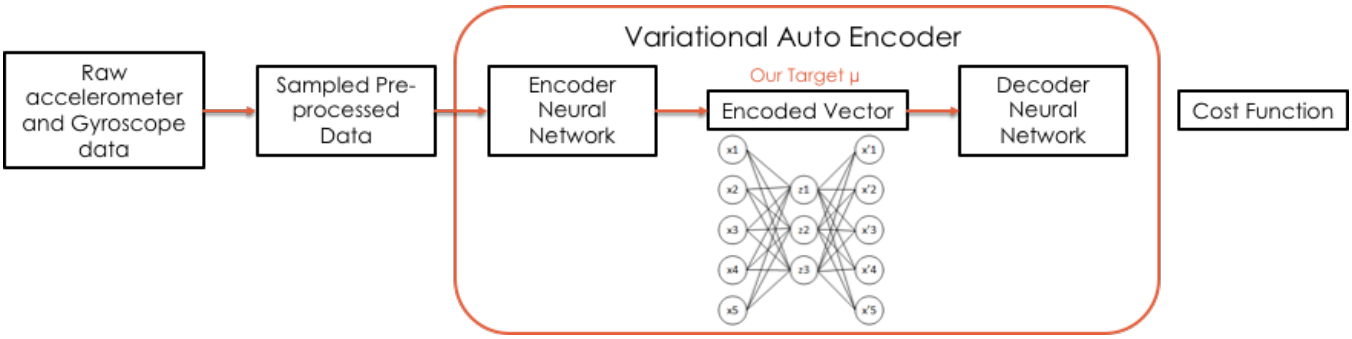

Figure 1: Deep learning architecture

Model Overview

The deep learning model used in this study is inspired by the variational autoencoder proposed by Hu et.al [7] which is applied in generating sketch drawings in a vector format using a recurrent neural network (RNN). In our work, we have made certain modifications to make the deep learning architecture fit with the context of our problem. Figure 1 shows this deep learning architecture. Raw sensor data is pre-processed and fed into an encoder neural network. The hidden layer is N-dimensional encoded vector which is used as the representation of the activity. The decoder neural network decodes the encoded vector and creates an output which has the exact same size as the input. The cost function is trying to minimize the difference between the input and output. The details of the deep learning network are described in the following subsections. After the model has been trained, the encoded vector is used as a good representation of the blocks of activity in the embedded space.

Input Data Preparation

The input data for the training model, and the validation dataset, consist of raw accelerometer and gyroscope data captured from the wearable wrist band. These datasets are pre-processed as described below.

Accelerometer and Gyroscope Sensors. In processing the sensor data, our aim is to convert every sensor data point with raw accelerometer and gyroscope data, into a data point that represents the relative change of the position and orientation of the sensing device. Essentially each data point is to be converted into a vector $M=(\Delta P_x, \Delta P_y, \Delta P_z, g_x, g_y, g_z)$ where $\Delta P_{x,y,z}$ represent the change in position, and $g_{x,y,z}$ represent change in orientation. Activity within a short time period is composed of a set of points.

Calculating the relative change of position. Extracting gravity from the raw data is used to identify a global reference frame. Extracting gravity from accelerometer data can be done using a low-pass / moving average filter: $\Delta G = G \cdot \alpha + (1 - \alpha) \cdot Acc_{input} \cdot (x)$

The relative change of position is to calculate the position relative to the previous position with respect to the global reference frame. In order to calculate the 3D relative position, the first step is to compute the linear acceleration from raw acceleration and the second step is to double integrate the linear acceleration. The input data to the auto encoder consists of a time window of multiple sensor data points (M_1, M_2,..., M_n).

Auto Encoder

The auto encoder consists of an encoder model that "compresses" the input data into a vector representation, and a decoder that uses the generated vector to "decompress" the data to its original form. The auto encoder is trained using a cost function that evaluate how well the encoding-decoding process works.

For our encoder, we use a bidirectional Long Short-Term Memory (LSTM) to encode the input blocks of pre-processed data.

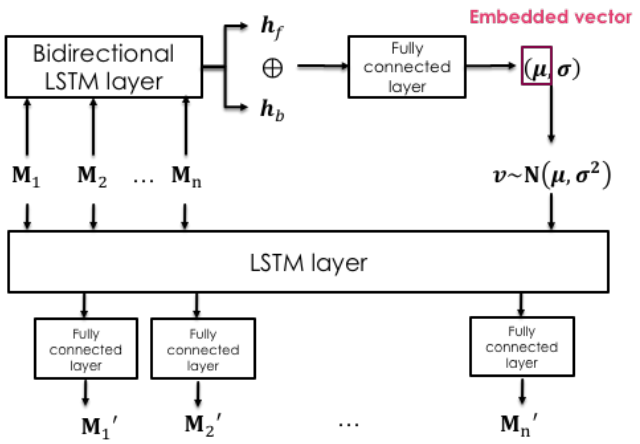

Figure 2: bidirectional LSTM

As seen in Figure 2, two hidden states h_f and h_b are generated. Then the concatenation of these hidden states will be used to generate two embedded vectors through a fully connected layer.

The embedded vector μ is our target in this study, it is a 128 dim vector and it is considered an appropriate representation of the input movement within a particular time period.

As part of the training process, the μ vector is used by the decoder which is responsible for regenerating the input data. In order to prepare the inputs for the decoder, a random vector v is created using μ and δ following a unit normalised distribution. The decoder takes the created vector v and all the inputs to generate the outputs.

Cost function and training details

Our model is trained through a combination of two parts of loss - content loss and KL-divergence loss [9]. As described in Equation 1, the content loss is to minimize the difference between the input M and output M'.

$$L_c = \frac{1}{N}\sum_{i=1}^{N}(M' - M)^2 \tag{1}$$

The KL divergence loss is described in the Equation 2 below:

$$L_{KL} = -\frac{1}{2N}\sum_{i=1}^{N}(1 + ln\sigma_i - \mu_i^2 - \sigma_i) \tag{2}$$

The final loss is the optimisation of the content loss and the KL-divergence loss (Equation 3).

$$L = L_c + \gamma L_{KL} \tag{3}$$

Annealing technique is used to adjust the latent lost: $0.1<\gamma<1$. The training result is shown in Figure 3. In each subplot, the x-axis is the number of the training iterations and y-axis is the value of training loss.

4 EVALUATION

Training Dataset

The training data has been captured as part of an Innovate UK funded project. The aim of the project was to provide support for people suffering from epilepsy, and it included a large deployment of wrist-worn devices to patients. During the study, each of the participants was given a Microsoft Band 2 and installed a mobile app able to log data from their wearable device. The app was also designed to collect the patients' self-reports on potential epileptic seizures; however, no other ground truth was captured regarding the daily activities of the participants. Wristband sensor data has been collected on 37 patients from May 2017 until August 2018.

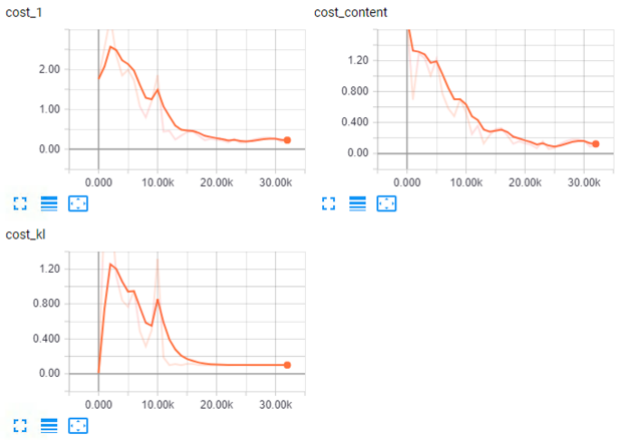

Figure 3: Loss graph for our model trained using our large cohort of data

Sensor data included tri-axis accelerometer and tri-axis gyroscope data, sampled at 31Hz. In this study, we use this dataset as our training datasets.

Figure 4: The system structure of the wearable data collection system

Evaluation Dataset

In evaluating our trained model, we needed wearable data with ground truth so as to evaluate the accuracy of the trained model. We relied on a combination of public HAR datasets, and in lab experiments.

Public Dataset. In order to evaluate our model, we selected the public HHAR dataset [15], a dataset collected from 9 participants. It contains data from 4 different models of smartwatches (2 LG watches, 2 Samsung Galaxy Gears). The sensor readings contain both accelerometer and gyroscope.

Data collection. Most of the public available datasets are focused in collecting general motion of the human during walking, running, cycling, sitting, etc. However, since our

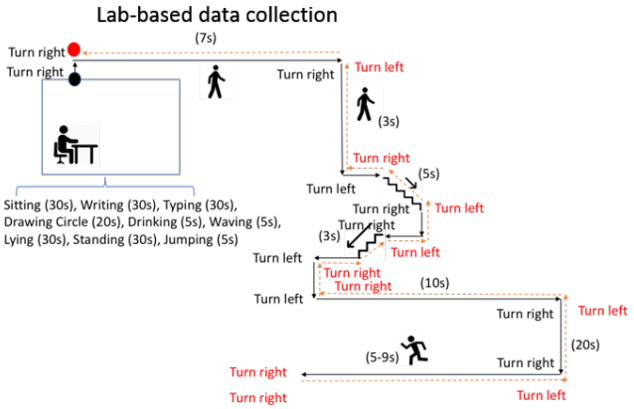

Figure 5: Data collection details for lab-based data collection

Table 1: The F1 Performance of different Datasets

Classifier	Public Dataset	Our lab dataset
C4.5	86.91 %	84.63%
KNN	90.21 %	75.73 %
Random Forest	91.54%	88.75%

wrist sensor is worn by the participant all the time, we are able to track some finer motion of the hand activity including drinking, waving hand, typing, etc. In order to explore a wider range of activities we generated our own labeled data, through a controlled experiment. 10 participants were invited for a 15 mins lab-based session where they were asked to perform certain tasks wearing a wristband.

The lab-based session experiment was carried out in a research room at the University of Kent. During this experiment session, the participants were asked to complete a range of different human daily activity including walking, running, typing, writing, etc. Figure 5 shows the experimental flow of the lab-based session. The ground truth was captured by the pad operated by the observer researcher. The app developed for capturing the activity was able to record the start and end time of each of the specific activities.

Evaluation Methods

For each dataset, through applying the trained model, we have gained 128 dimensional encoded vectors. One method to evaluate the performance is to use 128 dimension vectors as features. Three different classifiers have been used and the classification results are presented using F1 score. A comparison has been made with other researcher's work on the same dataset using hand crafted features.

Table 2: Activity recoginition Performance based on embedding dictionary

Performance	Public Dataset	Our lab dataset
Precision	87%	73%
Recall	88 %	73%
F1-Score	87%	72 %

Additionally, another method based on the Euclidean distance have been proposed to evaluate the results of the Embedding. 30% of data is used to create the word embedding dictionary, and the Euclidean distance is calculated using Equation (4) below for the distance between the test dataset and the embedding dictionary. The activity predicted on the test dataset is estimated as the activity of the closest vector in terms of Euclidean distance.

$$D = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4}$$

Results

Results Visualisation. In Figure 6, a visualisation of the Embedded space has been presented. Different activities are grouped together in the space and from the left to right, the activity type is less active. t-NSE is applied to reduce the dimenstions and visulise the 128 dimension vectors in the embeded space.

Classifier Validation. For both datasests, we have implement the first evaluation method using the embedded 128 dimension vectors as the features. We train 3 different classifiers and the performance F1 score is presented in Table 1. It is noted that the classifiers trained by using the embedded vectors as the features outperform the standard approaches with the handcrafted features from a previous study where the F1 scores of KNN and Random Forest are all below 90% while the F1 score of C4.5 is less than 85%. The F1 score of the lab-based dataset is slightly lower than that of the Public Dataset due to the fact that our collected lab-based dataset is much more noisy and contains more activities.

Euclidean Distance Validation. For both datasets, we have also implemented the evaluation using the Euclidean based method as described in the above Evaluation Methods section. The results of the performance are presented in Table 2 for two different datasets.

5 CONCLUSIONS

In this paper, we present a deep learning model for unsupervised activity recognition. The model has been trained using a large dataset from epileptic patient activity data. The

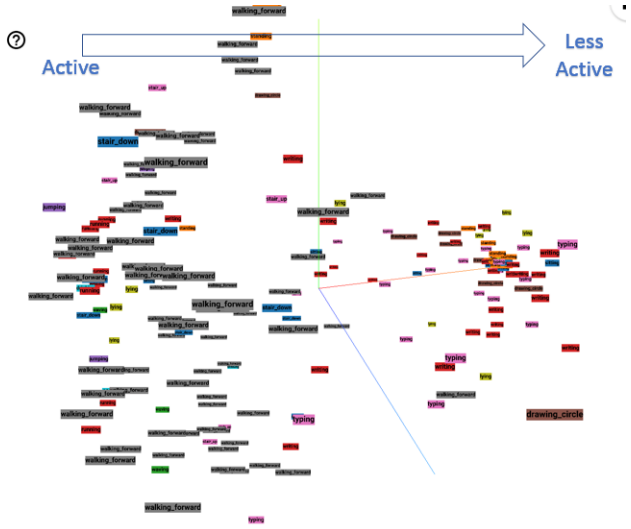

Figure 6: Visualisation of resutls in Embedded Space

experiments on public datasets and our collected datasets demonstrate the proposed model performance. In particular when limited labeled data is available, our model can achieve higher performance than traditional classification techniques, using hand-crafted features. In a fully unsupervised mode, our model can achieve accuracy of higher than 87% when tested on public datasets.

ACKNOWLEDGMENTS

This research was partially supported by Innovate UK (KTP-10829)

REFERENCES

[1] Ling Bao and Stephen S Intille. 2004. Activity recognition from user-annotated acceleration data. In *International conference on pervasive computing*. Springer, 1–17.

[2] Franco Cicirelli, Giancarlo Fortino, Andrea Giordano, Antonio Guerrieri, Giandomenico Spezzano, and Andrea Vinci. 2016. On the design of smart homes: A framework for activity recognition in home environment. *Journal of medical systems* 40, 9 (2016), 200.

[3] Christian Debes, Andreas Merentitis, Sergey Sukhanov, Maria Niessen, Nikolaos Frangiadakis, and Alexander Bauer. 2016. Monitoring activities of daily living in smart homes: Understanding human behavior. *IEEE Signal Processing Magazine* 33, 2 (2016), 81–94.

[4] Bruce H Dobkin. 2017. A rehabilitation-internet-of-things in the home to augment motor skills and exercise training. *Neurorehabilitation and neural repair* 31, 3 (2017), 217–227.

[5] Marcus Edel and Enrico Köppe. 2016. Binarized-blstm-rnn based human activity recognition. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 1–7.

[6] Donghai Guan, Weiwei Yuan, Young-Koo Lee, Andrey Gavrilov, and Sungyoung Lee. 2007. Activity recognition based on semi-supervised learning. In *13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2007)*. IEEE, 469–475.

[7] David Ha and Douglas Eck. 2017. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477* (2017).

[8] HM Hossain, MD Al Haiz Khan, and Nirmalya Roy. 2018. DeActive: scaling activity recognition with active deep learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 66.

[9] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[10] Narayanan C Krishnan and Sethuraman Panchanathan. 2008. Analysis of low resolution accelerometer data for continuous human activity recognition. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 3337–3340.

[11] Yongjin Kwon, Kyuchang Kang, and Changseok Bae. 2014. Unsupervised learning for human activity recognition using smartphone sensors. *Expert Systems with Applications* 41, 14 (2014), 6067–6074.

[12] David Minnen, Thad Starner, Jamie A Ward, Paul Lukowicz, and G Troster. 2005. Recognizing and discovering human actions from on-body sensor data. In *2005 IEEE International Conference on Multimedia and Expo*. IEEE, 1545–1548.

[13] Henry Friday Nweke, Ying Wah Teh, Mohammed Ali Al-Garadi, and Uzoma Rita Alo. 2018. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications* 105 (2018), 233–261.

[14] Ekaterina H Spriggs, Fernando De La Torre, and Martial Hebert. 2009. Temporal segmentation and activity classification from first-person sensing. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 17–24.

[15] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. 2015. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 127–140.

[16] Shuangquan Wang, Jie Yang, Ningjiang Chen, Xin Chen, and Qinfeng Zhang. 2005. Human activity recognition with user-free accelerometers in the sensor networks. In *2005 International Conference on Neural Networks and Brain*, Vol. 2. IEEE, 1212–1217.

[17] Ming Zeng, Le T Nguyen, Bo Yu, Ole J Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. 2014. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services*. IEEE, 197–205.

[18] Jakob Ziegler, Henrik Kretzschmar, Cyrill Stachniss, Giorgio Grisetti, and Wolfram Burgard. 2011. Accurate human motion capture in large areas by combining IMU-and laser-based people tracking. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 86–91.