Activity Recognition With Multiple Wearable Sensors for Industrial Applications

Adrien Malaisé, Pauline Maurice, Francis Colas, François Charpillet and Serena Ivaldi

Inria, Villers-lès-Nancy, F-54600, France
Université de Lorraine, Loria, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France
CNRS, Loria, UMR 7503, Vandoeuvre-lès-Nancy, F-54506, France
Email: firstname.surname@inria.fr

***Abstract*—In this paper, we address the problem of recognizing the current activity performed by a human operator, providing an information useful for automatic ergonomic evaluation for industrial applications. While the majority of research in activity recognition relies on cameras observing the human, here we explore the use of wearable sensors, which are more suitable in industrial environments. We use a wearable motion tracking suit and a sensorized glove. We describe our approach for activity recognition with a probabilistic model based on Hidden Markov Models, applied to the problem of recognizing elementary activities during a pick-and-place task inspired by a manufacturing scenario. We show that our model is able to correctly recognize the activities with 96% of precision if both sensors are used.**

Keywords–Activity recognition; Hidden Markov Model; Wearable sensors.

I. INTRODUCTION

In developed countries, work-related musculoskeletal disorders (MSDs) are a major health issue. MSDs affect almost 50 % of industrial workers and represent an important cost for companies [1]. In order to reduce the prevalence of work-related MSDs, the ergonomics of the workplace needs to be evaluated and improved. Standards ergonomic assessment methods rely on pen-and-paper worksheets filled by experts, such as the commonly used European Assembly Worksheet (EAWS) [2][3]. Some digital human modeling software provide automatic filling of these ergonomic worksheets [4], but this cannot be done directly from raw data (video, motion capture, *etc*.) because the scoring system depends on the activity that is being performed (*e.g.*, walking, bending, carrying an object). The software user has first to manually identify the different types of movements occurring in the task. Therefore, there exists no tool to inform a worker in real-time whether s/he is performing a task in an ergonomic way or not. Yet, such an evaluation could help reducing the risk of MSDs. This is one of the objectives of the European AnDy project [5].

The first step towards a fully automatic ergonomic assessment is to automatically identify the different activities within an industrial task. To address this problem, this paper proposes a method based on wearable sensors and Hidden Markov Model (HMM). We focus our activity recognition on a pick-and-place task inspired by a manufacturing scenario. The whole-body motions of the operator are recorded with inertial sensors embedded in a suit. Though motion-capture based activity recognition using HMM already exists [6][7], the recognition rate is not yet perfect and could be improved. This can be an issue for industrial applications. Therefore, we propose to complement the motion capture system with a glove embedding force sensors to detect hand contacts with the manipulated objects. This paper focuses on evaluating the benefit of using the contact information for the recognition performance with HMM-based models.

Figure 1. Wearable sensors used in the experiment: (a) XSens MVN suit [16]; (b) e-glove from Emphasis Telematics.

The paper is organized as follows: Section II presents state-of-the-art activity recognition methods based either on external sensors or on wearable sensors. Section III describes the proposed HMM-based recognition method and presents the experimental test-bed. The results of the comparison with *vs.* without contact information are presented in Section IV and discussed in Section V.

II. RELATED WORK

A. Motion capture based activity recognition

Human activity recognition methods can exploit external or exteroceptive sensors and wearable sensors [6][8][9].

Most external sensors approaches use vision-based systems, such as RGB-D cameras or optical motion capture systems. RGB-D cameras, such as Microsoft Kinect, require

 ISBN: 978-1-61208-616-3

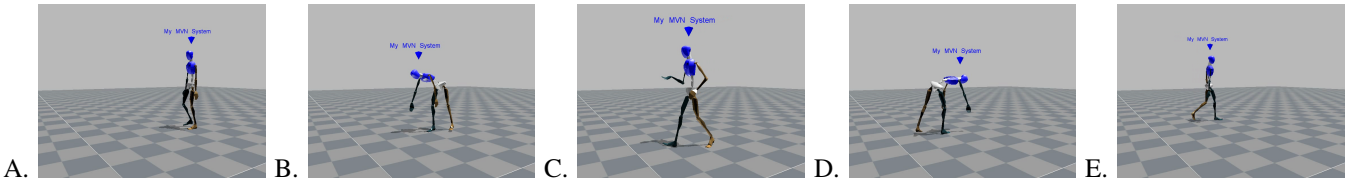

Figure 2. Examples of activities extracted from the MVN Studio software. A: WAIT, B: REACH, C: CARRY, D: PLACE, E: WALK.

image processing to extract motion features. Conversely, in optical motion capture, the 3D trajectories of markers placed on the user are directly retrieved. A major problem of vision-based systems is that the user must always be in the camera's field of view, which limits the applicability in complex spaces. Occlusion of markers is another issue for robust motion detection, especially in cluttered environments.

To avoid these issues that are frequent in industrial conditions, wearable sensors can be used. Such sensors are then directly attached to the user, and no external sensor is needed. Inertial sensors placed on the limbs and torso of the user are the most commonly used [10] [11].

B. Algorithms for activity recognition

Classification algorithms have been widely used to recognize human daily activities, such as walking, sitting or lying [12][10]. Using three inertial sensors placed on the chest, right thigh and left ankle, Attal et al. [10] compared the k-Nearest Neighbor (k-NN), Support Vector Machines (SVM), and HMM algorithms for both supervised and unsupervised learning. They showed that with supervised learning k-NN gives the best performance, whereas with unsupervised learning HMM performs best. They showed that main advantage of using HMM was that the model took the temporal aspect into account. Dubois and Charpillet [12] showed that HMM can efficiently discriminate falls from other daily life activities. Mandery et al. [13] used HMM to identify the best performing sets of features for dimensionality reduction of motion capture data. They showed that a small subset of features was sufficient to perform accurate recognition. Interestingly, the velocity of the whole-body center of mass was always included in the relevant features.

All the aforementioned studies only used motion capture data. Conversely, Wächter and Asfour [14] used optical devices to track not only humans but also objects motions. The distance between the human and tracked objects was used to detect contact and pre-segment the data before the recognition step. Coupeté et al. [15] addressed the problem of real-time activity recognition in an industrial environment using HMM and object-related information. They used depth-cameras to capture human motion and inertial sensors to track tools manipulated by the worker. They showed that the tool-related information improved the classification performance from 80% to 94%.

III. METHOD

A. Experimental protocol

1) Material: We used two wearable systems: the MVN Link suit from Xsens [16] (Figure 1a) and the e-glove from Emphasis Telematics [17] (Figure 1b). The XSens MVN suit was used to capture the whole-body human motion with 17 wireless inertial sensors embedded in a lycra suit. The suit information was combined with a glove containing force sensors on the fingertips of thumb, index and middle fingers and on the palm. However, in this paper, we used only the palm force information to detect contacts.

The sample rates of the MVN suit and the glove are 240 Hz and 100 Hz, respectively. To synchronize the data, both systems used the same wireless network during data collection, and each recorded sample was associated to an absolute time-stamps value.

2) Task description: To evaluate our method, we designed a pick-and-place task of a 6 kg bar, inspired by packaging tasks on assembly lines in manufacturing industry.

One male participant performed 8 sequences of the task, with each sequence consisting of 6 to 8 pick-and-place. Each sequence started and ended in the same neutral pose. The bar was initially placed at a height of 45 cm on a 100×50 cm flat support. The participant was instructed to take the bar with both hands, carry it to the other side of the support, place the bar there and return to the initial position to perform the next iteration. Each sequence lasted around one minute. In order to add variability in the data, the participant was instructed to change the position of his hands on the bar, and to follow two different paths when going to and coming from the bar final position.

We defined seven states/activities (Figure 2):

- WAIT: standing still
- REACH: bending forward, without the bar
- PICK: standing up straight while holding the bar
- CARRY: walking while carrying the bar in the hands
- PLACE: bending forwards with the bar in the hands
- RELEASE: standing up straight with empty hands
- WALK: walking without holding the bar

B. Activity recognition algorithm

1) Hidden Markov Model: We use HMM-based supervised learning to recognize the activities with the `hmmlearn` [18] library in Python. The model is defined by N states representing the activities, such as WALK or PLACE (all activities are presented in Section III-A2), with $S = \{s_1, s_2, ..., s_N\}$ being the set of possible states. Each recorded sequence $k \in [1, K = 8]$ is represented by a series of discrete states $Q^k = \{q_0^k, q_1^k, ..., q_t^k, ...q_T^k\}$ and a series of T observations $X^k = \{x_1^k, ..., x_t^k, ..., x_T^k\}$ corresponding to motion capture and glove data. For each instant, the goal is to infer the current hidden state, such as $q_t^k = s_i$.

Three parameters $\{\Pi, A, B\}$ represent the model. $\Pi = \{\pi_1, \pi_2, ..., \pi_N\}$ denotes the initial state probabilities and is learned from the training set. For each state, π_i is equal

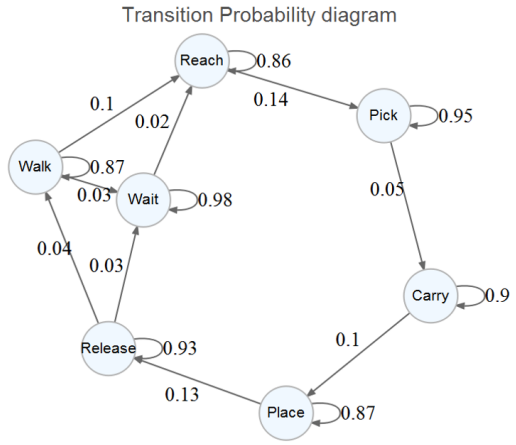

Figure 3. Probability transition diagram. The graph is not fully connected because some transitions are impossible. For instance, to PLACE an object, it first needs to be PICKED.

to the frequency of appearance of this state at the start of a sequence as in (1). $A = \{a_{ij}\}$ is the transition matrix probability, where a_{ij} is equal to the frequency of s_i following s_j in the training set (2) The transition matrix can also be represented by a probability transition diagram (Figure 3). $B = \{b_1, ..., b_N\}$ represents the emission distribution that models the observation (3). For each state, the emission b_i is computed as a multivariate Gaussian $\mathcal{N}(X, \mu_i, \Sigma_i)$ with μ_i and Σ_i the mean vector and the covariance matrix of the Gaussian variable, respectively. μ_i and Σ_i are learned from the observations X related to the state s_i in the training set.

$$\pi_i = p(q_0 = s_i) = \frac{\sum_{k=1}^{K-1} q_0^k = s_i}{K}, i \in [1, N] \quad (1)$$

$$a_{ij} = \frac{\sum_{k=1}^{K-1} \sum_{t=1}^{T} (q_t^k = s_j).(q_{t-1}^k = s_i)}{\sum_{k=1}^{K-1} \sum_{t=1}^{T} (q_{t-1}^k = s_i)}, i, j \in [1, N] \quad (2)$$

$$b_i = p(X|q_t = s_i) = \mathcal{N}(\mu_i, \sigma_i^2), i \in [1, N] \quad (3)$$

In this paper, we used $N = 7$, while T is different for each recorded sequence, from 444 samples for the shortest sequence to 715 samples for the longest one.

2) Normalization: As the data (observations) have different units (*e.g.*, Cartesian position, velocity, contact information), they need to be normalized. $\widetilde{x_t}$ is the vector containing the data x_t normalized within the range [-1,1]:

$$\widetilde{x_t} = 2.\frac{x_t - x_{min}}{x_{max} - x_{min}} - 1 \quad (4)$$

where x_{min} and x_{max} are constants computed from the data in the training set.

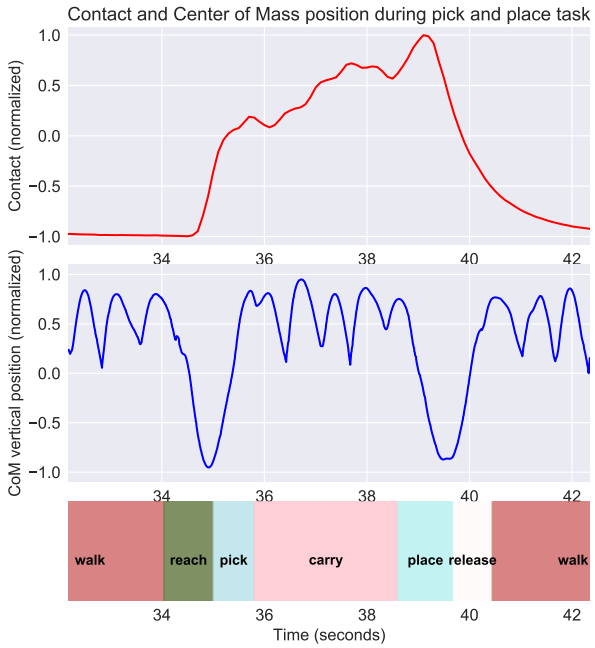

Figure 4. Example of normalized data used to train the model. Top: Time-series of the contact information from the e-glove; Middle: Time-series of the center of mass vertical position; Bottom: Manually labeled activities.

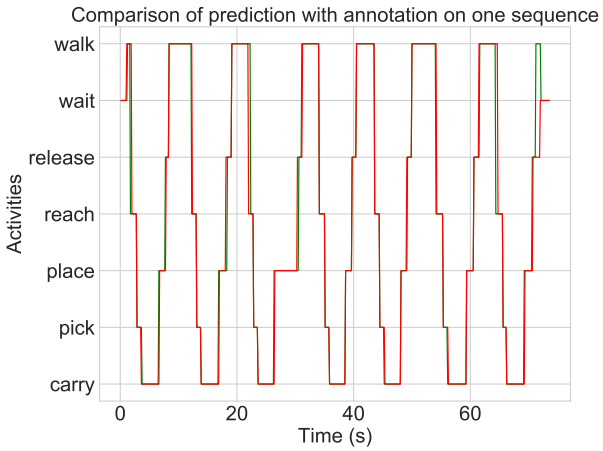

Figure 5. Comparison of manually annotated activities (red) and predicted activities (green).

3) Sliding window: In order to decrease the computational complexity and to reduce noise in the data, a sliding window filter is applied to the recorded motion capture and glove data. For each time window, the observation vector contains the average values of the data across this window. A 60 samples window is used, with an overlap of 30 samples between each window. As the frequency of the MVN Link system is 240 Hz, a window is 250 ms long. Given the 30 frames overlap, there is a new observation every 125 ms. This rate is sufficient since each activity lasts more than one second.

 ISBN: 978-1-61208-616-3

Figure 6. Recall and Precision scores for each state without (top) and with (bottom) contact information.

4) Data annotation: Given that we used supervised learning, the recorded sequences were manually segmented and annotated with the start and end of the different activities by using the Anvil annotation software [19]. .

C. Features selection

Based on the results of [13] and our pilot tests, we include the following motion-related features in the observation X: the vertical position of the center of mass, the 3D velocity of the center of mass, and the 3D velocity of both hands. The contact information is the normal force on the palm. Therefore, the observation vector x_t at each instant t is $x_t \in R^{10}$ without the contact information, and $x_t \in R^{11}$ with it.

D. Data analysis

1) Evaluation: The model is evaluated with cross-validation on the dataset consisting of $K = 8$ recorded sequences. At each iteration, the database is split between the training set that contains $K - 1$ sequences and the test set that contains 1 sequence. The training set is used to learn the parameters of the model, while the test set is used for the evaluation.

2) Metrics: In order to evaluate the advantage of adding the contact information, the recognition performances with and without contact information are compared on the same sequences. The model is evaluated with the Recall score, Precision score and F1-score (harmonic mean of *precision* and *recall*) as in (5 - 7):

$$\text{Recall}_i = \frac{\text{Number of samples correctly classified as class i}}{\text{Number of samples that belong to the class i}} \quad (5)$$

$$\text{Precision}_i = \frac{\text{Number of samples correctly classified as class i}}{\text{Number of samples classified as class i}} \quad (6)$$

$$\text{F1-score}_i = 2.\frac{\text{Precision}_i.\text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (7)$$

For each metric, an average score is computed on all the iterations of the cross-validation.

IV. RESULTS

A. First insight

Some patterns are observable when the data of the vertical position of the center of mass and contact information are compared to the annotated activities (Figure 4). When the participant performs the activities WALK, REACH and RELEASE,

TABLE I. RECOGNITION PERFORMANCE (%) WITH AND WITHOUT CONTACT INFORMATION (LEFT AND MIDDLE COLUMNS), AND PERFORMANCE COMPARISON WITH WILCOXON TEST (RIGHT COLUMN).

	No contact	Contact	*p-value*
Recall	90.07	93.38	0.06
Precision	94.08	96.12	0.16
F1-score	91.99	94.71	0.05

the contact information data are equal to the minimum possible value. Whereas during PICK, CARRY and PLACE activities, the contact data have positive values. We can also identify when the user bends forwards and stands up straight by looking at the center of mass position. During the WALK and CARRY states, there are oscillations around a value of 0.5.

B. Margin of error

Most activities are correctly identified (Figure 5), but the transition between two states does not always happen at the exact same frame on the real and the predicted case. This kind of error is not relevant, as our future application does not require a 125 ms accuracy. Therefore, the performance scores are computed with a margin of error of one 125 ms time window before and after each observation. The classification is considered correct if the predicted state at time t corresponds to the annotated state at either time $t-1$, t, or $t+1$.

C. Overall results

The performance for each score with and without glove is presented in Table I. For each score, the contact information improves the results. The results are compared with a Wilcoxon signed-rank test and we found a significant difference for the F1-score (*p-value* $= 0.05$).

D. Performance for each activity

Figure 6 presents the precision and recall scores for each states. There is mainly a confusion between the RELEASE and WALK activities. This is mainly due to the fact that at the end of each sequence, the subject returns to a neutral position with a little step after placing the last bar. In the annotation, it was labeled RELEASE then WAIT without WALK transition, but the step was classified by the model as *walking* (this error can also be seen in Figure 5).

Figure 7 presents a comparison of the scores for each state for both conditions. We computed a Wilcoxon signed-rank test on the measure performance of *recall*, *precision* and *F1-score*. For most of the measures, there is no significant difference with or without the use of the contact information. We found a significant difference with the F1-score of PICK task, the precision score of PLACE activity, the recall score of REACH task and the precision and F1-score of WALK activity. However, when the contact information is added, there is an improvement of the performances for the recognition of each activity and a reduction of the uncertainty.

Overall, our results show that using both kinetics information and contact information is beneficial for activity recognition with our HMM-based model.

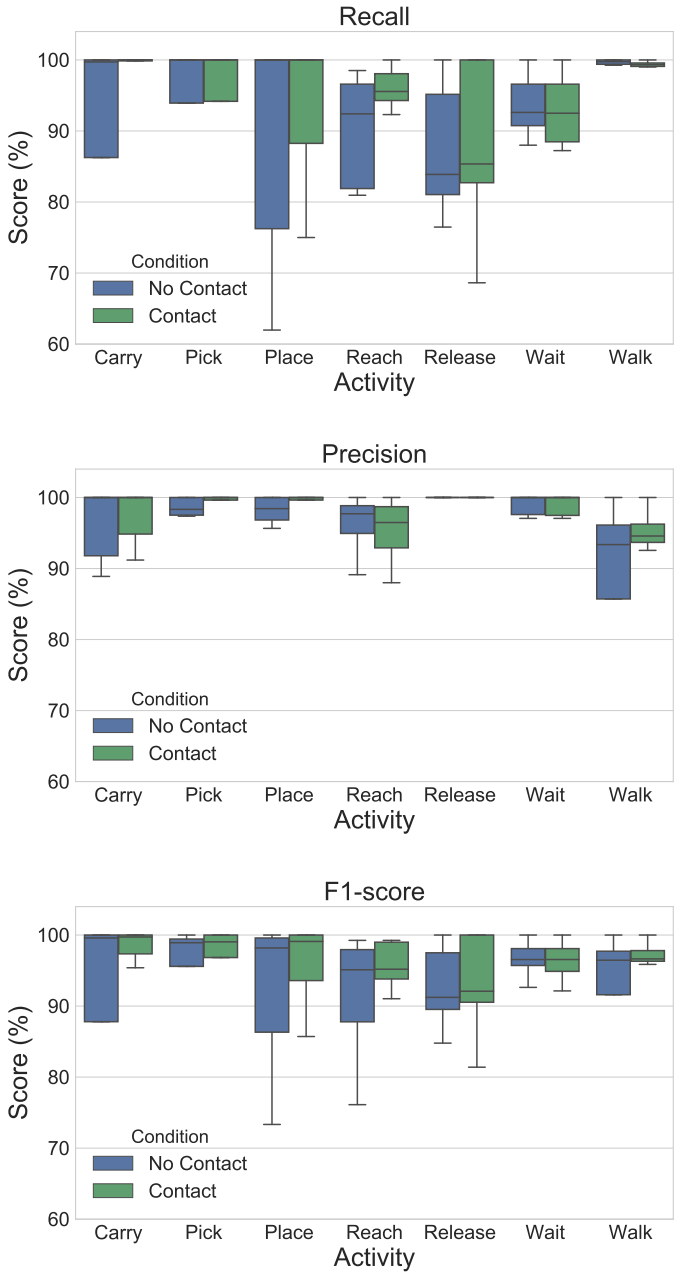

Figure 7. Classification performance with and without contact information for each activity: Recall score (top), Precision score (middle), and F1-score (bottom).

V. CONCLUSION

This paper presents a contextual use of Hidden Markov Model for activity recognition, applied to a pick-and-place task. We evaluate the benefit of using the hand-object contact information in addition to kinetics information to improve the classification performance. The overall performance is better by about 3 % when the contact information is added. The contact information could also increase the robustness of the recognition in case of data loss. For instance, if data corresponding to the PICK activity is unavailable, WALK and CARRY can easily be confused when the recognition is based

 ISBN: 978-1-61208-616-3

on motion features only. With the contact information, the differentiation becomes trivial.

Nevertheless, the dataset we used consists of a single task. In order to test if the usefulness of the contact information can be generalized, future work includes collecting a dataset with more tasks and more variability in the tasks and related activities. Finally, the automation of the recognition process could be increased by adding an automatic selection of relevant features among all the available ones.

ACKNOWLEDGEMENTS

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 731540 (project AnDy). The authors wish to thank Lars Fritzsche and Emphasis Telematics SA (Dr. Giorgos Papapanagiotakis, Dr. Michalis Miatidis, Panos Stogiannos, Giannis Kantaris, Dimitris Potiriadis) for their support with the e-glove.

REFERENCES

[1] E. Schneider, X. Irastorza, M. Bakhuys Roozeboom, and I. Houtman, "Osh in figures: occupational safety and health in the transport sector-an overview," 2010.

[2] G. Li and P. Buckle, "Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods," *Ergonomics*, vol. 42, no. 5, pp. 674–695, 1999.

[3] K. Schaub, G. Caragnano, B. Britzke, and R. Bruder, "The European Assembly Worksheet," *Theoretical Issues in Ergonomics Science*, vol. 14, no. 6, pp. 616–639, 2013.

[4] T. Bossomaier, A. G. Bruzzone, A. Cimino, F. Longo, and G. Mirabelli, "Scientific approaches for the industrial workstations ergonomic design: A review." in *ECMS*, 2010, pp. 189–199.

[5] S. Ivaldi, L. Fritzsche, J. Babic, F. Stulp, M. Damsgaard, B. Graimann, G. Bellusci, and F. Nori, "Anticipatory models of human movements and dynamics: the roadmap of the andy project," in *Proc. International Conf. on Digital Human Models (DHM)*, 2017.

[6] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, 2013.

[7] H. Junker, O. Amft, P. Lukowicz, and G. Tröster, "Gesture spotting with body-worn inertial sensors to detect user activities," *Pattern Recognition*, vol. 41, no. 6, pp. 2010–2024, Jun. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320307005110

[8] J. K. Aggarwal and L. Xia, "Human activity recognition from 3D data: A review," *Pattern Recognition Letters*, vol. 48, no. Supplement C, pp. 70–80, Oct. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865514001299

[9] L. Lo Presti and M. La Cascia, "3D skeleton-based human action classification: A survey," *Pattern Recognition*, vol. 53, no. Supplement C, pp. 130–147, May 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320315004392

[10] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31 314–31 338, 2015.

[11] A. Y. Yang, S. Iyengar, P. Kuryloski, and R. Jafari, "Distributed segmentation and classification of human actions using a wearable motion sensor network," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*. IEEE, 2008, pp. 1–8.

[12] A. Dubois and F. Charpillet, "Human activities recognition with RGB-Depth camera using HMM," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 4666–4669.

[13] C. Mandery, M. Plappert, J. Borras, and T. Asfour, "Dimensionality reduction for whole-body human motion recognition," in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 355–362.

[14] M. Wächter and T. Asfour, "Hierarchical segmentation of manipulation actions based on object relations and motion characteristics," in *Advanced Robotics (ICAR), 2015 International Conference on*. IEEE, 2015, pp. 549–556. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/7251510/

[15] E. Coupeté, F. Moutarde, S. Manitsaris, and O. Hugues, "Recognition of Technical Gestures for Human-Robot Collaboration in Factories," in *The Ninth International Conference on Advances in Computer-Human Interactions*, 2016.

[16] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors," *Xsens Motion Technologies BV, Tech. Rep*, 2009.

[17] "e-glove | Emphasis Telematics," URL: http://www.emphasisnet.gr/e-glove/ [accessed: 2018-01-30].

[18] "hmmlearn 0.2.1 documentation," URL: http://hmmlearn.readthedocs.io/ [accessed: 2018-01-30].

[19] M. Kipp, L. F. von Hollen, M. C. Hrstka, and F. Zamponi, "Single-person and multi-party 3d visualizations for nonverbal communication analysis." in *LREC*, 2014, pp. 3393–3397.

 ISBN: 978-1-61208-616-3