Vol.2019-HCI-185 No.3
Vol.2019-UBI-64 No.3
2019/12/10

A Plan for a Cooking Activity Recognition Challenge with Micro and Macro Activities

Paula Lago[1,a)] Sayeda Shamma Alia[1,b)] Shingo Takeda [1,c)] Kohei Adachi [1,d)]
Moe Matsuki[1,e)] Sozo Inoue[1,2,f)]

Abstract: For seniors, the ability to cook is a strong indicator of cognitive health and the ability to live independently. Cooking, one of the instrumental activities of daily living, involves several steps and a plan to follow such steps. Understanding and recognizing these steps enables a wide range of assisted living applications. For example, if an elderly person keeps missing some steps in cooking, we can find some patterns and suggest them a memory screening or other tests. Also a voice assistant can remind the person of the next step if they forgot. In this challenge, participants will design methods to recognize cooking steps (micro) and cooking activities (macro) from inertial sensors, motion capture sensors and other environmental sensors. The data will be collected during cooking sessions of different foods. It involves several steps performed with a different sequence. In this work, we present the data collection experiment and the lessons learned from previous experiences.

Keywords: activity recognition, nurse care, machine learning, data collection

1. Introduction

Although sensor-based activity recognition has many potential applications in health-care such as automating nursing records [7], remote health care monitoring [8] and work scheduling [9], the competence of current approaches for activity recognition in complex scenarios is limited. One main issue for research advancement is that we cannot compare different approaches to each other as their evaluation is done in different datasets. Therefore, we cannot conclude which approaches are have better performance with confidence.

For scenarios of complex activity recognition, the lack of shared datasets has hindered these comparisons. In fields like image recognition [4], there are open datasets to evaluate the proposal methods equally. These datasets are important to advance these technologies. On the other hand, there are few open datasets for activity recognition. The existing open datasets focus on simple and repetitive activities, such as ”walk”, ”run” and ”stand” [5, 16, 17]. However, complex activities are composed of sequences of actions that are not periodic, nor repetitive, and the evaluation on such datasets gives no hint of their performance on complex activities.

Therefore, we aim to contribute to the activity recognition field by publishing the dataset of complex activities. We are committed to collecting high quality activity recognition datasets that can be used by the community to advance the state of the art and compare different approaches under the same settings.

In this paper, we summarize the results of the previous challenge with nursing activities and present the plan for collecting a cooking activity dataset. Our datasets use motion capture and inertial sensors for measuring human motion during the activities with different precision and multiple body parts information. The cooking dataset will enable research in cooking activity recognition, micro and macro activities recognition and zero-shot learning.

2. Summary of previous challenge results and lessons learned

On June, we presented the Nurse Care Activity Recognition Challenge and it was conducted from July until August. The results were presented during the Hasca Workshop in Ubicomp [12]. In this challenge, 4 teams participated. Table 1 summarizes the approaches and results used by participants and the proposed baseline.

Even if few teams participated, we draw some conclusions from these results. The first lesson to draw is that simple approaches can outperform more complex networks. In this challenge a simple approach based on Nearest Neighbors (KNN) was the winning approach, even when other teams used deep learning and other state of the art approaches. There are several factors for this result. First, the dataset has a high number of dimensions but a not so large number of instances. Deep neural networks require large amounts of data to learn patterns and they are known for not being robust to variations. Since the testing approach was using a new subject, variations are expected. However, KNN uses sim-

1 Kyushu Institute of Technology, Kitakyushu, Fukuoka, Japan
2 Riken AIP
a) paula@mns.kyutech.ac.jp
b) alia@sozolab.jp
c) takeda@sozolab.jp
d) adachi@sozolab.jp
e) matsuki@sozolab.jp
f) sozo@brain.kyutech.ac.jp

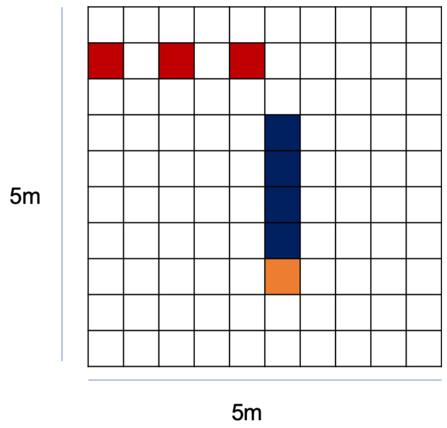

Fig. 1: Plan of the studio in Kyutech and position of furniture. Sink (orange), table (blue), drawer, refrigerator and shelf (red)

ilarity metrics, which shows that the variations are not coming from different activities becoming similar. The second lesson is that the winning team used strong features, based on angles for body posture and location on the room. These features are robust to variations in person size and the location is a strong indicator of the activity.

We also draw some lessons about the data collection experiment. The first is about the quality of data. We found out that the accelerometer data had a slower sampling rate than expected (expected was 20Hz, real was 4Hz) and that there was some missing data. One reason for this is the We also found some body marker labeling errors in the motion capture data. Regardless of this, motion capture was used with success. This is mainly because the errors were in the lower part of the body, whereas the most important information for the activities is given by the upper part of the body.

The second lesson is about ensuring good practices for participants. Two participants had overfitting and their results in the test datasets were low. We designed the evaluation of the challenge in a new user, and if the evaluation in the training dataset is not careful, such errors can take place. We will now advise participants to use a leave-one-subject out cross-validation for their evaluations in the training dataset.

3. Why Macro and Micro Activities?

In this new experiment we will collect data about cooking activities. Although it may seem unrelated to the healthcare field we are focusing on, the reason is that we want to explore the relationship between micro and macro activities. Researchers have defined activities as being composed of actions [3,13,14]. Recognizing both micro (actions) and macro (activities) levels is important for improving accuracy. Specifically, macro activities may vary in the order of the micro activities and it is important to be able to recognize this micro activities to ensure the correct order is followed. Also, if one micro activity is missing, the macro activity might still be recognized, but the quality is lowered. For example, if hand washing at the end of the activity is missing. Therefore, detecting micro activities is important for quality assessment.

Cooking provides a good scenario for this combination because it involves several steps that can be combined in different order to create different recipes. Recognizing the micro activities in cookking is challenging because they involve action and object. For instance, 'cut vegetable' is not the same as 'cut fruit' or 'cut meat'. However, the semantics of the action, 'cut', is the same, so it is important to recognize as a single micro activity.

Macro activities can follow different order but have the same set of primitives. Therefore, the correct identification of the macro activities, depends on the correct identification of the order of the micro activities.

4. Experiment Description

In this section, we describe the activities and micro-activities selected for this dataset. We describe the experiment settings, sensors used and risk mitigation plan. The purpose of the dataset is to study the role of micro activities in the detection of macro activities. Current datasets label only granular activities or only coarse activities like routines, but few have both levels of detail (only OPPortunity dataset to the best of our knowledge [17]). This dataset will be collected by Sozolab (Japan) and Larsen team (France) in two different locations. In this paper, we describe the settings of the part that will be collected in Japan.

4.1 Activities

The dataset contains data about 10 micro activities and 4 macro activities. The macro activities correspond to recipes. For the success of the experiment, we have the following restrictions:

- Recipes should not involve heating as it can not be done in the laboratory for safety.
- Each micro activity should be present in at least 2 macro activities.
- The duration of the recipe should be short and easy so volunteers can easily do even if they have not done it before.

Considering the previous restrictions, we chose the following recipes:

- ***Sandwich*** A cheese and vegetables sandwich. Although it includes a 'spread' micro activity for spreading mayonnaise, we don't consider it as it is not a part of any other activity.
- ***Salad*** A vegetable salad including 3 different vegetables that must be cut and then mixed with dressing.
- ***Cereal***: Pouring milk and cereal into a bowl. We include banana into the cereal to include cut and peel micro activities.
- ***Fruit salad***: A fruit salad including 3 different fruits that must be cut and then mixed with yogurt.

We designed the micro-activities in accordance to previous datasets labels (Section 6). In those datasets, the action is accompanied by the object, if it is relevant. Figure 2 shows a summary of the micro activities involved in each recipe.

4.2 Experimental settings

The dataset will be recorded in the Smart Life Care Unit of the Kyushu Institute of Technology in Japan (Figure 1). This unit is located in the Wakamatsu Campus of the University. A second part of the dataset will be recorded in the smart apartment of LORIA, Nancy in France.

Data collection in Kyutech will take three days total. Each day,

Table 1: Summary of submissions made to challenge ordered by accuracy in the test dataset

Team	Sensors used	Method	Training Accuracy	Test Accuracy
Team IITDU [10]	Motion Capture and Meditag data	KNN	66.11%	80.2%
TDU-DSML [1]	Motion Capture	Spatio-temporal Graph Convolutional Network	57%	64.6%
Baseline DL	All modalities	CNN	100%	46.5%
Baseline ML	Acceleration	Random Forest	60%	43.1%
Dark Shadow [6]	Motion Capture and Meditag	Gated Recurrent Unit	66%	29.3%
Data Digger [15]	Acceleration	Random Forest	82.86%	18.1%

	Take			Wash	Cut		Peel	Put	Pour	Mix
	shelf	drawer	refrigerator		Vegetable	Fruit				
Sandwich	X	X	X	X	X			X		
Cereal	X	X	X			X	X	X	X	
Salad	X	X	X	X	X			X	X	X
Fruit salad	X	X	X	X		X	X	X	X	X

Fig. 2: Micro activities involved in each macro activity

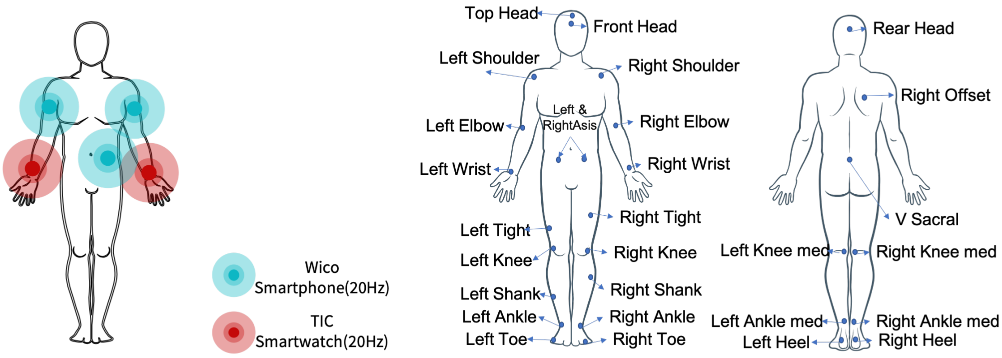

Fig. 3: Placement of smartphones, smartwatches and motion capture markers in the experiment

Fig. 4: Table, sink, shelf, and drawers used to create the kitchen environment for the experiment

one or two voluntary student participants come to the laboratory and prepare the recipes, 6 times each. All participants have already agreed to the sensor data collection for this experiment and will not be compensated.

The participants are male students volunteers. Most of them do not cook by themselves, so the simple recipes were chosen considering this.

We performed an initial test for the data collection experiment in the smart apartment of LORIA, in Nancy, France (Figure 5) and in Kyutech (Figure 6). During this test, we decided the activities should follow the same order and we should have specific placements for the objects during the recipe, to reduce marker occlusion for the motion capture and from the open pose sensor. Therefore, we decided to have 4 constrained runs and 1 free run of each recipe.

Fig. 5: Test experiment in LORIA

For each activity, first the instructions about the procedure will be given, so that the activity comprises all necessary steps. One person will guide the volunteer so that the actions follow the same order in the first four runs and the last run is free for the volunteer to decide the order of the actions. The experiment room is equipped with the kitchen furniture illustrated in Figure 4. The shelf, refrigerator and drawers were selected so that the open and close movements are different for each.

4.3 Sensors used

We will collect data from motion capture, open pose and accelerometer sensors. Video will also be recorded for each run, but it will not be released due to privacy concerns of the participants. The description of each data source is given below.

Motion Capture: We use the motion capture system form Motion Analysis Company[*1]. The setup used consists of 29 body markers located as in Figure 3. The markers are tracked using 16 infrared cameras (Kestrel Digital Real Time as in Figure **??**). The three dimensional position of each marker is recorded with a frequency of 100Hz. The markers may be labeled incorrectly in some cases due to the complex setting.

Open Pose: [2] Open Pose is an open source system to detect 135 keypoints of the human body, hand, facial, and foot using a single COTS camera. We use the keypoints from body (25). The main purpose is to compare results of motion capture with those of open pose, which is less costly than motion capture.

Accelerometer sensor: We will use three smartphones and two smartwatches. The smartphones are Wico W V600 model [*2], which has Android v8 installed and 3-axis accelerometer. The smartwatch is TIC Watch E, which uses Google wear and connects to the smartphone via bluetooth. For the measurements of the smartwatch, we collect two timestamps, that of the smartwatch and that of the smartphone when it receives the data.

4.4 Risk mitigation plan

Based on our previous experience we have the following plan for improving data quality and posterior analysis.

Sensor synchronization was made by clock synchronization las time. We plan to use the same method since it gave good performance. However, we also plan to use initial and end pose calibrations so that the start of the action can be better synchronized.

Orientation calibration In many datasets, the orientation of the sensor is not known. This makes the analysis difficult since the measurements depend on the orientation. We will have an initial pose holding for 0.5 seconds so that the initial data can be used for orientation calibration of all accelerometers.

Missing data Since Android uses a best effort approach for sampling, missing data can be high when using it. Although last time we used a setting expecting a 20Hz sampling rate, we will use the smallest delay available in the API this time to reduce this risk. We have also updated the version of Android.

Motion capture labeling errors: We will follow the same approach, correcting all upper body errors while leaving lower body labeling errors. We have extended the recording area to ensure the quality of the labeling.

5. Challenge description

The challenge will be open until April 30, 2020. Participants should follow the instructions in the webpage https://abc-research.github.io/cook2020/. The results will be evaluated based on accuracy metric. Participants should submit a paper to the workshop, describing the data sources selected, their preprocessing pipeline, the features used and their results. All submissions must follow the workshop format.

We want to increase participation so we will announce in different events, mailing lists and social media.

6. Related datasets

In this section, we describe other datasets that are publicly available and comprise cooking scenarios and micro and macro activities.

CMU Lifestyle dataset: This dataset was recorded using dif-

[*1] http://motionanalysis.com/movement-analysis/

[*2] https://www.handsetdetection.com/device-detection-database/devices/wiko/w-v600/

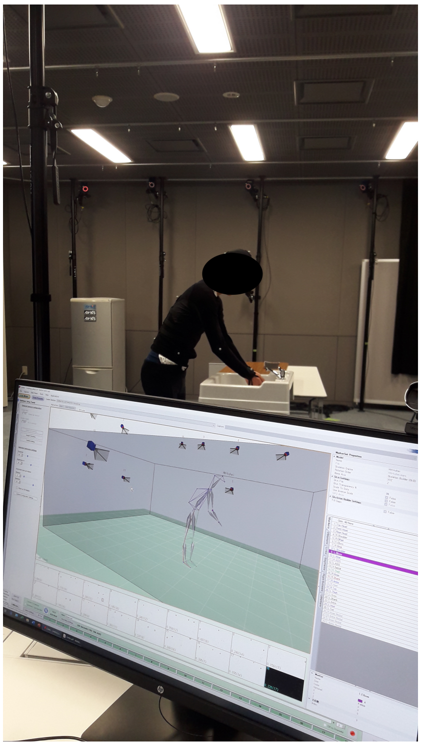

Fig. 6: Test experiment in Kyutech

ferent modalities including: video, audio, wearable watch, motion capture system, accelerometers and gyroscopes. It comprises of the activities conducted during cooking and food preparation. There were five subjects and five different recipes: brownies,pizza, sandwich, salad and scrambled eggs.

Cooking activities dataset [11]: The data was collected using motion capture system based on wearable inertial measurement units (five positions). The scope was the activities done during meal time, many micro activities and five macro activities: cooking meal, setting table, eating meal, cleaning up and putting away utensils. The subjects of this experiment followed an experimenter who described the tasks, so there was a slight dependency although subjects were free to chose their order of actions. Only one recipe is done.

Opportunity dataset [17]: contains information about gestures that occur during some high level activities. Similar to micro and macro activities. The data was collected in environment rich with sensors. There were 72 sensors in the environment and on the body of the subjects. The dataset includes complex cooking activity mainly associated with breakfast (preparing a sandwich). Four subjects performed 17 micro activities in a predefined scenario.

7. Recommendations for future datasets

To promote the shareability and ability to use together different datasets, we following recommend this advice:

- *Follow similar labeling patterns:* For instance, in this experiment we follow labeling similar to the Cooking dataset [11] where they label both the action and the object.
- *Use similar sensor position:* Some public datasets use multiple sensors but their positions are not compatible, hindering their use together.
- *Label orientation:* As we mentioned before, the orientation of the accelerometer is important so that preprocessing techniques can be applied more confidently. We suggest to label the orientation or have a calibration pose at the beginning of the experiment

Since this experiment is being held in France and Japan, we will follow these recommendations so that the two datasets can be combined into a single dataset with multiple recipes and different micro-activities.

8. Future applications

The dataset we will collect focuses on cooking activities. For seniors, the ability to cook is a strong indicator of cognitive health and the ability to live independently. Cooking, one of the instrumental activities of daily living, involves several steps and a plan to follow such steps. Understanding and recognizing these steps enables a wide range of assisted living applications. For example, if an elderly person keeps missing some steps in cooking, we can find some patterns and suggest them a memory screening or other tests. Also a voice assistant can remind the person of the next step if they forgot. The correct classification of the steps is crucial for the success of these applications.

In addition, research in recognizing micro and macro activities is important in other domains as well. In nursing activities the order of actions is a strong indicator of quality and compliance with safety and other rules. For instance, hand washing must be done at specific moments, which are different for each activity. In such cases, it is not only important to recognize the step of hand washing but also the macro activity during which it was done and the order within that activity. Therefore, to investigate the possibility of recognizing micro activities, their correct order and the macro activity to which they belong is important for these applications.

A third challenge we want to investigate by using this dataset is zero-shot learning. In zero-shot learning activities are usually described by their attributes. These attributes can be thought of as micro-activities. However, the performance of zero-shot learning depends on the design of these micro-activities and their successful recognition. Therefore, datasets labeled with micro and macro activities are important for this application.

9. Conclusion

In this paper, we have presented an experiment to collect a dataset for micro and macro activities involved in cooking scenarios. Our main motivation is to bridge the gap on activity recognition research for complex activities.

The goal of our experiment is to evaluate the feasibility of recognizing micro activities based on body movement and location with respect to a kitchen room. The second goal is to study the effect of recognizing such micro activities to recognize macro activities.

We will open the data through the Cooking Activity Recognition Challenge. The challenge participants can use the data freely and are encouraged to submit papers to the workshop we will organize.

The application of activity recognition technology in the healthcare domain opens innovative services for improving quality of care, work satisfaction, and others.

References

[1] Xin Cao, Wataru Kudo, Chihiro Ito, Masaki Shuzo, and Eisaku Maeda. Activity recognition with st-gcn for 3d motion data. In *HASCA '19*, Ubicomp '19, 2019.

[2] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*, 2018.

[3] Alexandros André Chaaraoui, Pau Climent-Pérez, and Francisco Flórez-Revuelta. A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living. *Expert Systems with Applications*, 39(12):10873–10888, 2012.

[4] Jia Deng, R. Socher, Li Fei-Fei, Wei Dong, Kai Li, and Li-Jia Li. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, volume 00, pages 248–255, 06 2009.

[5] H. Gjoreski, M. Ciliberto, L. Wang, F. J. Ordonez Morales, S. Mekki, S. Valentin, and D. Roggen. The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access*, 6:42592–42604, 2018.

[6] Nazmul Haque, Mahir Mahbub, Hasan Tarek, Lutfun Nahar Lota, and Amin Assan Ali. Nurse care activity recognition: A gru-based approach with attention mechanism. In *HASCA '19*, Ubicomp '19, 2019.

[7] Sozo Inoue, Paula Lago, Tahera Hossein, Tittaya Mairittha, and Nattaya Mairittha. Integrating activity recognition and nursing care records: the system, deployment, and a verification study. *IMWUT*, 2019. In Press.

[8] Sozo Inoue and Xincheng Pan. Supervised and Unsupervised Transfer Learning for Activity Recognition from Simple In-home Sensors. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services - MOBIQUITOUS 2016*, volume 28-Novembe, pages 20–27, New York, New York, USA, 2016. ACM Press.

[9] Sozo Inoue, Naonori Ueda, Yasunobu Nohara, and Naoki Nakashima. Mobile activity recognition for a whole day: Recognizing real nursing activities with big dataset. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 1269–1280, New York, NY, USA, 2015. ACM.

[10] Eusha Kadir, Pritom Saha Akash, Sadia Sharmin, Amin Ahsan Ali, and Mohammad Shoyaib. Can a simple approach identify complex nurse care activity? In *HASCA '19*, Ubicomp '19, 2019.

[11] Frank Krüger, Martin Nyolt, Kristina Yordanova, Albert Hein, and Thomas Kirste. Computational state space models for activity and intention recognition. a feasibility study. *PLOS ONE*, 9(11):1–24, 11 2014.

[12] Paula Lago, Sayeda Shamma Alia, Shingo Takeda, Tittaya Mairittha, Nattaya Mairittha, Farina Faiz, Yusuke Nishimura, Kohei Adachi, Tsuyoshi Okita, François Charpillet, and Sozo Inoue. Nurse care activity recognition challenge: Summary and results. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, pages 746–751, New York, NY, USA, 2019. ACM.

[13] Paula Lago, Claudia Jiménez-Guarín, and Claudia Roncancio. Contextualized behavior patterns for change reasoning in Ambient Assisted Living: A formal model. *Expert Systems*, 34(2):e12163, apr 2017.

[14] Ye Liu, Liqiang Nie, Li Liu, and David S. Rosenblum. From action to activity: Sensor-based activity recognition. *Neurocomputing*, 181:108–115, mar 2016.

[15] Nazmus Sakib Patwary, Protap Kumar Saha, and Ifthakhar Ahmed. Nurse care activity recognition challenge using a supervised methodology. In *HASCA '19*, Ubicomp '19, 2019.

[16] A. Reiss and D. Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th International Symposium on Wearable Computers*, pages 108–109, June 2012.

[17] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, H. Sagha, H. Bayati, M. Creatura, and J. d. R. Millàn. Collecting complex activity datasets in highly rich networked sensor environments. In *2010 Seventh International Conference on Networked Sensing Systems (INSS)*, pages 233–240, June 2010.