Bento Packaging Activity Recognition Based on Statistical Features

Name of First Author and Name of Second Author

Abstract Due to the fast advancements of low-cost micro-embedded sensors and MoCap sensors, human action recognition has become an essential study topic and is garnering a lot of interest in different sectors. Recently, it is drawing a lot of attention in human-robot collaboration to assist human to preform regular tasks step-wise because it is difficult to obtain human labor at a lower wage to monitor industrial works. In this work, we have presented a straightforward machine learning paradigm to recognize 10 different *Bento* (lunch-box) packaging activities in real-time world. Unlike other skeleton-based human activity recognition domain, it is a very challenging task due to the absence of lower-body marker information. Under these circumstances, we have provided an in-depth statistical analysis of different Bento packaging activities. After feature extraction process, we have used several machine algorithms and obtained best results in Random Forest Classifier using hyperparameter tuning. We have achieved 64.9% validation accuracy using leave-one-out method.

1 Introduction

Human Action Recognition (HAR) in real-time applications is both an exciting and difficult problem because it differs from the diverse settings and behaviors of a person. It has become a popular research area with numerous applications in surveillance, healthcare monitoring, robot vision, fall detection, sports analysis, entertainment, autonomous driving, security, and smart homes and so on [1, 2]. The research domain of HAR can be divided into 2 categories: vision-based or sensor-based techniques.

Name of First Author
Name, Address of Institute, e-mail: name@email.address

Name of Second Author
Name, Address of Institute e-mail: name@email.address

The majority of the research focuses on the spatial and temporal aspects of an action recognition video sequence recorded by RGB cameras. Illumination, varied angles of view, color videos reduce the overall performance significantly in vision based HAR. Skeleton based HAR has made remarkable progress because of the ease of environmental setups, faster execution speed, and low cost of Kinect sensors. In the skeleton-based HAR domain, research on aged people's daily activity monitoring [3] and physical rehabilitation exercise assessment of patients [4] has recently gained a lot of traction. It is very useful since it will track their regular activities, provide real-time feedback on their physical health status in daily basis, and supply physicians with current information regularly without costing transportation problem and huge healthcare cost burdens [5, 6].

Researchers have made remarkable progress explored several models for automatic recognition and assessment of human actions in real-time applications. Machine learning (ML) techniques such as Support vector Machine [7], Random Forest [8], and K-NN [9] have been used to tackle the HAR problem for a long time. In addition, Bayesian Networks [10] and Markov-models [11] have been used with promising results. Deep learning models have sparked a lot of interest recently, and they deliver excellent results in the HAR domain with their superior feature extraction abilities. Researchers in [12] provided 3D CNN model to classify human actions in the real-time surveillance videos. Wentao Zhu et al. [13] proposed a deep LSTM model to extract the concurrence features in HAR domain. I. Lee et al. [14] proposed a temporal sliding LSTM model for skeleton based HAR. Besides, R. Chaudhry et al. [20] provided a 3D discriminative skeleton features to recognize human actions.

Very recently, HAR has gained a lot of interest for its use in tracking the actions of workers in many sectors. There have been some research progress of HAR domain in real-time applications for hospitals and nursing homes [17, 18]. Unlike conventional activities such as jogging, walking, running, standing, sitting, and so on [22], recognizing nursing services towards is a difficult and challenging process because traits of an activity rely on both patients and nurses and so, there is a lot of intra-class diversity in nursing activities. Besides, nursing activities, complex cooking activity recognition [19] was arranged in 2020. When it comes to cooking, for example, HAR is quite good for reminding a lonely old person of a missing step or ensuring that a nutritious diet is being maintained. This sort of industrial activity dataset collection is really difficult to come by. In SHL Challenge [20], HAR has been employed to classify 8 different locomotion and transportation activities.

This is a major issue for a variety of sectors that demand a large number of workers at a cheap wage. HAR has a significant influence on human-robot collaboration. For example, When cooking in a industry, the procedures and materials applied in the dish are entirely up to the user. However in packaging task, it is extremely usual to insert an in-gradient. If robots can be utilized to help people with these kind of small tasks, they might be a perfect solution to this issue. In this paper, we provide a simple machine learning framework to recognize Bento packaging activities. Our

machine model is very efficient and can be trained with low computational cost. The paper is organized as follow: In section 2, we have provided a short description of this challenge dataset [18]. In section 3, we represent our proposed workflow to solve this problem. The later section, 4 provides our advance results. The final section 5 provides a summary of our work and some future works.

2 Dataset Description

The dataset [18] for this Bento Packaging Activity Recognition Challenge comprises a high number of *NaN* values and solely contains skeletal Joint 3D coordinates obtained by MoCap sensor. The train dataset contains 10 different activities performed by 3 subjects. Each action contains 5 repetitions performed by each subject. The participants are instructed to do activities in 5 different scenarios such as normal, forgot to put ingredients, failed to put ingredients, turn over Bento-box and fix/rearranging ingredients. Length of each activity segment is about 50 to 70 seconds. The dataset is perfectly balanced which is very rare in other HAR domain. Skeleton joints coordinates were captured at 100 Hz and no preprocessing technique is applied to this data. The dataset contains no information about the lower body joints. The dataset is difficult to work with since it contains just 13 joints from the upper body parts and a high number of empty values. The test data consists of one subject and the format is same as train dataset

Activity No	Scenarios	Patterns
1	Normal	Inward
2		Outward
3	Forgot to put ingredients	Inward
4		Outward
5	Failed to put ingredients	Inward
6		Outward
7	Turn over bento-box	Inward
8		Outward
9	Fix/rearranging ingredients	Inward
10		Outward

Fig. 1: Activity list of Bento packaging activity recognition

3 Methodology

In general skeleton based HAR, 3D coordinates of different markers in body are provided. Human behaviors can be described as time sequences of skeleton positions, according to our basic premise. Our main purpose of this work is to classify 10 different Bento packaging activities in real-time applications. In this section, we will introduce our proposed approach to classify 10 different Bento Packaging Activities. A simple block diagram our proposed approach to solve this challenge is shown in 2 Our method consists of 3 different major steps: 1. Data Preprocessing, 2. Feature Extraction, and 3. Model training.

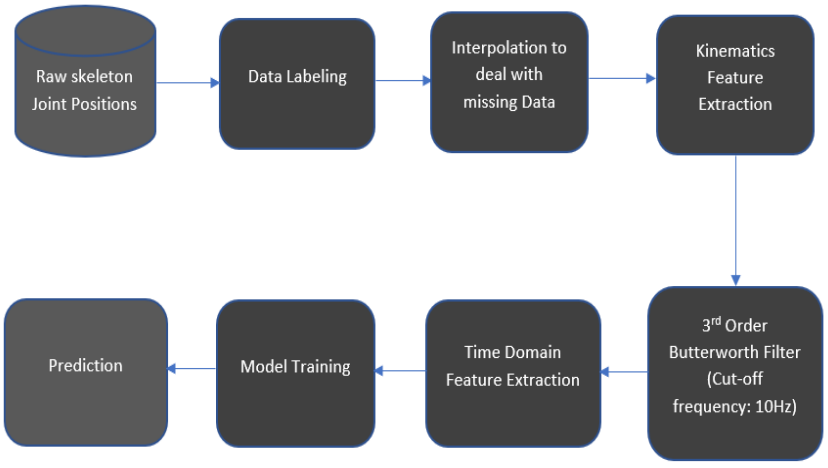

Fig. 2: A simple system flow diagram of the employed method

3.0.1 Data Preprocessing

The skeleton joint positions acquired by MoCap sensor are preprocessed via several steps. First, we collect raw skeleton joints positions and their corresponding action labels from different folders and merge them into a single dataframe. In most cases, the issue of missing data is overlooked while collecting data from MoCap sensors. In this challenge, almost every repetition of an action has a large number of missing values. To solve this problem, we employed the linear interpolation approach to fill the empty values considering real-time circumstances. We did not apply any

segmentation method because the initial and last segments of an action frequently show different characteristics due to both technical errors and human mistakes.

3.0.2 Feature Extraction

Feature Extraction is a dimensional reduction procedure that converts raw input data into a set of features that may be used to distinguish certain actions. In this challenge, the dataset was provided with only skeleton joint positions. This is a very challenging dataset because only 13 upper-body joint coordinates are available. We have extracted various joint-angles, joint-distance and joint-vectors. For angles, we have extracted both elbow and shoulder angles and their corresponding difference. For joint distances, we have extracted elbow-elbow, wrist-wrist and both shoulder-wrist distances. We have extracted different joint vectors in all 3D coordinates. After extracting all the motion features, we have employed a 3rd order Butter-worth low-pass filter to remove random noises and spikes of skeleton data. In total, we used 47 different spatial features including joint-distances, joint-angles and joint-vectors.After that, we used several important time-domain features which are likely relevant to solve our classification problem. In general, we employed minimum, maximum, mean, and standard deviation, median absolute deviation, peak to peak range, mean to peak range, as in earlier HAR research. Besides, we have used some other time-domain features. These features are listed below along with their definitions and mathematical expressions:
25th Percentile: It's also described as the first quarter ($Q1$). The 25th Percentile refers to a data value that is less than or equal to 25 percent of the sorted dataset.

$$Q1 = (\frac{25}{100})n \tag{1}$$

Here, n is the total number of samples in a repetition.
75th Percentile: It's also known as the third quartile ($Q3$). The 75th Percentile is the data value that is less than or equal to 75 percent of the sorted dataset.

$$Q3 = (\frac{75}{100})n \tag{2}$$

Interquartile Range (IQR): It is the calculation of the statistical variation which denotes the difference between 75th and 25th percentiles.

$$IQR = Q3 - Q1 \tag{3}$$

Skewness: Skewness calculates the asymmetry of a real-valued random variable's probability distribution around its mean. The skewness value can be any of the following: positive, zero, negative, or undefined.

$$Skewness = \frac{n}{(n-1)(n-2)} \sum \frac{(Xi - \bar{X})^3}{\sigma^3} \tag{4}$$

After all spatial and time-domain feature extraction, we have found total 615 features. Then we train our dataset with various machine learning algorithms such as Random Forest Classifier (RFC), Extra Tree Classifier (ETC), Support Vector Machine (SVC) and K-Nearest Neighbors (K-NN).

4 Result and Analysis

We have suggested a simple machine learning paradigm to solve the classification problem of Bento packaging activities. The overall performance of our suggested method employing multiple machine learning algorithms will be described in this section. We used hyperparameter tuning to get the best reults in validation data. We tuned max_depth, n_estimators, min_samples_split and n_jobs. Among them, max_depth is the most important parameter and it refers to the depth of each tree in the forest. if it is very high, then the model overfits the test data. Only tuning this parameter increases an average accuracy of 10%. By tuning all the mentioned hyperparameters, we were able to increase the overall performance by 17%. The train dataset consists of total 3 subjects. We have used leave one out method to test the robustness of our proposed method. In this way we get three different accuracy using three different combinations of data. We have obtained highest accuracy in RFC model which is 64.9%. Random Forest is a ensemble machine learning model in which each individual tree in the forest produces a class prediction and with the class with the most votes becoming the model's prediction. . Besides, we have found 61.6% , 58% and 57% in ETC, SVC and K-NN model respectively. We have obtained a promising result on validation dataset.

5 Conclusion

Skeleton based HAR is drawing a lot of attraction in human-robot interaction in various industrial applications. However, Research progress in this domain has not been flourished yet due the lack of publicly available challenging dataset. In this paper, we have provided a simple yet efficient machine learning framework to recognize complex Bento packaging activities. Our proposed method is very practical and can be trained costing less computational resources. Our experimental result illustrates that our proposed workflow can provide a decent accuracy. The primary difficulty we encountered when working on this difficult data was large missing data, which we solved using the linear interpolation approach. In future, we will add some more complex features to improve the robustness of our model. In addition, we will explore several hybrid Deep Neural Networks to improve overall performance for automatic recognizing of packaging activities. We are optimistic that our method will yield superior outcomes on the final test data as well.

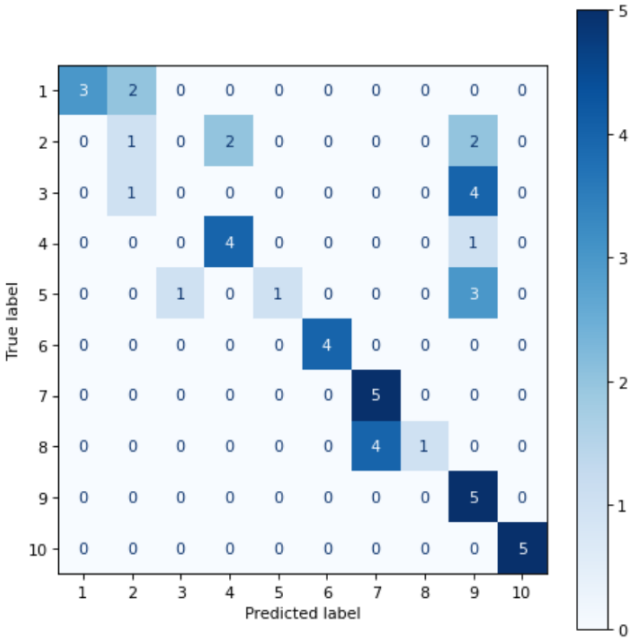

Fig. 3: Confusion matrix while employing leave one out cross-validation for subject 1

References

1. Ahad, Md Atiqur Rahman. (2020). Vision and Sensor-Based Human Activity Recognition: Challenges Ahead. 10.4018/978-1-7998-2584-5.ch002.
2. Md Atiqur Rahman Ahad, Anindya Das Antar, Omar Shahid; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2019, pp. 1-11.
3. Abbate, Stefano Avvenuti, Marco Corsini, Paolo Light, J. Vecchio, Alessio. (2010). Monitoring of Human Movements for Fall Detection and Activities Recognition in Elderly Care Using Wireless Sensor Network: a Survey, Wireless Sensor Networks: Application - Centric Design, Geoff V Merrett and Yen Kheng Tan, IntechOpen, 10.5772/13802.
4. Yalin Liao, Aleksandar Vakanski, Min Xian, David Paul, Russell Baker, A review of computational approaches for evaluation of rehabilitation exercises, Computers in Biology and Medicine, Vol. 119, 2020.
5. Ivy K. Ho, , Kenneth R. Goldschneider, Susmita Kashikar-Zuck, Uma Kotagal, Clare Tessman, and Benjamin Jones. "Healthcare Utilization and Indirect Burden among Families of Pediatric Patients with Chronic Pain". Journal of Musculoskeletal Pain, Vol. 16, No. 3 (2008): 155-164.
6. Steven R. Machlin, Julia Chevan, William W. Yu, Marc W. Zodet, Determinants of Utilization and Expenditures for Episodes of Ambulatory Physical Therapy Among Adults, Physical Therapy, Vol. 91, No. 7, 2011, pp. 1018–1029.
7. Bilal M'hamed Abidine, Lamya Fergani, Belkacem Fergani, and Mourad Oussalah. 2018. The joint use of sequence features combination and modified weighted SVM for improving daily activity recognition. <i>Pattern Anal. Appl. 21, 1 (February 2018), 119–138.
8. C. Hu, Y. Chen, L. Hu, X. Peng, A novel random forests based class incremental learning method for activity recognition, Pattern Recognition 78 (2018) 277–290.

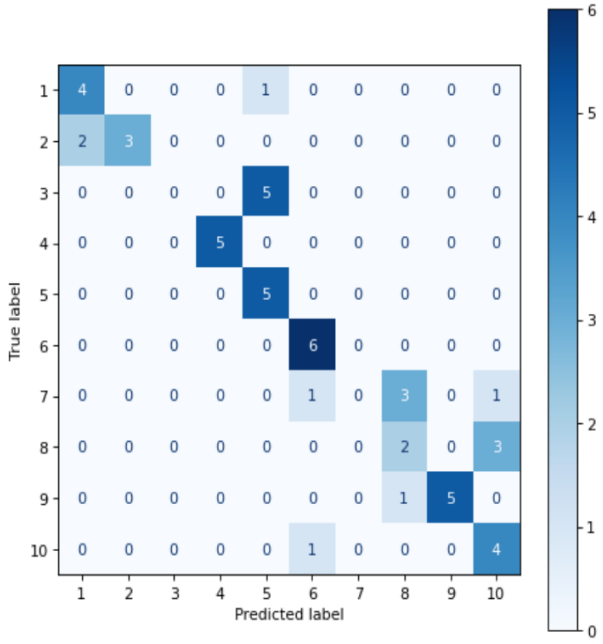

Fig. 4: Confusion matrix while employing leave one out cross-validation for subject 2

9. K. Förster, S. Monteleone, A. Calatroni, D. Roggen and G. Troster, "Incremental kNN Classifier Exploiting Correct-Error Teacher for Activity Recognition," 2010 Ninth International Conference on Machine Learning and Applications, 2010, pp. 445-450.
10. Xiao, Q., Song, R. Action recognition based on hierarchical dynamic Bayesian network. Multimed Tools Appl 77, 6955–6968 (2018).
11. P. Sok, T. Xiao, Y. Azeze, A. Jayaraman and M. V. Albert, "Activity Recognition for Incomplete Spinal Cord Injury Subjects Using Hidden Markov Models," in IEEE Sensors Journal, vol. 18, no. 15, pp. 6369-6374, 1 Aug.1, 2018.
12. S. Ji, W. Xu, M. Yang and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221-231, Jan. 2013.
13. Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., and Xie, X. (2016). Co-Occurrence Feature Learning for Skeleton Based Action Recognition Using Regularized Deep LSTM Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 30(1).
14. I. Lee, D. Kim, S. Kang and S. Lee, "Ensemble Deep Learning for Skeleton-Based Action Recognition Using Temporal Sliding LSTM Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1012-1020.
15. R. Chaudhry, F. Ofli, G. Kurillo, R. Bajcsy and R. Vidal, "Bio-inspired Dynamic 3D Discriminative Skeletal Features for Human Action Recognition," 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 471-478.
16. Sayeda Shamma Alia, Kohei Adachi, Nazmun Nahid, Haru Kaneko, Paula Lago, Sozo Inoue, June 3, 2021, "Bento Packaging Activity Recognition Challenge", IEEE Dataport, doi: https://dx.doi.org/10.21227/cwhs-t440.s
17. Paula Lago, Sayeda Shamma Alia, Shingo Takeda, Tittaya Mairittha, Nattaya Mairittha, Farina Faiz, Yusuke Nishimura, Kohei Adachi, Tsuyoshi Okita, François Charpillet, and Sozo Inoue.

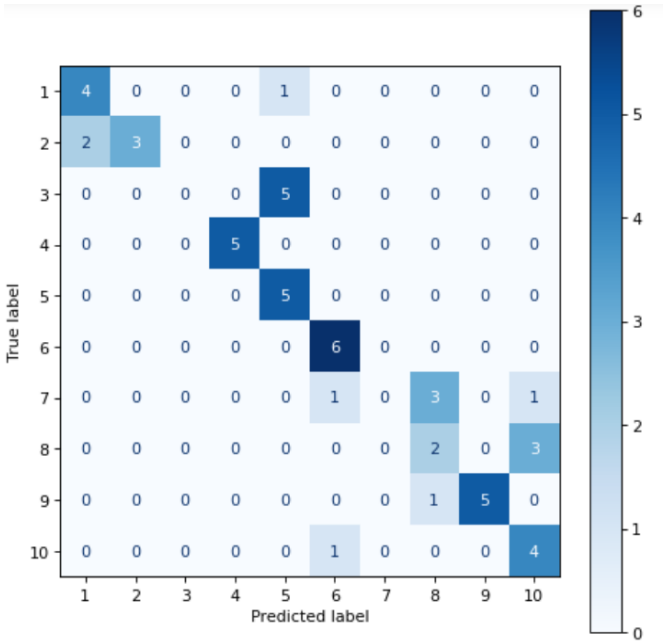

Fig. 5: Confusion matrix while employing leave one out cross-validation for subject 3

2019. Nurse care activity recognition challenge: summary and results. UbiComp/ISWC'19 Adjunct, 746–751.

18. Sayeda Shamma Alia, Paula Lago, Kohei Adachi, Tahera Hossain, Hiroki Goto, Tsuyoshi Okita, and Sozo Inoue. 2020. Summary of the 2nd nurse care activity recognition challenge using lab and field data. In Adjunct Proceedings of UbiComp-ISWC'20, 378–383.
19. Alia S.S. et al. (2021) Summary of the Cooking Activity Recognition Challenge. In: Ahad M.A.R., Lago P., Inoue S. (eds) Human Activity Recognition Challenge. Smart Innovation, Systems and Technologies, vol 199. Springer.
20. Lin Wang, Hristijan Gjoreskia, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2018. Summary of the Sussex-Huawei Locomotion-Transportation Recognition Challenge. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp'18), 1521–1530.
21. Md Atiqur Rahman Ahad, Masud Ahmed, Anindya Das Antar, Yasushi Makihara, Yasushi Yagi, Action recognition using kinematics posture feature on 3D skeleton joint locations, Pattern Recognition Letters, Vol. 145, 2021, pp. 216-224.
22. Pinilla, Macarena, Javier Medina, and Chris Nugent. 2018. "UCAmI Cup. Analyzing the UJA Human Activity Recognition Dataset of Activities of Daily Living" Proceedings 2, no. 19: 1267.