Interactive and Interpretable Online Human Activity Recognition

1st Yale Hartmann
Cognitive Systems Lab
University of Bremen
Bremen, Germany
yale.hartmann@uni-bremen.de

2nd Hui Liu
Cognitive Systems Lab
University of Bremen
Bremen, Germany
hui.liu@uni-bremen.de

3rd Tanja Schultz
Cognitive Systems Lab
University of Bremen
Bremen, Germany
tanja.schultz@uni-bremen.de

***Abstract*—This demo paper puts forward our interactive and inspectable real-time recognition demo for human activity recognition. The demo captures online data from wearable sensors such as inertial measurement units, recognizes the performed human activity using Hidden Markov Models, and displays the search's current state along with the recognition results in real-time. Therefore, allowing users to interact with the system and probe its recognition by altering their actions and comparing their expectations with the demo's performance. Students, Researchers, or the general public can easily extend the detected classes by utilizing the integrated recording, segmentation, and re-training tools — all while running on moderate hardware or even IoT devices like a Raspberry Pi. The demo is titled "ASK2.0" and provides an easy-to-use and powerful way to teach and explain real-time human activity recognition and machine learning.**

***Index Terms*—Human Activity Recognition, real-time, Hidden Markov Model, interaction, interpretable**

I. Introduction

Various types of sensors, especially wearable sensors, are being more and more adequately used in the research and application of Human Activity Recognition (HAR) to acquire appropriate biosignals emitted from human bodies, including biomechanical signals from Inertial Measurement Units (IMUs) and electrogoniometers, bioelectrical signals such as electromyography (EMG) and electrocardiography (ECG), among others. Research results in related fields of sensor-based HAR have been emerging endlessly, such as wearable multimodal sensor signal acquisition [1]–[3], HAR research pipeline [4], Machine Learning (ML)-based human activity modeling, training, decoding, and evaluation [5]–[10], signal (pre-)processing, and feature extraction [11], feature space reduction [12], [13], manual, automatic, or semi-automatic data segmentation [14], [15]. Based on the offline outcomes of the above-listed aspects, real-time HAR systems are becoming increasingly applicable and stable for a wealth of practical application scenarios like auxiliary health care, sports, and edutainment.

For both offline and real-time HAR research, Neural Network (NN) and Hidden Markov Model (HMM) are the two most applied ML models with advantages and inconveniences in different tasks [7]–[9], [12], [13], [16], [17]. For example, algorithms such as Convolutional and Recurrent NN (CNN, RNN), superior in training simplicity and resource procurability, are currently inferior to HMM in terms of model interpretability [7].

Grounded by the experience in our initial version of the online HAR software [17] whose performance and functionalities have been academically recognized, our next real-time HAR system is proposed in this paper. This system is endowed with interactivity, interpretability, and expandability, enabling it to serve as an educational tool, a modeling research helper, and an algorithm visualizer.

II. Software

ASK2.0 provides three main modes: "Recording and Annotation", "Online Recognition", and "Plug and Play", in which data and labels can be recorded and stored, performed activities can be recognized in real-time, and new activities might be added to a recognizer. These functionalities are built from several easily swappable modules for recording devices, recognition models, and mode-dependent graphical user interfaces. The devices currently used in the demo are either the *biosignalsplux*[1] hubs with selectable multimodal sensors or the *BITalino R-IoT*[2] IMUs with Wi-Fi connection. The recognition and recognition visualization currently relies on HMMs as these are capable of real-time recognition and can be easily visualized and interpreted, but technically any ML model that matches these requirements could be plugged in. The three modes are introduced in the following subsections.

A. Recording and Annotation

Recording and annotating new data is supported by selecting an annotation stream, either a wearable sensor like a push button or a software button displayed onscreen. Asked to perform a (new) selected activity a few times, the user (or an experiment controller) presses, holds, and releases the button, similar to [16]. After the interface closes, the data is automatically placed in the selected folder, ready for further scientific processing or re-training in an upcoming recognition session.

[1]biosignalsplux.com
[2]https://plux.info/kits/376-bitalino-r-iot.html

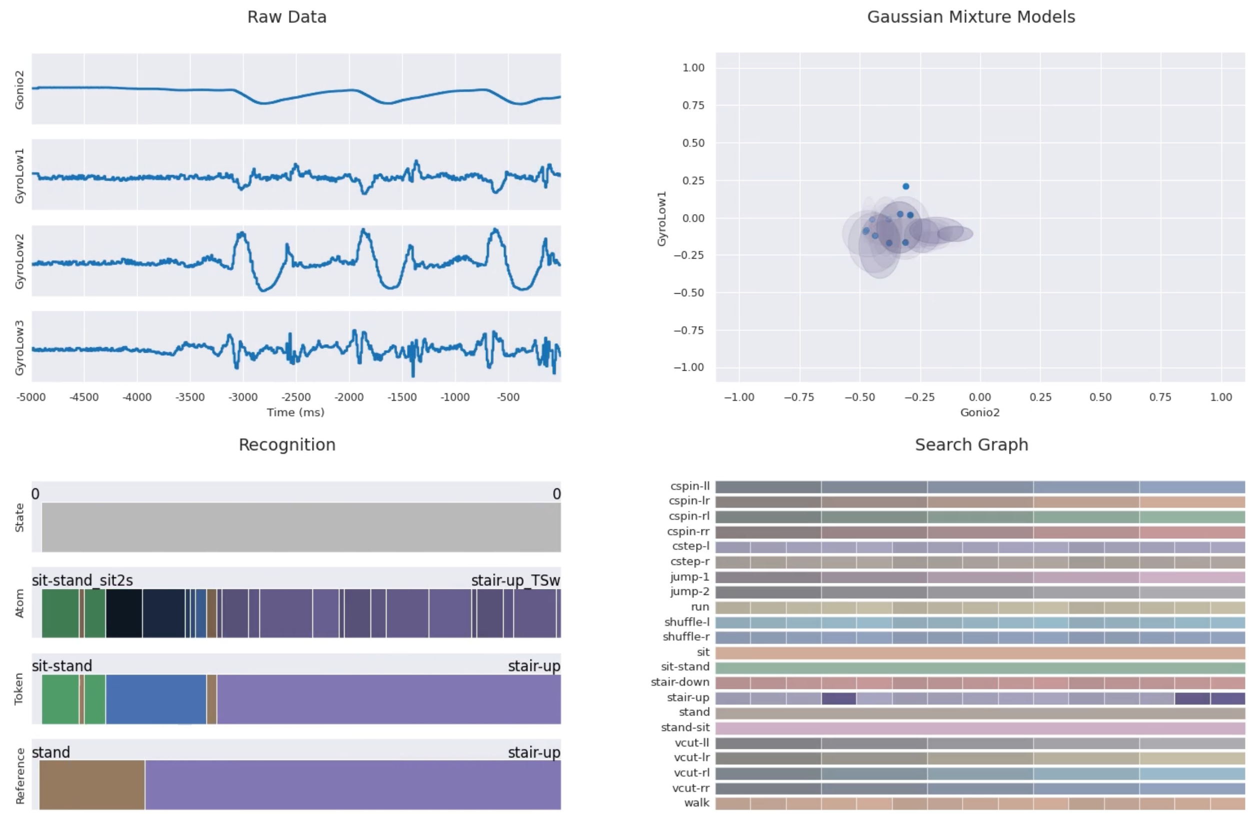

Fig. 1. Online Recognition Mode. The four subplots illustrate the raw data (upper left), the GMMs of the current top three hypotheses (top right), the search graph with the current top three hypotheses (lower right), and the final recognition results (lower left), respectively.

B. Online Recognition

The online recognition function is the core of the demo. It updates a user-selected pre-trained model every frame. The demo relies on HMMs with Gaussian Mixture Models (GMMS) as these have proven potent for HAR and allow for good interpretability and visualization. However, any other models or preprocessing steps could be incorporated and displayed. The HMM modeling of known activities is adapted from previous work by [7], [12], [13], [16], [17], while new activities are modeled with a single state. The states could be automatically inferred from known activities henceforth.

The recognition is the most demanding view running at 30 FPS on moderate hardware and 10 FPS on a *Raspberry Pi* due to the complex visualization requiring most of the computation power. For a good viewing experience and user interaction, the figure size was set to 16x10 inches. With smaller figure sizes or fewer graphs, higher FPS are possible, reaching 55 FPS with an 8x5 inch figure and 95 FPS if combined with only showing the recognition output. Figure 1 shows the recognition mode of the demo.

The upper left plot shows the last five seconds of data for several user-selected channels. In this case, the goniometer and three gyroscope axes are selected, both recorded at 100 Hz.

The upper right plot shows the first two of these channels to each other. The last goniometer value is plotted against the last gyroscope value in a scatter plot. Time is visualized as alpha values: the older a value, the more transparent it is. The plot additionally shows the Gaussians of each current node in the HMM search graph according to the top three hypotheses. The alpha value indicates weight: the more important a Gaussian is in the mixture, the less transparent it is.

The lower right plot displays the search graph itself. Each activity is given a bar, where each segment corresponds to a state in the HMM. In this plot, "Stair up" is modeled as three steps up a stair with five phases each. Most researchers are likely accustomed to observing the search graph as a Directed Acyclic Graph (DAG), as is often depicted in the literature. However, the bars remove complexity, making live observation and color changes easier to follow. The three most likely states are highlighted in the same color as the Gaussians in the plot above ("Gaussian Mixture Models"). In this example, the states are all in the "Stair up" activity.

The lower left plot depicts the final recognition result in sync with the plot above ("Raw Data"), i.e., the last five seconds. In the live demo, recognition results may change if newly observed data implies a more likely other explanation, but the older a result, the more stable it becomes. The result uses the same previously established colors and combines four sub-plots. The lowest is the reference (only available in playback of recorded data with known labels). The "Token"-level represents complete activities built from activity phases, called "Atom" here. The state-level is unused in the example but would represent the start/middle/end of a phase. All plots are updated in real-time and can be enabled/disabled by the user.

C. Plug and Play

The demo supports easy model extension and re-training. Data of a new activity may either be added into the corresponding folder or newly recorded (see section II-A). In a background task, the model is re-trained on-device with new activities, and the new model is loaded when switching into recognition mode – making the new activity directly recognizable. On-the-fly activity addition without re-training is viable with models that excel in zero-shot learning or utilize an automatic combination of already known activity phases. An on-the-fly zero-shot learning extension utilizing activity attributes adapted from [13] for straightforward interpretation is underway.

D. Recognition Accuracy

ASK2.0 can load different machine learning models. The model used for Figure 1 is built with the CSL-SHARE dataset [3], which contains 22 activities recorded with 19 sensor channels from 20 subjects. The GMM-HMM-based model was evaluated in a 5-fold person-independent cross-validation predicting the whole sequence and performing at $84.5 \pm 5\%$ frame-wise recognition accuracy on the labeled segments. This performance compares well to the segment-wise classification performance of 93.7% reported in [12].

E. Interaction

These components, "recording and annotation", "online recognition and visualization", and "plug and play" combined with live input from body-worn sensors, allow students, researchers, and the general public to interact with HAR in real-time. The demo is intended for people to probe the models and experience the capabilities and limitations of real-time HAR and ML first-hand with the guidance and explanations of a HAR-Expert.

III. Future Work

Many potential functionality upgrades can be further developed in the future, like changing the selected channels and graphs on the fly, models that support zero-shot learning or can model new activities automatically, intermediate renderings of features or pose estimations, or even a slow-motion plotting mode that catches up afterward or uses a longer timeline. Applying the system to other domains for teaching and visualization would provide a further understanding of needed features. Provided models from that domain are used and visualizations of features are adjusted, domains like (silent-)speech, attention, or video processing would be possible. However, most important is demonstrating the system and learning from interactions, explanations, and intuitions from students, researchers, and non-experts to learn and grow the demo, with which we aim to render HAR and ML vividly understandable and interactively experienceable.

References

[1] B. Hu, E. Rouse, and L. Hargrove, "Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals," *Frontiers in Robotics and AI*, vol. 5, p. 14, 2018.

[2] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha *et al.*, "Collecting complex activity datasets in highly rich networked sensor environments," in *INSS 2010 - 7th International Conference on Networked Sensing Systems*. IEEE, 2010, pp. 233–240.

[3] H. Liu, Y. Hartmann, and T. Schultz, "CSL-SHARE: A multimodal wearable sensor-based human activity dataset," *Frontiers in Computer Science*, 2021.

[4] ——, "A practical wearable sensor-based human activity recognition research pipeline," in *BIODEVICES 2022 — 15th International Conference on Biomedical Electronics and Devices*, INSTICC. SciTePress, 2022, forthcoming.

[5] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, "Human Activity Recognition using Inertial, Physiological and Environmental Sensors: a Comprehensive Survey," vol. 1, no. 1, apr 2020.

[6] Ó. D. Lara, M. A. Labrador, O. D. Lara, and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[7] H. Liu, Y. Hartmann, and T. Schultz, "Motion Units: Generalized sequence modeling of human activities for sensor-based activity recognition," in *EUSIPCO 2021 — 29th European Signal Processing Conference*, 2021, pp. 1506–1510.

[8] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.

[9] J. Suto, S. Oniga, C. Lung, and I. Orha, "Comparison of offline and real-time human activity recognition results using machine learning techniques," *Neural computing and applications*, vol. 32, no. 20, pp. 15 673–15 686, 2020.

[10] W. S. Lima, E. Souto, K. El-Khatib, R. Jalali, and J. Gama, "Human activity recognition using inertial sensors in a smartphone: An overview," *Sensors (Switzerland)*, vol. 19, no. 14, pp. 14–16, 2019.

[11] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "TSFEL: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[12] Y. Hartmann, H. Liu, and T. Schultz, "Feature space reduction for human activity recognition based on multi-channel biosignals," in *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies*, INSTICC. SciTePress, 2021, pp. 215–222.

[13] Y. Hartmann, H. Liu, S. Lahrberg, and T. Schultz, "Interpretable high-level features for human activity recognition," in *BIOSIGNALS 2022 — 15th International Conference on Bio-inspired Systems and Signal Processing*, INSTICC. SciTePress, 2022, forthcoming.

[14] D. Bouchard, "Automated time series segmentation for human motion analysis," *Center for Human Modeling and Simulation, University of Pennsylvania*, 2006.

[15] E. Guenterberg, S. Ostadabbas, H. Ghasemzadeh, and R. Jafari, "An automatic segmentation technique in body sensor networks based on signal energy," in *Proceedings of the Fourth International Conference on Body Area Networks*. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2009.

[16] H. Liu and T. Schultz, "ASK: A framework for data acquisition and activity recognition," in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOSIGNALS*, INSTICC. SciTePress, 2018, pp. 262–268.

[17] ——, "A wearable real-time human activity recognition system using biosensors integrated into a knee bandage," in *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 1: BIODEVICES*, INSTICC. SciTePress, 2019, pp. 47–55.