HIERARCHICAL ACTIVITY CLUSTERING ANALYSIS FOR ROBUST GRAPHICAL STRUCTURE RECOVERY

Namita Lokare, Daniel Benavides, Sahil Juneja and Edgar Lobaton
North Carolina State University
NC STATE UNIVERSITY Electrical & Computer Engineering

ABSTRACT

- We propose a hierarchical activity clustering methodology, which incorporates the use of topological persistence analysis, to capture the hierarchies present in the data.
- The key innovations presented in this research include the hierarchical characterization of the activities over a temporal parameter as well as the characterization and parameter selection based on stability of the results using persistence analysis.

DATASET

- Motion capture data was recorded for activities shown in Table 1. We collected individual activities as well as some transitions

Activity	Time (minute)
Bicycle	1
Walking	1
Sitting	1
Golf	1
Waving	1

Table 1. Protocol

METHOD

Assumptions:

- Let $x(t)$ be the sensor observations at time t.
- Let $\gamma_{k,\tau}: [0,\tau] \to \mathbb{R}^N$ be sensor observation trajectories over $[t_k, t_k+\tau]$.
- That is, $\gamma_{k,\tau}(t) := x(t+t_k)$
- We define a dissimilarity metric $D(\gamma_{k_1,\tau}, \gamma_{k_2,\tau})$ and construct point clouds as in Fig.1

$$D(\gamma_{k_1,\tau}, \gamma_{k_2,\tau}) = \underset{a\,\in\gamma_{k_1,\tau}}{mean}\{\min_{b\in\gamma_{k_2,\tau}}\{\boldsymbol{d(a,b)}\}\}$$

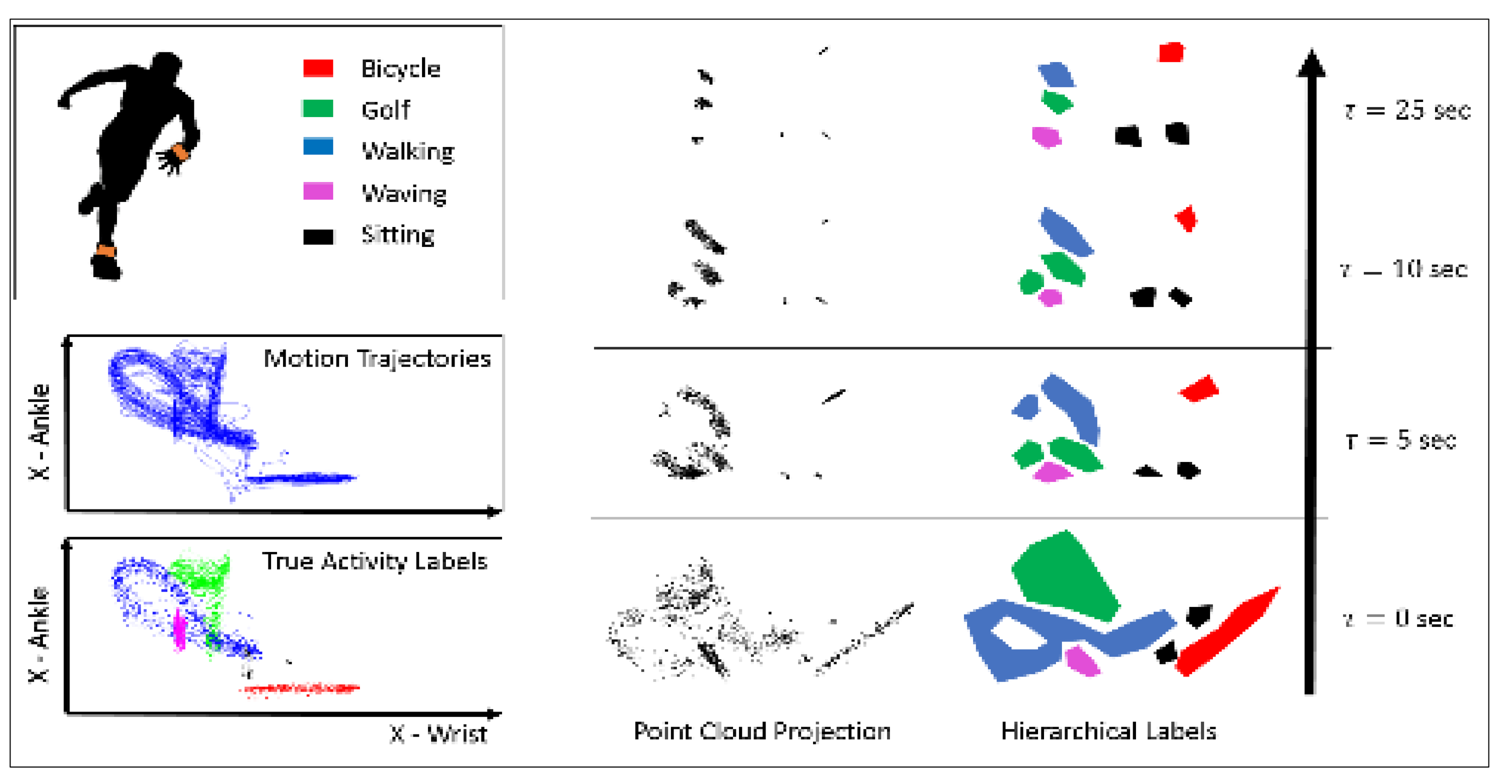

Fig. 1 Construction of point clouds from data streams. Motion trajectories and corresponding activity labels

METHOD Continued

Approach:

- We perform a filtration of the space for each by using the dissimilarity metric for a fixed τ.
- A multi-scale representation is obtained by considering the structure of the data over τ.
- Persistence homology allows us to keep a record of which clusters persist as shown in Fig.2.

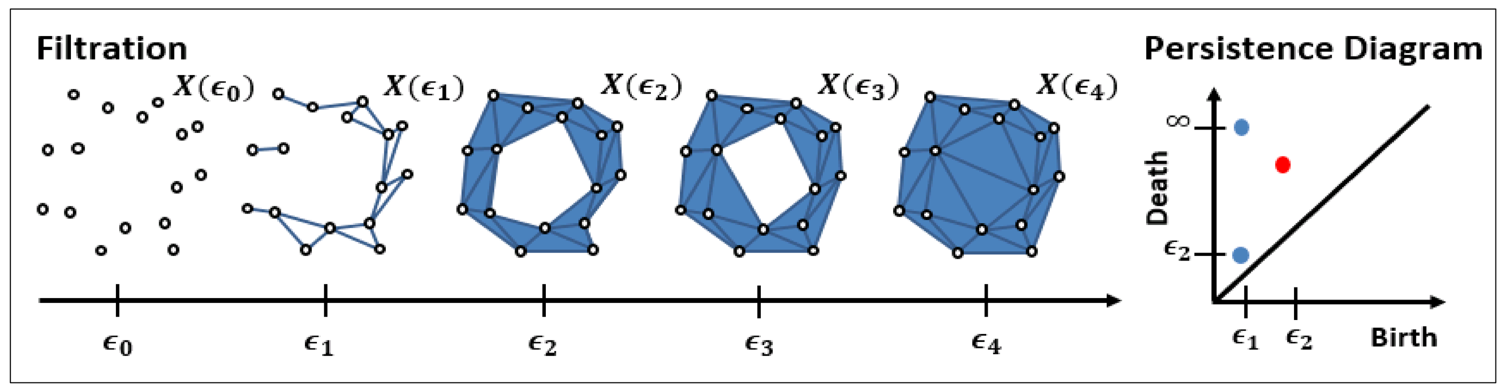

Fig. 2. Persistent Homology. A filtration of set of points based on a specified metric and its persistence diagram.

Training

- We divided the dataset into test (activities and transitions) and train(individual activities).
- Hierarchical clustering is performed using the point clouds for each value of τ $(1-40)$as the clustering parameter ϵ (3-60)is increased.
- Regions with higher densities are clustered into a single cluster and labeling was performed by using the majority vote rule.

Aggregated Persistence Diagrams

- We construct the hierarchical model by connecting the overlapping clusters over consecutive values of τ. Persistence diagrams from τ levels are aggregated in a diagram.
- Regions in the diagram corresponding to the three values of ϵ that produce the most robust graphical models are shown in (d), (e) and (f).
- The selected models will not change if ϵ is perturbed within the specified ranges. These ranges are also highlighted as pink regions in the diagram.

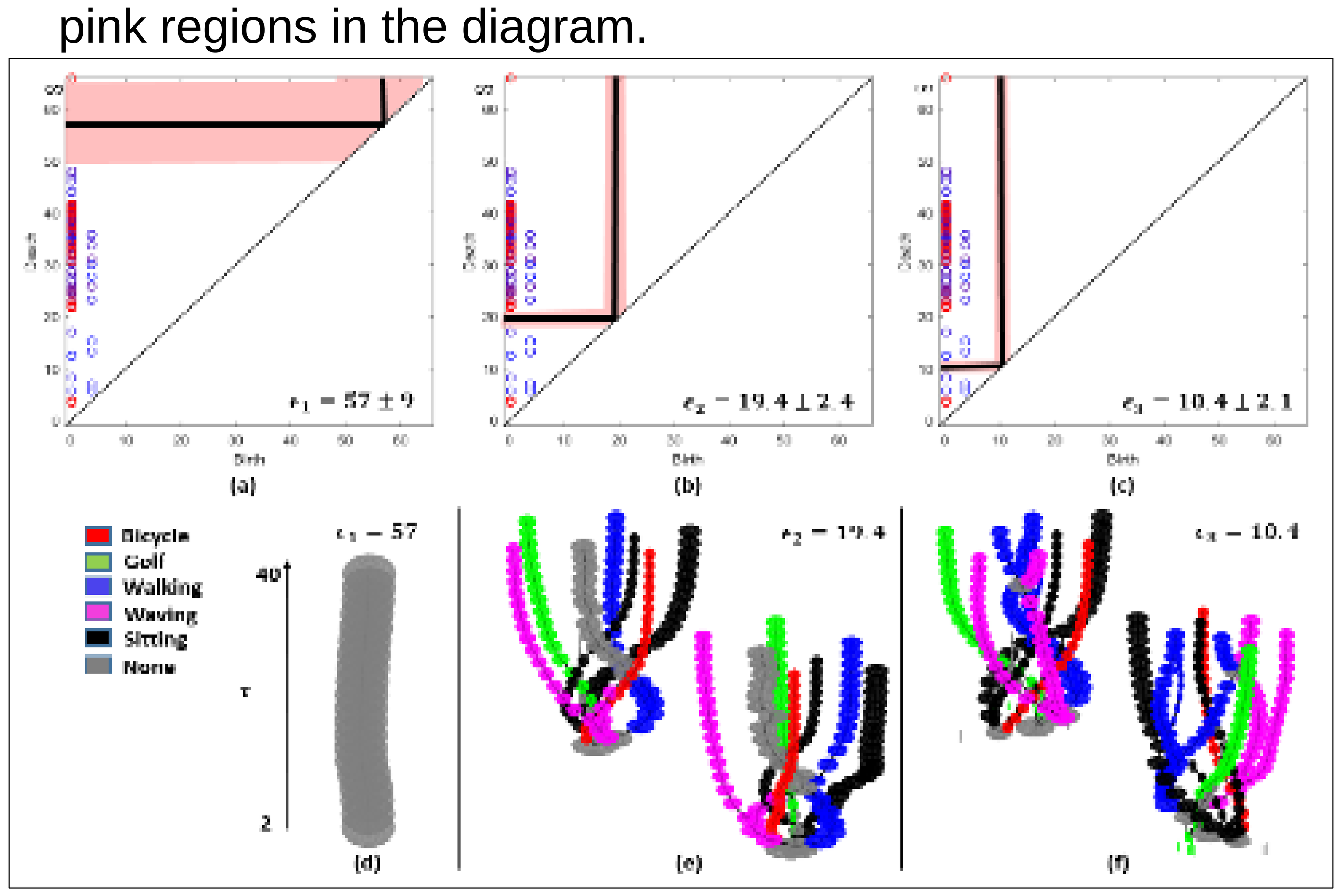

Fig. 4 Aggregated persistence diagrams shown in (a),(b) & (c). Correspond to the connected components for all levels of τ .

RESULTS

Results & Analysis

- Classification results shown in Fig.5, show the average class accuracy.
- Picking a small τ (i.e., the size of the data window) and a large enough ϵ (i.e., the clustering radius) gives us a better classification accuracy.
- The models are stable over the bound specified in the diagrams in Fig. 4.

$$Class\ Accuracy = \frac{True\ Positive + False\ Positive}{True\ Positive + True\ Neagtive + False\ Positive + False\ Negative}$$

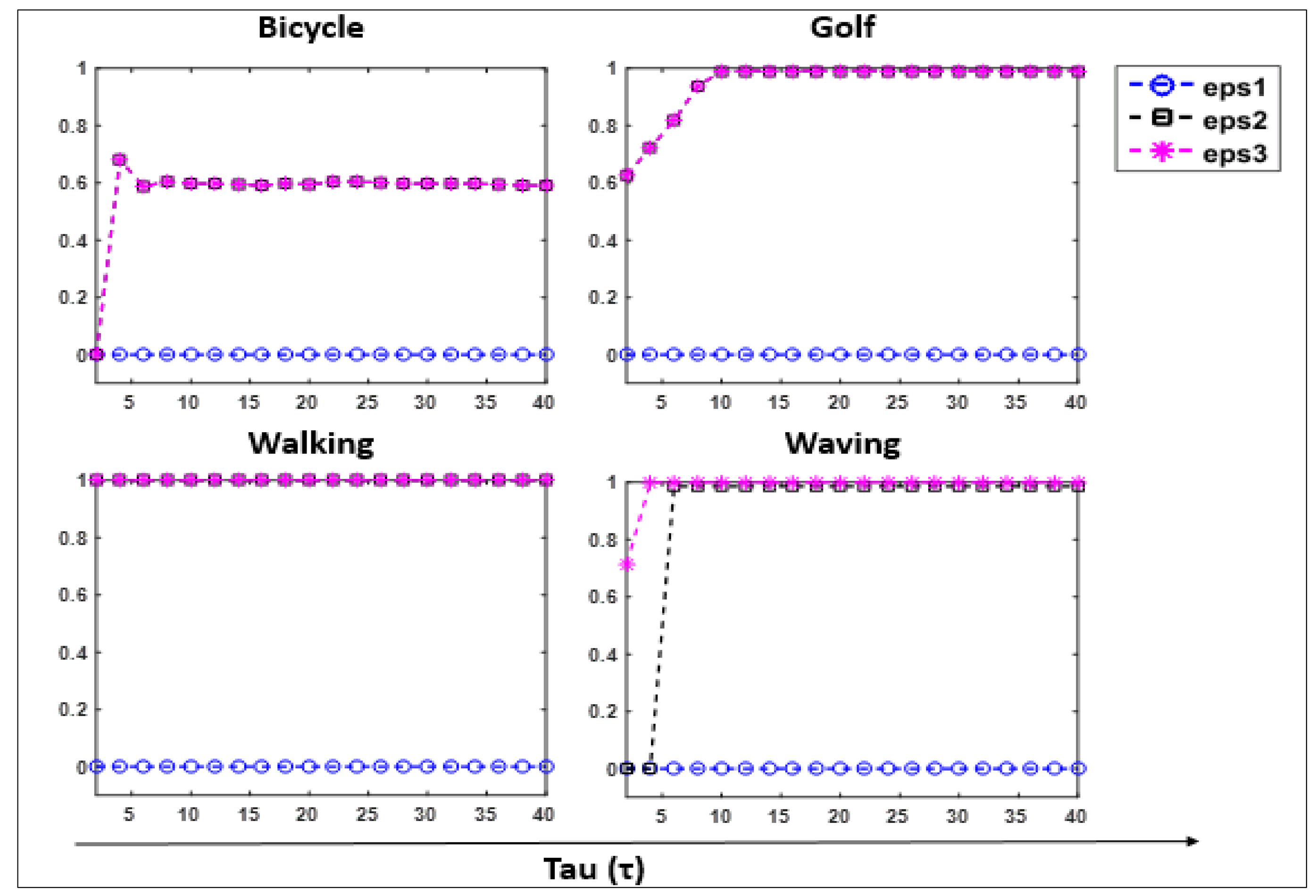

Fig. 5. Average class accuracy for each activity

- We conclude that values of τ anywhere between 10-40 and ϵ = 19.2 would give us an accuracy of 0.60 for bicycling, 1 for golf, 1 for waving and 1 for walking
- We see lower performance in the bicycling activity because for one of the repetitions bicycling was misclassified as sitting.

CONCLUSION

- We show how persistence diagrams can help reduce computation time and help choose stable models for our hierarchical representations.
- Furthermore, we are able to characterize the stability of our results via persistence analysis.
- Our future work will involve testing our method on other datasets and comparing it with other existing algorithm.

This work was partially supported by the National Science Foundation (NSF) ASSIST Nanosystems ERC under Award Number EEC-1160483.