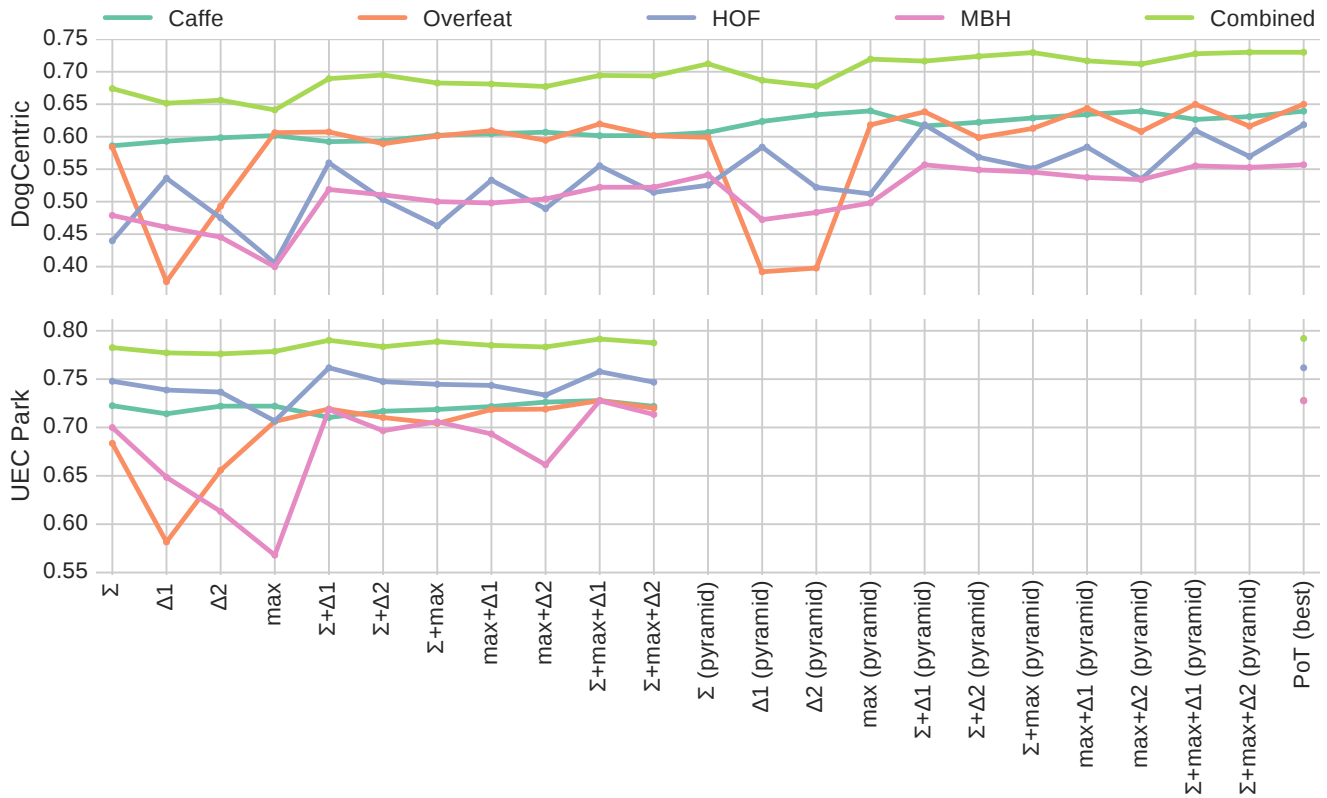# Pooled Motion Features for First-Person Videos
- supplementary material -

M. S. Ryoo, Brandon Rothrock , and Larry Matthies

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

`mryoo@jpl.nasa.gov`

Figure 1: Performance of each feature and pooling method for the DogCentric and UEC Park datasets. Temporal pyramid pooling was not used for UEC Park due to the arbitrary segmentation of the videos: these videos were segmented every 2 seconds without considering any temporal alignment.

Appendix

In this appendix, we provide detailed experimental results of our *pooled time series* (PoT) representations. We show activity recognition performances of our PoT per each different temporal pooling operator, as mentioned in Section 3.2 of the paper: sum pooling, max pooling, and two types of our new 'histogram of time series gradients' pooling. Results with individual pooling operators as well as possible combinations of them are reported. Both the Dog-centric activity dataset and the UEC Park dataset were used.

Figure 1 illustrates the results with different pooling operators. The top graphs are for the DogCentric activity dataset, and the bottom graphs are for the UEC Park dataset. These graphs not only suggest that using multiple different types of temporal pooling operators helps overall recognition (compared to using just one), but also implies that our PoT is a good representation to combine different types of feature descriptors (e.g., HOF, CNN features, and so on).