

# TPA 11: Categorization of Human Actions from Videos

January 26, 2017

**Problem Statement:** The aim of the project is to build a system that can categorize large number of videos according to a set of complex human actions being performed. The videos feature large intra-class variations as well as several challenges in the form of camera motion, jitter, multiple persons and low quality.

**Input:** A video of arbitrary time length.

**Expected Output:** Table of classification accuracy and signature map of action area.

## Dataset:

- **UCF-101:** This dataset contains 13320 annotated videos belonging to 101 classes having 180 frames/video on average. The evaluation will be done on the three standard train/test splits following the scheme of THUMOS challenge.
- **HMDB-51:** A large collection of real-world videos gathered from various sources such as web and movies. This dataset comprises of 6766 videos across 51 classes, where each class consists of at least 100 variable length video clips.

## References

- X. Peng, C. Zou, Y. Qiao, and Q. Peng. Action recognition with stacked fisher vectors. In European Conference on Computer Vision, pages 581-595. Springer, 2014.
- Z. Lan, M. Lin, X. Li, A. G. Hauptmann, and B. Raj. Beyond gaussian pyramid: Multi-skip feature stacking for action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 204-212, 2015.

- K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems*, pages 568-576, 2014.
- Mahasseni, Behrooz, and Sinisa Todorovic. Regularizing long short term memory with 3D human-skeleton sequences for action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3054-3062, 2016.