

# Learning to Observe, Query and Justify Behavior from Human Activity Video Clips

Patrick Emaase<sup>o</sup> and Byoung-Tak Zhang  
Seoul National University

## Abstract

How do people effortlessly infer an activity from video clips? How can one retrograde a given ultimate activity? These are valid questions to explore and exploit the RGB-D video clips of human activities. Human activities is a rich combination of diverse actors, actions, subjects, objects, context and inherent relationships defined over a period of time, hence difficult to regenerate for learning. In this paper, we focus on a special case of activity reason generation given activity video clips, where the goal is to generate an unambiguous activity descriptions of that applies to exactly one video clip or collection of video frames especially from a robot's perspective in order to learn human behavior. Such description is known as "activity reasoning". This approach has a major advantage over image captioning and activity simulation since it has a well-defined performance metric: activity reasoning is considered to be good if it uniquely describes the activity reasons within its context congruent to the subject, objects and inferred objective. In addition, activity reasoning is detailed necessary for incremental abstraction of facts and novel plannable procedures to realize truly intelligent system with discriminative power. Further, video based reasoning would be practical to deploy in an interactive activity monitory and unambiguous recommendation.

We consider three problems: (1) Activity localization – we localize activities in video clips as they progress by decomposing activities into actionlets – action constituent parts that richly describe the increment transitions. Later we generate an entire activity as composition of various components. (2) Activity description – this is where we discriminatively select action part to attend and interactively using relevant referential expressions to describe activity details and progression and (3) Activity reasoning – to learn dynamic reason in an end-to-end manner.

## Acknowledgement

This work was partly supported by the Korea government (IITP-R0126-16-1072-SW.StarLab, KEIT-10044009-HRI.MESSI, KEIT-10060086-RISF), and the ICT at Seoul National University.

## References

- Gal, Y., & Ghahramani, Z. (2015). Bayesian convolutional neural networks with Bernoulli approximate variational inference. arXiv preprint arXiv:1506.02158.
- Zhang, B. T. (2008). Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory. IEEE computational intelligence magazine, 3(3).