A Benchmark Dataset and Comparison Study for Multi-modal Human Action Analytics

JIAYING LIU, SIJIE SONG, CHUNHUI LIU, YANGHAO LI, and YUEYU HU,
Institute of Computer Science and Technology, Peking University, China

Large-scale benchmarks provide a solid foundation for the development of action analytics. Most of the previous activity benchmarks focus on analyzing actions in RGB videos. There is a lack of large-scale and high-quality benchmarks for multi-modal action analytics. In this article, we introduce PKU Multi-Modal Dataset (PKU-MMD), a new large-scale benchmark for multi-modal human action analytics. It consists of about 28,000 action instances and 6.2 million frames in total and provides high-quality multi-modal data sources, including RGB, depth, infrared radiation (IR), and skeletons. To make PKU-MMD more practical, our dataset comprises two subsets under different settings for action understanding, namely Part I and Part II. Part I contains 1,076 untrimmed video sequences with 51 action classes performed by 66 subjects, while Part II contains 1,009 untrimmed video sequences with 41 action classes performed by 13 subjects. Compared to Part I, Part II is more challenging due to short action intervals, concurrent actions and heavy occlusion. PKU-MMD can be leveraged in two scenarios: action recognition with trimmed video clips and action detection with untrimmed video sequences. For each scenario, we provide benchmark performance on both subsets by conducting different methods with different modalities under two evaluation protocols, respectively. Experimental results show that PKU-MMD is a significant challenge to many state-of-the-art methods. We further illustrate that the features learned on PKU-MMD can be well transferred to other datasets. We believe this large-scale dataset will boost the research in the field of action analytics for the community.

CCS Concepts: • **Computing methodologies → Activity recognition and understanding**; *Supervised learning by classification*;

Additional Key Words and Phrases: Benchmark, multi-modal, action detection, action recognition

ACM Reference format:
Jiaying Liu, Sijie Song, Chunhui Liu, Yanghao Li, and Yueyu Hu. 2020. A Benchmark Dataset and Comparison Study for Multi-modal Human Action Analytics. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 2, Article 41 (May 2020), 24 pages.
https://doi.org/10.1145/3365212

This work was supported by National Natural Science Foundation of China under contract No. 61772043, Beijing Natural Science Foundation under contract No. 4192025, Microsoft Research Asia (FY19-Research-Sponsorship-115) and Peking University Tencent Rhino Bird Innovation Fund.

Authors' addresses: J. Liu, S. Song, C. Liu, Y. Li, and Y. Hu, Institute of Computer Science and Technology, Peking University, Zhongguancun North Street 128#, Haidian, Beijing, China; emails: {liujiaying, ssj940920, liuchunhui, lyttonhao, huyy}@pku.edu.cn.

1551-6857/2020/05-ART41 $15.00
https://doi.org/10.1145/3365212