A Decade of Action Quality Assessment: Largest Systematic Survey of Trends, Challenges, and Future Directions

Hao Yin[1,2†], Paritosh Parmar[3†], Daoliang Xu[1,2], Yang Zhang[2], Tianyou Zheng[2*], Weiwei Fu[1,2*]

[1*]School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China.
[2*]Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Science, Suzhou, Jiangsu, China.
[3]Institute of High-Performance Computing, Agency for Science, Technology and Research, Singapore.

*Corresponding author(s). E-mail(s): zhengty@sibet.ac.cn; fuww@sibet.ac.cn;
Contributing authors: hao_yin@mail.ustc.edu.cn; paritosh.parmar@alumni.unlv.edu; daoliang@mail.ustc.edu.cn; zhangyang@sibet.ac.cn;
†These authors contributed equally to this work.

Abstract

Automated Action Quality Assessment (AQA)—the ability to quantify the quality of human motion, actions, or skill levels and provide feedback—has far-reaching implications in areas such as low-cost physiotherapy, sports training, and workforce development. As such, it has become a critical field in Computer Vision and Video Understanding over the past decade. Significant progress has been made in AQA methodologies, datasets, and applications, yet there remains a pressing need for a comprehensive synthesis of this rapidly evolving field. In this paper, we present a thorough survey of the AQA landscape, systematically reviewing over 200 research papers using the PRISMA framework. We begin by covering foundational concepts and definitions, then move to general frameworks and performance metrics, and finally discuss the latest advances in methods and datasets. Our survey provides a detailed analysis of research trends, performance comparisons, challenges, and future directions. Through this work, we aim to offer a valuable resource for both newcomers and experienced researchers, promoting further exploration and progress in AQA.

Keywords: action quality assessment, skills assessment, action understanding, video understanding, computer vision, deep learning, survey