World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:17, No:07, 2023

A Transformer-Based Approach for Multi-Human 3D Pose Estimation Using Color and Depth Images

Authors : Qiang Wang, Hongyang Yu

Abstract : Multi-human 3D pose estimation is a challenging task in computer vision, which aims to recover the 3D joint locations of multiple people from multi-view images. In contrast to traditional methods, which typically only use color (RGB) images as input, our approach utilizes both color and depth (D) information contained in RGB-D images. We also employ a transformer-based model as the backbone of our approach, which is able to capture long-range dependencies and has been shown to perform well on various sequence modeling tasks. Our method is trained and tested on the Carnegie Mellon University (CMU) Panoptic dataset, which contains a diverse set of indoor and outdoor scenes with multiple people in varying poses and clothing. We evaluate the performance of our model on the standard 3D pose estimation metrics of mean per-joint position error (MPJPE). Our results show that the transformer-based approach outperforms traditional methods and achieves competitive results on the CMU Panoptic dataset. We also perform an ablation study to understand the impact of different design choices on the overall performance of the model. In summary, our work demonstrates the effectiveness of using a transformer-based approach with RGB-D images for multi-human 3D pose estimation and has potential applications in real-world scenarios such as human-computer interaction, robotics, and augmented reality.

Keywords : multi-human 3D pose estimation, RGB-D images, transformer, 3D joint locations
Conference Title : ICSLP 2023 : International Conference on Speech and Language Processing
Conference Location : Stockholm, Sweden
Conference Dates : July 06-07, 2023