

•Editorial•

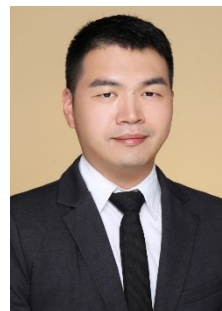
3D vision for virtual reality and augmented reality

Guofeng ZHANG

State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310058, China

zhangguofeng@cad.zju.edu.cn

3D vision is a kind of technology that allow computers to perceive, reconstruct and interact with the 3D world based on vision sensors. It is not only a hot academic topic in computer science, but also crucial to many applications, such as virtual reality (VR) and augmented reality (AR). For achieving high-quality effects of VR and AR, we need to recover the 6DoF camera pose, 3D structure of the scene, and even the human interaction, so that people not only can see the lifelike virtual objects/scenes, but also the seamless fusion of virtual and real contents and even interact with them. We can leverage 3D vision technology to achieve this objective. For example, the simultaneous localization and mapping (SLAM) technique allows users to walk freely in a digital world wearing a VR/AR helmet by recovering the 6DoF movement of the helmet, and is also able to see the seamless mixture of real and virtual contents, and even the effect of their interaction with each other by reconstructing the geometry, textures and materials of the real environment. In this special issue, we have selected 6 papers (3 review papers, 2 research articles and 1 case report) covering 3 most relevant 3D vision techniques for the VR/AR applications, including SLAM, 3D reconstruction, object segmentation and tracking.



SLAM is a fundamental problem in 3D vision and robotics community. The traditional SLAM methods are based on multi-view geometry and non-linear optimization, and most of the recent interests are to leverage the power of deep learning to improve the robustness. Yan and Zha gave a review of the evolution of SLAM in detail, covering both the traditional methods based on multi-view geometry and the emerging learning based methods. The review specifically focused on the nature of temporal continuity (denoted as flow-based nature), which could be leveraged to improve the robustness. Zou et al. gave another review focusing on the SLAM techniques applied to multiple independently moving agents, named collaborative SLAM in literature. Compared to typical SLAM deployed on a single agent, the information in collaborative SLAM can be exchanged or shared among different agents, so better performance can be achieved. In addition, with collaborative SLAM, multiple users holding VR/AR devices in the same environment can be registered together and are able to interact with each other. The article paper from Dong and Sheng focused on the recovery of depth and camera motion by learning from training data. They found that the success of many recent approaches shows that the tricks (e.g., data augmentation and learning objectives) in the training procedure can usually improve the performance without introducing additional computational cost. This article investigated a collection of such tricks, by both theoretical examination and empirical evaluation of the final accuracy variation of visual odometry, to get a comprehensive analysis.

Typical visual SLAM techniques represent the 3D environment with sparse (at most semi-dense) points. In order to interact with the real world more naturally, the 3D environment should be densely reconstructed. Xu et al. surveyed the latest developments of 3D reconstruction based on depth cameras,

including algorithmic components of camera pose estimation, 3D reconstruction of objects or scenes, and texture mapping. Gao et al. presented a case report focusing on the reconstruction of large-scale scenes, specifically the ancient Chinese architecture, which is somehow difficult due to the architectural complexity and structural delicacy. It suggested to tackle this difficulty by fusing multi-source data, including ground and aerial images and LiDAR. It also demonstrates the effectiveness of the fusion of ground and aerial images, the fusion of images and LiDAR data, scene surface reconstruction and semantic modeling based on the acquired multi-source data.

In some VR/AR applications, it is required to obtain the pose and depth for a non-rigid moving object rather than the static scene. Zhao et al. presented a system integrating segmentation and SLAM techniques to track the 6DoF camera pose relative to a specified human body part. With the pose and the segmentation results, the system is able to control the virtual objects or add virtual backgrounds.

Besides VR/AR, 3D vision is also crucial to some other applications like robotics, autonomous driving, unmanned aerial vehicle, etc. Although this special issue is focusing on VR/AR, we hope that the publication will also benefit other related applications and research communities.

Guofeng ZHANG

8 October 2019