Video Content Analysis for Augmented and Mixed Reality

Kiyoshi Kiyokawa, Osaka University

Augmented reality (AR) is a technology that overlays computer information onto the real world, thereby augmenting human capability. Mixed reality (MR) refers to a broader concept, the integration and merging of the real and virtual environments, that ranges between the completely virtual and the completely real. AR and MR have been intensively studied in the last few decades and have become fundamental technologies in computer systems for a wide variety of applications including medical, military, architecture, manufacturing, education, art and entertainment. Still, there are a number of issues with AR and MR technologies. Azuma identified three common characteristics of AR scenes; combination of the real and virtual, interaction in real-time, and scene registration in 3D. To maintain a consistent MR environment, geometric, photometric and temporal consistencies need to be taken into account. That is, virtual entities should appear at the correct location in the real environment, shading and illumination of the virtual entities should match to those in the real environment, and motions of the virtual and the real entities should be coordinated.

Video content analysis is a key technology in addressing these challenges in actual AR and MR systems. In this talk, trends and prospects in the field of video content analysis for AR and MR will be introduced, focusing primarily on the following three topics; geometric consistency, photometric consistency and scene understanding.

First, a brief history and cutting-edge examples for geometric consistency will be introduced. Geometric consistency is achieved by camera calibration and relative pose estimation between a camera and the scene. Major pose estimation methods include marker-based, model-based and natural feature based tracking. Natural feature tracking is a more advanced approach and is now main stream, as it assumes no prior knowledge of geometric features.

Second, several cutting-edge examples and challenges for photometric consistency will be introduced. Photometric consistency requires a more complex process, from geometric model acquisition, to reflection property acquisition, to lighting environment estimation. Camera properties such as depth of focus, motion blur, dynamic range and white balance should also be considered in rendering realistic virtual objects.

Lastly, scene understanding technologies will be briefly introduced as a frontier of video content analysis for AR and MR. Most current AR and MR applications simply overlay virtual content onto the real world background without considering what the user is actually paying attention to. However overlaid information will be useless or even harmful if it is presented at the wrong time or in the wrong situation. Scene understanding and context estimation using video based object recognition and gaze tracking will be very important in future AR and MR applications.