Scene Reconstruction using Corresponded Depth

Daniel G. Aliaga

Capturing and modeling 3D scenes is an important goal for several applications in computer graphics, computer vision, and geometric modeling. Such a digital model serves to enable telepresence, gaming and simulations, and several forms of virtual reality. Single viewpoint scene reconstruction is limited to devices that provide depth (e.g., time of flight lasers) and to the reconstruction of the surfaces visible from the single viewpoint. To support reconstruct without using per-pixel time-of-flight systems, a multi-view reconstruction system uses two or more calibrated views to reconstruct the scene. The at least two views are separated by a known baseline where higher accuracy acquisition is obtained by a larger baseline. However a large baseline capture reduces the amount of surfaces visible from both (all) viewpoints and, in fact, reduces to less surfaces than those visible from a single viewpoint. This fundamental limitation necessitates acquisition from additional viewpoint sets. The additional viewpoint sets must be either a priori calibrated relative to the first (e.g., static calibration) or calibrated on-the-fly (e.g., dynamic calibration, pose estimation, tracking, etc.). This significantly complicates the reconstruction process and requires both depth via triangulation and relative position and orientation estimation of the additional views. In this work, we overcome this fundamental limitation, by enabling a multi-view reconstruction of arbitrarily large scenes supporting arbitrary baselines and supporting compositing the reconstructions together without having to know the relative positions and orientations of the reconstructions. This provides a huge simplification to the reconstruction process and permits acquiring a scene by simply taking a set of pictures.