Performance Assessment and Calibration of the Kinect 2.0 Time–Of–Flight Range Camera for Use in Motion Capture Applications

Jeremy Steward, Derek Lichti, Jacky Chow, Reed Ferber and Sean Osis (Canada)

Key words: Low cost technology; Photogrammetry; Remote sensing; time of flight range camera; sensor calibration; motion capture; human motion

SUMMARY

Robust, 3D geometric information is a powerful analytical tool to assist in the understanding of planning operations during practical surveys. Specifically, it is of interest to be able to determine the size, shape, and geometric properties of objects in the real world when performing planning surveys, deformation analyses, or motion capture surveys. Traditional photogrammetric reconstruction techniques require multiple sensors for data capture and retro-reflective markers or targets to be attached to patients during activity. The Microsoft Kinect 2.0 sensor provides an onboard time-of-flight (ToF) ranging sensor based on the Canesta technology. Given the price for the Kinect 2.0 is $200 USD, it shows potential to become a cost-effective, single-sensor solution for capturing full 3D geometric information in place of costly, multi-sensor techniques requiring invasive or otherwise difficult to place markers. This study examines the performance characteristics and calibration of the Kinect 2.0 sensor in order to determine the feasibility of its use in 3D imaging applications; particularly that of human motion capture. The Kinect 2.0 sensor was tested under controlled conditions in order to determine the warm-up time, distance measurement precision, target reflectivity dependencies, residual systematic errors, and the quality of human body reconstruction when compared to a device of known quality. The sensor in question proved promising, showing similar precision to other ToF imaging systems at a mere fraction of the price. By using a standard photogrammetric bundle adjustment approach, systematic errors identified were successfully modelled. Over the course of this testing, it was found that negligible warm-up time is required before the geometric measurement performance stabilizes. Furthermore, a distance measurement precision of approximately 1.5mm is achievable when imaging highly reflective, diffuse target surfaces. Small cyclic errors were detected on the order of approximately 5mm; however this did not prove to have a large impact on human body measurements. Beyond the performance characteristics of the sensor itself, a self-calibration of the sensor for un-modelled lens distortions improved image measurement residuals by 78%, and likewise improved the range measurement precision by 68%. Despite these results, factors beyond the user's control such as scene-dependent distortions, and inhomogeneity in depth accuracy across the image plane continue to limit the potential performance of the sensor. Thus, the following "best practice" guidelines were put forth, in order to achieve the best performance possible. In simple terms: only the inner 300x300 pixels about the centre of the sensor should be used, due to loss in signal strength near the periphery of the image; and ensure that the object of interest is contained within the foreground of the scene, ideally at a range approximately 2m away from the sensor. Finally, highly-reflective, diffuse objects provide better precisions on range measurements, and should be preferred to darker or shiny objects in the captured scene if possible.