CLOUD-BASED DEPTH SENSING QUALITY FEEDBACK FOR INTERACTIVE 3D RECONSTRUCTION

Kar-Han Tan   *John Apostolopoulos*

Hewlett-Packard Laboratories

ABSTRACT

In this paper we propose a cloud-based approach to improve the 3D reconstruction capability of handheld devices with real-time depth sensors. We attempt to characterize the quality of 3D information captured by real time depth sensing devices, and in particular examine how sensors from Prime Sense and Canesta measure distances, and derive simple analytical models on performance limitations for each. We also study the factors that affect depth sensing quality when these devices are used to incrementally build larger or denser 3D models. Empirical experiments confirm our analysis. Our findings allow us to design a quality metric which can interactively inform users to guide them on how to optimize the quality of their captured 3D content.

Index Terms— TOF, Depth Sensing, interaction, quality feedback, 3D reconstruction

1. INTRODUCTION

With the successful introduction of affordable real time depth sensors into the consumer video gaming market[1], reliable depth sensing has become immediately accessible to researchers and practitioners. The availability of reliable depth information in addition to color enables the synthesis of new views in image-based rendering [2]. Real time time-of-flight(TOF) depth sensors can also be synergistically fused with multiview stereo for fast high quality 3D reconstruction [3].

As depth sensors become more power efficient, mobile devices can potentially turn into 3D acquisition instruments, allowing interactive reconstruction of 3D scenes. By repositioning a depth sensor and combining data from the different viewpoints, a mobile 3D acquisition device can overcome field-of-view limitations and create compelling, immersive experiences [1].

Another key component of a mobile 3D acquisition system is the computational requirements of creating 3D models from depth data. In this paper, we propose a cloud-based system for interactive 3D reconstruction using mobile depth sensors. Like in [4], the user uploads raw data to the cloud for processing. In our system, in addition to creating 3D models we also provide feedback to the user on the quality of the captured data. This allows users to easily see which parts of the 3D models are good and which parts can be improved. Based on the quality feedback, users can then capture additional depth data which can in turn be fused with previously captured data to produce a higher quality 3D model.

2. REALTIME DEPTH SENSING TECHNOLOGIES

To derive our quality feedback metric, we briefly review the leading realtime depth sensing technologies, as well as related work for combining a sequence of point clouds into larger 3D models.

2.1. Structured Light

The sensor used in the Kinect is made by PrimeSense [5], and captures a depth map by projecting a fixed pattern of spots with infrared light [6]. An IR camera captures the scene illuminated with the dot pattern and depth can be estimated based on the amount of displacement. Since the PrimeSense setup requires a baseline distance between the light source and camera, there is a minimum distance that objects need to be at, resulting in a ‘dead zone’ of about 0.8m in the case of the Kinect.

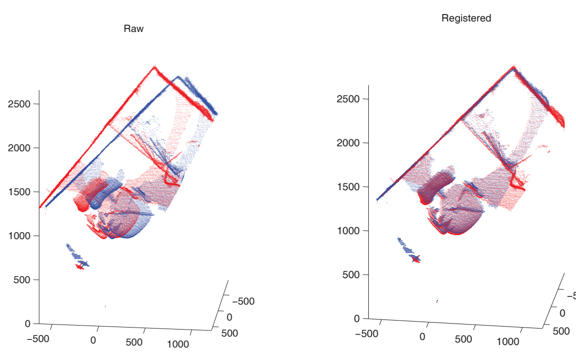
2.2. Per-pixel Time-of-flight.

A different class of depth sensors also uses infrared light sources, but instead of using spatial light patterns they send out temporally modulated IR light and measure phase shift of the returning light signal. The Canesta [7] and MESA [8] sensors employ custom CMOS/CCD sensors while the 3DV ZCam [9] employ a conventional image sensor with a gallium arsenide-based shutter. As the IR light sources can be placed close to the IR camera these ‘time of flight’ sensors are capable of measuring shorter distances.

2.3. Point Cloud Registration

Often it is desirable to piece together, or *register* depth data captured from a number of different positions. For example, to measure all sides of a cube, at least two depth maps captured from the front and back are necessary. While there are many techniques for aligning partially overlapping 3D points, perhaps the most widely used technique are variations of the Iterative Closest Point (ICP) algorithm [10]. At each step, the algorithm finds correspondence between a pair of 3D points and computes the rigid transformation which best aligns the point clouds. Fig. 1 shows an example of ICP registration. While it is typically rather computationally intensive, ICP optimized for GPUs [11] have reduced the computation time to enable interactive operation. With GPU acceleration and additional optimization, it has been shown that realtime interactive registration is possible [1].

The GPU acceleration approach is appropriate for desktop or workstation-class settings. To realize the potential of lightweight mobile depth sensing devices, we propose that a cloud-based service for registering 3D points be employed. This approach has previously been used in the PhotoSynth [4], a photo sharing web site where users can upload image collections of a scene and a cloud-based service applies structure-from-motion algorithms to recover a sparse set of 3D points and allow users to navigate the photo collection with an immersive 3D interface. We aim to provide a cloud-based ICP service that in addition to aligning the point clouds from mobile depth sensors, it provides interactive quality feedback to guide users.

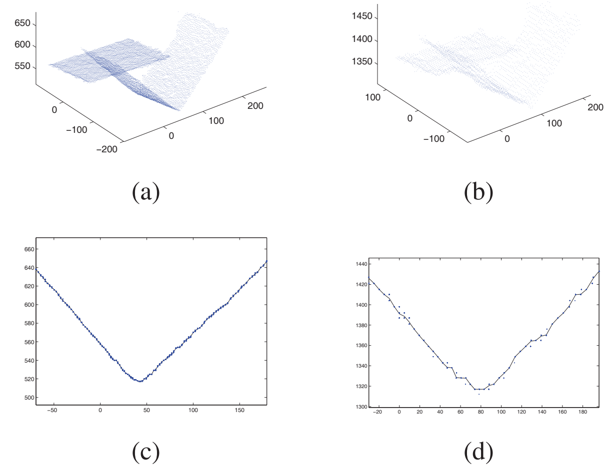


(a) Raw 3D Points (b) Registered

Fig. 1. Point cloud registration with ICP.

3. QUALITY OF DEPTH MEASUREMENTS

Consider the two point clouds of the same object shown in Fig. 2, captured using the same sensor. However it is clear that the two point clouds are different, with the left column showing a much higher quality data set than the right column. In particular, the left column has higher spatial density and lower noise levels. Given that these were captured with the same sensor, what caused the difference in the two point clouds? We found that the quality of raw depth data captured is heavily influenced by the following factors: sensor distance, sensor motion, and infrared signal strength.



(a) (b) (c) (d)

Fig. 2. Variations in the quality of captured depth: (a) and (b) are 3D scans of the same object, captured with the same depth sensor. (c) and (d) show cross section views of the same data. Clearly data in the left column is denser and less noisy.

Sensor Distance Fig. 3 shows a simplified illustration of how distances are measured in a structured light-based depth sensor such as the PrimeSense product. The light source and camera are positioned along the x-axis (drawn vertically), with a baseline separation B. When the camera observes a scene point at depth z illuminated by the light ray (drawn horizontally), the light spot is projected onto a location with coordinate x on the camera sensor. The geometry gives us the following:

$$\begin{aligned}\frac{B+x}{z} &= \frac{x}{F}\\ F(B+x) &= zx\\ FB &= x(z-F)\\ x &= \frac{FB}{z-F}\\ \frac{dx}{dz} &= \frac{-zFB}{z^2} = \frac{-FB}{z}\end{aligned} \tag{1}$$

Equation (1) shows that as the scene point distance z goes to infinity, the rate of change in observation coordinate goes to 0. This shows that a structured light-based depth sensor measuring distances by triangulation becomes less precise with distance, and therefore more susceptible to noise. Per-pixel time-of-flight sensors do not use triangulation, but instead rely on measuring the intensity of returning light. Of course, the intensity of light is inversely proportional to the square of the distance from the source.

Sensor Motion Relative motion between the sensor and the scene can also degrade depth measurements. In the case of structure light sensors, observations of the light spots may become blurred, making detection difficult and also making

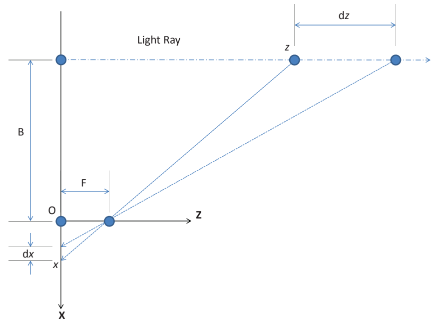


Fig. 3. Geometric relationships in a structured light setup.

localization less precise. In the case of TOF sensors, motion violates the assumption that each pixel is measuring a single scene point distance.

Signal Strength In addition to light fall off with distance, different parts of the scene may reflect varying amounts of infrared light that the sensors need to capture. If an object absorbs and does not reflect IR light, it becomes challenging for structured light sensors to observe the light spots. For TOF sensors, the diminished intensity reduces the precision of the sensor.

4. INTERACTIVE QUALITY FEEDBACK

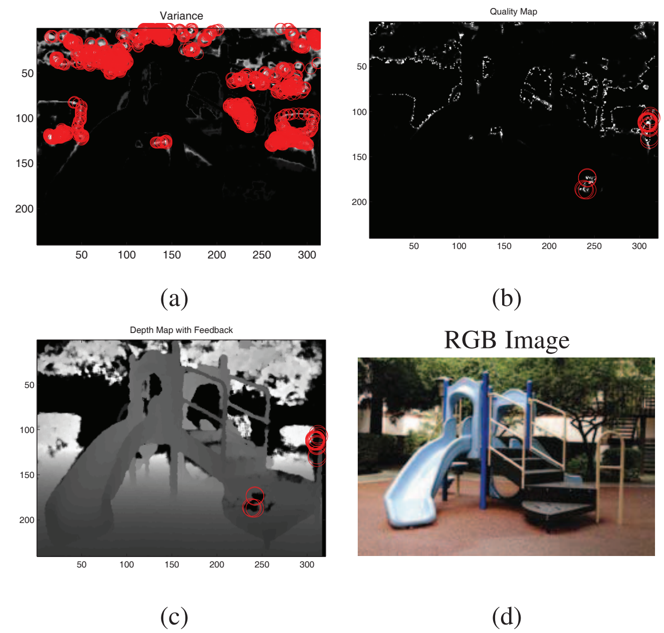
For experiments we constructed a prototype mobile 3D capture device using a notebook computer and a PrimeSense-based depth sensor. The system, shown in Fig. 4, captures a sequence of depth maps which is then sent to a server for registration. The resulting reconstruction along with the depth map quality feedback information is sent back to the device for display.

Fig. 4. Experimental mobile sensing setup.

In order for the user to know which parts of the 3D scene needs more careful, closeup capture, we need a way to visually indicate the regions of the depth map with the highest uncertainty. However as discussed in the previous section, there are systematic sources of uncertainty in depth sensor measurements. Therefore if we simply use the variance in depth measurement for quality feedback, the resulting feedback as shown in Fig. 5(a) often is not very useful. Instead we use the following metric for each position x in a depth map:

$$Uncertainty(x) = \frac{var(x)}{\bar{x}^2 + \epsilon} \qquad (2)$$

Where $\bar{x} = mean(x)$. In our experiments we set $\epsilon = 0.001$. This offsets the systematic uncertainties and produces the quality map in Fig. 5(b). Discarding points beyond a manufacturer-rated maximum measurable distance and thresholding the quality map at the 70 percentile gives us the red marker positions, which corresponds to parts of the scene with high levels of uncertainty, as shown in Fig. 5(c). Fig. 6(b) shows a second, closeup capture of one of the parts of the scene with high uncertainty. The second capture was able to recover the finer-scale features.



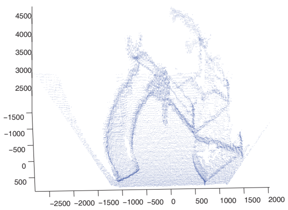
(a) (b) (c) (d)

Fig. 5. Depth quality measure feedback. Red circles indicate regions with (a) high variance (b,c) low quality.

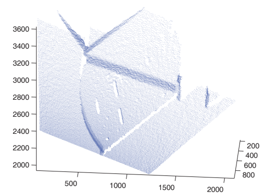
For visualizing the 3D points, we use a voxel representation [12]. Each voxel is colored according to the maximum uncertainty of 3D points the voxel contains. This allows the user to incrementally build the 3D model, guided by feedback, and registered with ICP in the cloud.

5. DISCUSSION AND FUTURE WORK

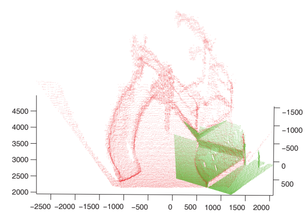
We have presented a experimental system for 3D reconstruction, guided by a quality feedback measure which takes into account the systematic uncertainties inherent in realtime depth sensors. This results in an intuitive, highly usable cloud-based system which allows users to efficiently build



(a) Initial capture



(b) Second capture



(c) Registered

Fig. 6. Feedback-guided 3D reconstruction.

large 3D models. In the near future, ICP-type applications can be run in real-time in the cloud providing low-latency quality feedback to the user. While we have described a number of sources of uncertainties that can be incorporated into the quality metric we have not exploited all of them. For example, some sensors provide a confidence measure based on signal strength and other sensors commonly available on mobile devices can also be used to improve the visual feedback. Emerging depth sensors such as [13] can also leverage a cloud-based system for computing the depth maps, and for these sensors the deeper algorithmic integration may allow even richer feedback to be provided to users.

6. REFERENCES

[1] Shahram Izadi, Richard Newcombe, David Kim, Otmar Hilliges, David Molyneaux, Steve Hodges, Pushmeet Kohli, Andrew Davison, and Andrew Fitzgibbon, "Kinectfusion: Real-time dynamic 3d surface reconstruction and interaction," in *SIGGRAPH Talks*. ACM, August 2011.

[2] Ha T. Nguyen and Minh N. Do, "Image-based rendering with depth information using the propagation algorithm," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, 18-23, 2005, vol. 2, pp. 589 – 592.

[3] Qingxiong Yang, Kar-Han Tan, Bruce Culbertson, and John Apostolopoulos, "Fusion of active and passive sensors for fast 3d capture," in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, October 2010, pp. 69 –74.

[4] "Photosynth," http://www.photosynth.net.

[5] "Primesense," http://www.primesense.com.

[6] Barak Freedman, Alexander Shpunt, Meir Machline, and Yoel Arieli, "Depth mapping using projected patterns," *US Patent Application*, , no. US20100118123A1, 2010.

[7] S. Burak Gokturk, Hakan Yalcin, and Cyrus Bamji, "A time-of-flight depth sensor - system description, issues and solutions," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. IEEE Conference on*, june 2004, p. 35.

[8] MESA Imaging AG, *Swiss Ranger SR4000 User Manual*, http://www.mesa-imaging.ch.

[9] "3dv zcam," http://en.wikipedia.org/wiki/ZCam.

[10] Paul J. Besl and Neil D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 239–256, 1992.

[11] Toru Tamaki, Miho Abe, Bisser Raytchev, and Kazufumi Kaneda, "Softassign and em-icp on gpu," in *Networking and Computing (ICNC), 2010 First International Conference on*, nov. 2010, pp. 179 –183.

[12] Anirban Ray Chaudhuri, A. Basu, Kar-Han Tan, S. Bhandari, and B. B. Chaudhuri, "An efficient set estimator in high dimensions: consistency and applications to fast data visualization," *Comput. Vis. Image Underst.*, vol. 93, pp. 260–287, March 2004.

[13] Ahmed Kirmani, Andrea Colao, Franco N. C. Wong, and Vivek K. Goyal, "Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor," *Optics Express*, vol. 19, no. 22, pp. 21485–21507, October 2011.