Week 1 Update

Rajat Saxena

February 6, 2015

0.1 Kinect Skeleton Tracking

0.1.1 Introduction

- In skeleton tracking, a human body is represented by 17 joints (Head, Neck, Torso, Left and Right Collar, L/R Shoulder, L/R Elbow, L/R Wrist, L/R Hip, L/R Knee and L/R Foot) along with the tracking confidence.
- Each joint is represented by its 3d coordinates.
- Kinect uses per pixel, body part recognition as an intermediate step to track skeleton. Evaluating each pixel separately avoids a combinatorial search over the different body joints.

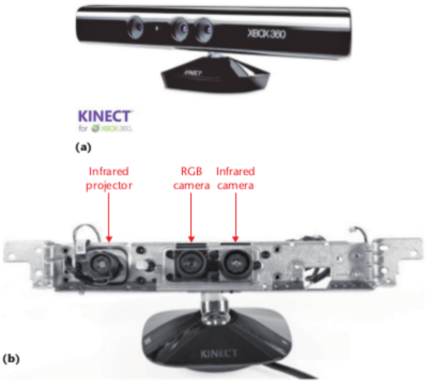

Figure 1. Microsoft Kinect sensor. (a) The Kinect sensor for Xbox 360. (b) The infrared (IR) projector, IR camera, and RGB camera inside a Kinect sensor.

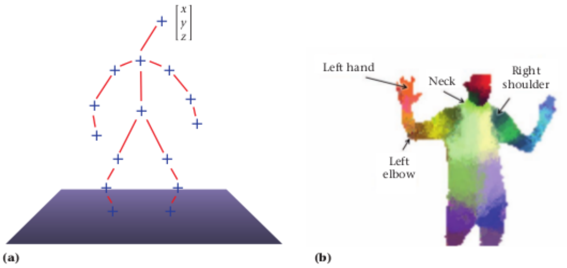

Figure 5. Skeletal tracking. (a) Using a skeletal representation of various body parts, (b) Kinect uses per-pixel, body-part recognition as an intermediate step to avoid a combinatorial search over the different body joints.

0.2 Microsoft SDK - Skeleton Tracker

0.2.1 Introduction

The Microsoft Kinect provides a convenient and inexpensive depth sensor and with the SDK, a skeleton tracker. This doc covers the evalaution of noise, accuracy, resolution and latency of the skeleton tracking software.

0.2.2 Range of Kinect

Experiments were conducted to find out how far and close user can be from the imaging sensor in order to be track skeleton. Below are the results:

- The angular range of the device is 57°x 43°(horizontal x vertical)
- This can be extended vertically by using the software controls for tilt motor, this has a range of 54°
- Microsoft recommends an optimal range of 1.2-3.5m from the sensor. But after testing on the system, results indicated that a skeleton could be acquired within a range of 0.85-4m from the camera.

0.2.3 Noise

Tracker noise causes the rendered image to jitter on screen. 1000 samples of the central position for a tracked skeleton standing 2.0m from the sensor were taken and mean position and standard deviation were calculated.
Results:

- 3D noise was found be 1.3m with a sd=0.75mm at 1.2m
- 3D noise at 3.5m was 6.9mm and sd=5.6mm
- Noise differed by dimension: x averaged 4.1mm, y 6.2mm, and z 8.1mm.

0.2.4 Accuracy

Focus on relative accuracy was more rather than absolute accuracy. To test, a straight wooden meter stick positioned 2m from the sensor was taken as reference which was running approximately along the x axis. A marker was affixed to a user's wrist to give a consistent position relative to the physical skeleton and the marker was placed along the meter stick. 25 samples were taken per point to reduce the effects of noise and measured distances between 100mm and 500mm.
Results are as follows:

- The average error in the tests was 5.6mm, with a sd=8.1mm
- To test scaling of accuracy with additional users in view, a long metal bar was used, 250mm segments, and a tape measure for reference. Error grew from 1.4mm with one user to 1.8mm with two users to 2.4mm with three users.
- No differences were found with respect to dimensions, including dpeth even after multiple tests.

0.2.5 Latency

Latency was measured using USB mouse relatively. Minimum latency from windows(system used to test) comes to be around 20ms. A test was conducted with two skeletons using a simple pendulum (a hand moving shoulder-to-shoulder).
Results are as follows:

- When the program was running at its normal frame rate of 30 Hz, the relative latency was found to be 106 ms on average, with a standard deviation of 23 ms and a maximum of 156 ms.
- When the program was running more slowly, the relative latency was found to average 202 ms, with a standard deviation of 26 ms and a maximum of 270 ms.
- With a single skeleton to track, the program generally maintained a 30 Hz update rate, but it would on occasion drop.
- With two or three users, the frame rate was generally between 18-20 Hz.
- with one skeleton, mean latency= 146ms (max=243ms)
- with two skeleton, mean latency = 234ms (max=386ms)
- with three skeleton, mean latency = 205ms (max=490ms)

0.2.6 Resolution

To establish resolution, error values less than the accuracy measurements with standard deviation less than the noise measurement were taken. This was done separately for depth(z) and one lateral dimension(x). Measurement were taken 2m from the sensor using the protocol of relative accuracy test, with distances of 1-5mm.
Results:

- Lateral resolution was found to be 3mm (0.086°), in agreement with PrimeSense, makers of the depth sensor.
- Also measured depth resolution, 2mm, was much better than the specification of 10mm because of interpolation of joints data.

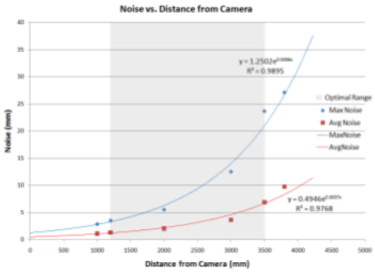

Figure 1: The mean and maximum noise as a function of distance from the sensor showed an exponential fit. The shaded region denotes the optimal depth range.

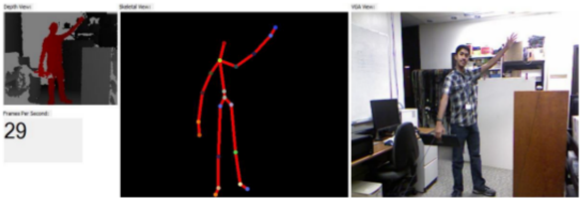

Figure 2: Three data streams provided by the Kinect for Windows SDK, from left: depth with user index, skeleton, and video.

0.2.7 Conclusion for above section

- The latency is the most problematic performance characteristic. Even the best condition produces a latency of 106 ms relatively or approximately 125 ms end-t- end latency. With multiple users or a single user close to the sensor, the number of pixels being processed for tracking of human forms increased, and latency increased correspondingly. A maximum latency of 500ms was observed which is large enough to be disturbing for a user.

- As with any most structured light system to recover depth, Kinect's performance degrade in difficult lighting conditions. Bright fluorescent lighting can increase the noise, but still allows tracking in dark rooms.