Estimating Anthropometry with Microsoft Kinect

M. Robinson & M. B. Parkinson

The Pennsylvania State University

May 10, 2013

Abstract

Anthropometric measurement data can be used to design a variety of devices and processes with which humans will interact; however, collecting these data is a very time-consuming process. In order to facilitate the rapid and accurate collection of anthropometric data, a novel system is under development which makes use of multiple Microsoft Kinect sensors to estimate body sizes. The Kinect is an appealing sensor for this application because it can be purchased for around 250 dollars, can be quickly set up in many different environments, and can easily be interfaced with a Windows computer. This paper will present the experimental setup and procedure used as well as some preliminary results.

Keywords: Anthropometry measurement; Microsoft Kinect

1 Introduction

In order to adequately design products with which people interact, accurate anthropometric measurements are needed; however, obtaining accurate measurements of people has traditionally been a very labor-intensive task. One of the biggest challenges is that the high variability in human sizes necessitates that large samples be measured, and measuring each participant can take a substantial amount of time. Additionally, some of the anthropometric landmarks, such as the acromion and the trochanterion, are difficult to palpate and measure repeatably, which can lead to measurement errors. Several past studies, such as ANSUR and CAESAR, have collected large amounts of anthropometric data. ANSUR represents the U.S. military at a particular point in time. CAESAR is a convenience sample in the early 2000s. Any data are associated with a particular sample at a particular time. The effects of secular trends and changing demographics make data obsolete. Additionally, many design populations are different than those for which we have data. All of these factors have led to the current interest in measuring anthropometry automatically from 3D images. This study investigated the fitness of the Microsoft Kinect sensor for collecting anthropometric data.

1.1 Anthropometric Measurements from Point Clouds

Given the promise of rapidly collecting anthropometric data, many other researchers have looked into obtaining anthropometry directly from 3D images, often known as point clouds. Prior studies have used machine learning and image processing techniques to estimate anthropometry (Ben Azouz et al., 2006), (Maier, 2011). These studies have used high-quality point clouds generated by expensive equipment in ideal environments. For example, the CAESAR data are widely used because they provide both high-fidelity point clouds sampled under highly controlled environments and corresponding anthropometric data obtained by physical measurement. This study, by comparison, aims to extract anthropometry with a low-cost sensor and with participants dressed in regular clothing.

1.2 Kinect Sensor

The Kinect sensor belongs to the class of devices known as depth cameras. Specific details about the measurement principles of the sensor itself can be found in Khoshelham (Khoshelham, 2011). The Kinect combines a standard RGB camera with a depth camera, which can simultaneously measure the distance to thousands of points in a scene. Although the Kinect's point cloud capability is impressive, its built-

in ability to estimate joint center locations has made it a revolutionary device.

Details on how the Kinect estimates joint centers can be found in Shotton et. al (Shotton et al., 2011). Compared to many traditional motion capture systems, the Kinect does not require that markers be placed on participants before data logging. This feature substantially reduces data collection time, but also reduces accuracy. Many groups have found that substantial errors can be present in the Kinect's pose estimation (Ye et al., 2011), (Clark et al., 2012). Ye et. al. presented a more accurate tracking algorithm for the Kinect; however, this increase in performance comes at the cost of execution speed. This paper will use the native Kinect motion tracking algorithm and will demonstrate a method for extracting anthropometry information from the Kinect's noisy and biased signal data.

1.3 Research Focus

This study aims to explore how low-cost and easy-to-use sensors, such as the Microsoft Kinect, could be used to collect anthropometric measurements. The primary focus of this research is on collecting data that are sufficiently accurate, and can be collected quickly from participants in standard clothing. This paper will focus on the experimental setup and will present some preliminary results for estimation of stature, acromion-radiale length and radiale-stylion length.

2 Experimental Setup

Anthropometric data for this experiment were collected using both physical measurements and four Kinect sensors. A list of the measures recorded can be found in Table 1. A total of 17 undergraduate students participated in this study. With the exception of height and mass, two researchers independently recorded each measurement and agreed on the final values. Measurements were taken using the CAESAR experimental procedures whenever procedures were available (Blackwell et al., 2002). All applicable measurements except for seated shoulder height were taken on the right side of the body. Measurements were compared to values predicted by proportionality constants based on stature, and any measurements that were more than three standard deviations off were resampled to maintain accuracy.

Four Kinect sensors were simultaneously used to record joint positions. Additionally each sensor captured two point clouds of the participant. An overhead
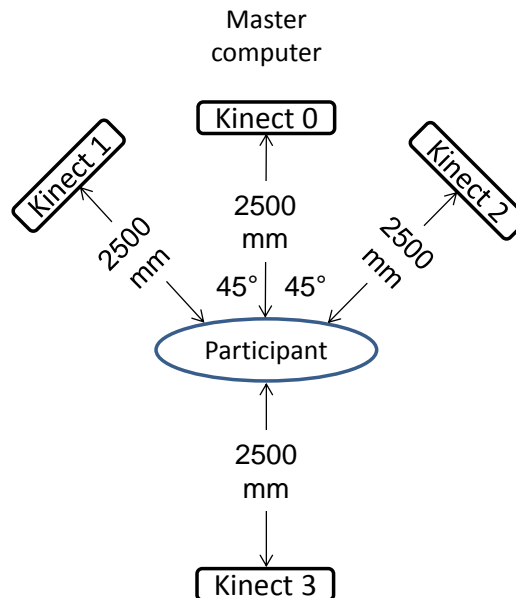

Figure 1: Overhead view of experiment

view of the experimental setup can be seen in 1. Joint locations were obtained using the built-in functionality of the Kinect.

Participants were asked to perform a number of tasks during the Kinect measurement portion of the experiment. The following list breaks the experiment into five stages intended to create a diverse set of conditions to observe the performance of the anthropometry estimation algorithm.

1. Walk slowly and complete two circles around a pair of markers on the floor
2. Place feet on two marks provided on the floor, face forward and swing arms slowly
3. Stand in a "T" and a "psi" pose (shown in Figure 2 A and B), touch head with both hands, touch shoulders, knees, and toes
4. Stand in pose C (shown in Figure 2) while a point cloud was captured
5. Sit in a backless chair in pose D (shown in Figure 2) while a point cloud was captured

Each Kinect sensor was paired with a dedicated computer that controlled the sensor and recorded data. It is possible to connect multiple Kinect sensors to one computer, but the decision to use one computer per sensor was motivated by a number of factors. First, a computer must have one USB 2.0 controller per
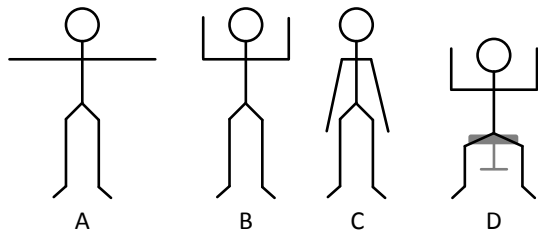
Figure 2: Poses used during Kinect data capture (A) "T" pose (B) "psi" pose (C) standing point cloud pose (D) seated point cloud pose

Kinect sensor, which is a limiting factor for many machines. Second, recording and displaying data coming from the Kinect can be a computationally intensive process and using multiple Kinects on one computer may cause the control program to slow down. Last, as of Kinect for Windows SDK version 1.5, the Kinect software does not allow for skeleton tracking using more than one Kinect at a time, which was a major component of this experiment. Running each sensor on a separate computer addresses all of these issues, but necessitates a method for synching up all of the data from the individual sensors.

Communication over a TCP/IP network was used to synchronize all of the computers used in this experiment. A local wired network was created exclusively for this experiment by connecting all four host computers to a network router. One computer was used to control the experiment and sent commands to the other three computers. Time synchronization was accomplished by having the control computer send the current program time to all of the other computers, which used that value as a time stamp on measurements. It is possible for one of the computers to take multiple measurements before receiving a new time stamp from the control computer, but all measurements should be within a few frames of each other.

3 Experimental Results

Summary statistics for the anthropometric measures obtained using traditional methods can be seen in Table 1. It is anticipated that methods similar to those described below could be used to estimate more of the dimensions listed, but this paper will focus on a subset of these measures. Specifically, stature will be examined as a measure that can be estimated from the Kinect point cloud as well as acromion-radiale length and radiale-stylion length, which show how segment lengths can be estimated from the raw Kinect joint data.

Table 1: Anthropometric measures and summary statistics obtained using traditional measurement methods

Measurement	Mean	St. Dev
Stature	1692.5	84.4801
Mass	70.7	16.1248
Trochanterion ht.	878.7	42.857
Sitting ht.	976.9	45.6685
Shoulder ht. sitting	685.8	31.9737
Shoulder breadth	430.5	41.3976
Outside arm breadth	464.1	47.7198
Hip breadth sitting	384.6	35.3607
Buttock popliteal lg.	489.8	28.2678
Acromion-radiale lg.	309.5	21.8321
Radiale stylion lg.	254.5	20.4148
Knee ht. sitting	515.2	31.7318
Max hip circ.	808.2	58.7848
Max hip circ. ht.	1012.5	79.3474
Chest circ.	940.3	95.8819
Head circ.	562.2	15.1308

3.1 Stature Estimation

Stature was estimated directly from the standing point clouds obtained in this experiment. Intuitively, stature could be estimated by finding the highest data point in the point cloud and subtracting the height of the lowest point; however, this method would be sensitive to outliers in the data caused by measurement errors.

Statistical data from the point cloud can be used to accurately estimate the height of the floor. If the Kinect's measurements were perfect, all of the points belonging to the floor would be at -1000 mm for this experiment. Instead the floor points are distributed around the true value. This can be observed in a histogram of the Z values recorded for a standing point cloud as shown in Figure 3. For this reason, the Z value with the highest number of measurements was selected as the floor level. While the floor height can easily be measured before taking measurements, uncertainties about the calibration of the Kinect made it preferable to estimate the floor height as described above.

In addition to Gaussian errors, the Kinect occasionally registers false measurements, an example of which can be seen on the top of the point cloud in Figure 4. Some of these errors occur because the Kinect occasionally has trouble measuring points on a user's hair. The interaction of multiple Kinect's can also cause errors, which will be described further in section 4. To reduce sensitivity to false measurements, the average value of the highest 100 Z points

was selected as the maximum height. This method was found to perform acceptably, but a more rigorous method would likely lead to reduced variability in results.

Results obtained using the methods described above to identify the floor location and the top of the participant's head can be seen in Table 2. Two outliers were removed from this data set: The Kinect failed to correctly obtain a point cloud for one participant and the other was likely mismeasured. In order to correct for bias errors in height estimation, an experimentally determined correction factor was applied to each sensor to force the average error to be zero.
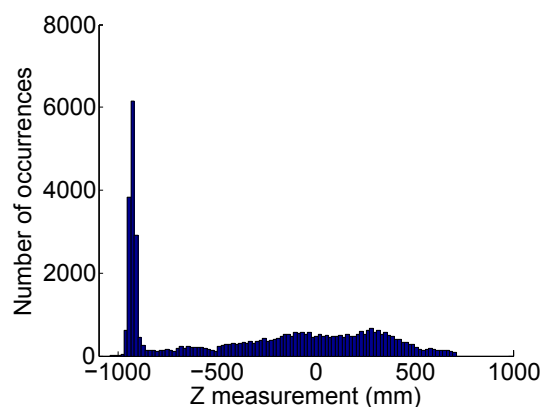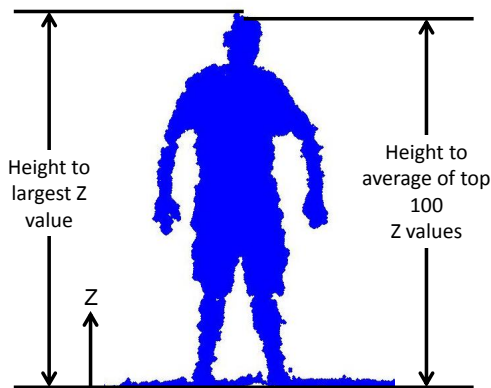

Figure 3: Histogram

Figure 4: Comparison of multiple methods for estimating height

3.2 Arm Segment Estimation

The Kinect's built-in joint tracking ability was used to estimate participant's right-side acromion-radiale length and radiale-stylion length. For each frame of joint data where the right arm was tracked, the shoulder-elbow and elbow-wrist distances were calculated. The raw joint center distances needed to be scaled to match anthropometric measurements. An experimental scale factor was found that minimized the error for this sample of data. For acromion-radiale length, a scale of 1.209 was used and a scale of 1.039 was used for radiale-stylion length. Table 3 shows the results obtained and compares those results to estimations based on proportionality constants obtained from Drillis and Contini (Drillis and Contini, 1966).

4 Discussion of Results

Stature estimation results were favorable with a maximum error for the data set without outliers of 1.3 percent and an average root-mean-squared error of 0.65 percent. The effect of averaging the height estimates from all four sensors was similar to averaging the front and back sensors. Additionally, averaging the front three sensors did little to improve the accuracy of the estimate. On the basis of these data, a sensor in front and behind a participant appears to be sufficient for height estimation. Adding extra sensors to the system not only increases cost and complexity, but can also deteriorate the quality of measurements.

Measurement errors caused by the interaction of multiple Kinect sensors degraded the quality of the point clouds obtained. These interactions are caused by the way the Kinect measures distance. In short, the Kinect projects a pattern of infrared dots and measures the displacement of that pattern with an infrared camera. If the patterns from two sensors overlap, the sensors may record erroneous measurements. This effect was found in prior studies, but it was not previ-

Table 2: Height estimation error results with outliers removed. Note that for cases where multiple sensors were used, the average height from all of the sensors was used to calculate error. 15 participants' data are included in this set.

Sensor(s)	Max. error (mm)	St. dev (mm)
0	35.6	18.3
1	49.8	29.2
2	42.3	19.1
3	70.9	26.4
0,1,2	41.3	18.3
0,3	25.6	10.4
0,1,2,3	23.0	11.6

Table 3: Maximum absolute error and standard deviation of error using the Kinect for estimation and proportionality constants based on participant's physically measured stature.

Measurement	Kinect estimation		Proportionality constant estimation	
	Max. error (mm)	St. dev (mm)	Max. error (mm)	St. dev (mm)
Acromion-radiale lg.	24.9	11.0	21.6	11.2
Radiale-stylion lg.	16.1	8.2	26.3	10.9

ously known how pronounced these effects would be when measuring humans (Berger et al., 2011). Interactions between the sensors produced artificial peaks in the point cloud data, as can be seen in Figure 5. It remains to be seen exactly how these errors will affect the accuracy of anthropometry estimated from these point clouds. One consequence of these errors is that methods of estimating anthropometry based on the extreme points, such as estimating stature by finding the highest point in the cloud, will be susceptible to large errors.
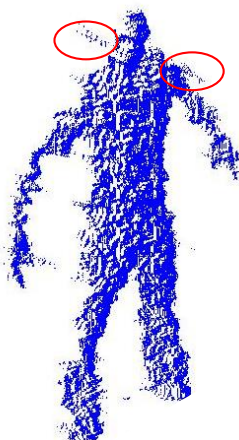

Figure 5: Measurement errors (circled in red) caused by the interaction of multiple Kinect sensors

Based on the data collected in this study, the method proposed above for estimating joint lengths was found to provide mixed results. The estimation for acromion-radiale length was not improved over using proportionality constants based on participant stature. This may partially be due to the fact that the acromion can be difficult to palpate, possibly leading to errors in the physical measurements used in this study. On the other hand, the radiale-stylion estimate shows a promising improvement when compared with proportionality constants. Maximum absolute error

was reduced by over 10 mm and the standard deviation of the error was reduced by over 2.5 mm. It should be noted that the sample size used for this experiment is too small to validate the accuracy of this system, but these results appear promising.

Errors in the Kinect's estimation of joint position were a substantial source of uncertainty in the segment length estimations. These errors were also found to be dependent on the pose of the segment being measured. Figure 6 shows how the instantaneous acromion-radiale length varied as Participant 9 moved through the experimental actions and poses as described in Section 2. As an example of how the anthropometry estimation changes with joint angle, there is a substantial difference between the estimate while the participant posed for the standing point cloud (Stage 4 in Figure 6) and the pose for the seated point cloud (Stage 5). It can also be seen that the estimation was reasonably stable when the participant was still, as was the case during stages 4 and 5, but changes rapidly while the participant moved around. This apparent pose dependence was not a major issue for this study as all of the participant performed the same actions; however, in an unstructured environment, this may cause more significant estimation errors.

It is also interesting to examine the effects of using multiple Kinect sensors for segment length estimation. Table 4 shows the radiale-stylion estimate error statistics for each individual sensor, as well as for combinations of sensors. Note that for each case presented, a scale factor that minimized estimation error was found, so these results can be considered a best-case scenario. It can be seen that the lowest maximum error and variance was recorded for Sensor 2. This result matches intuition, because Sensor 2 was facing the right side of the participant. Similar to the results found for height estimation, the result obtained by averaging all four sensors was not necessarily the most accurate. The front and back Kinect sensors were most useful for stature estimation, while the right camera was most useful for estimating segment lengths on the right side of the body. This result indicates that there may be value to using multiple Kinect
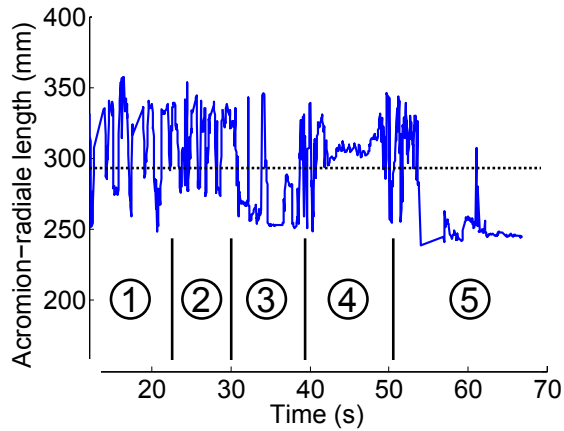
Figure 6: Example of acromion-radiale length data broken into five distinct stages as described in Section 2. The true acromion-radiale length is indicated by the dashed line.

sensors to collect data and utilizing different sensors based on the measurement being estimated.

Table 4: Error statistics for the Kinect estimation of right-side radiale-stylion length using individual sensors and combinations of multiple sensors.

Sensor(s) used	Max. error (mm)	St. dev (mm)
0	20.2	8.3
1	27.0	13.9
2	12.7	7.3
3	18.9	9.9
0,3	17.1	8.8
0,1,2,3	16.1	8.2

5 Conclusions and Further Areas of Investigation

The method presented in this paper was found to be relatively accurate at estimating stature, acromion-radiale length and radiale-stylion length. It was also found that using multiple Kinect sensors improved accuracy in certain cases, but was detrimental in other cases.

Moving forward, more data will be needed to verify the accuracy of this system. The results presented in this paper use the data collected to estimate biases in the Kinect. In order to reach more definitive conclusions, more data will need to be evaluated using the biases found in this study. Additionally, the methods presented here may be improved by more sophis-

ticated estimation techniques.

Acknowledgements

References

Ben Azouz, Z., Shu, C. and Mantel, A. (2006), 'Automatic locating of anthropometric landmarks on 3d human models'.

Berger, K., Ruhl, K., Brümmer, C., Schröder, Y., Scholz, A. and Magnor, M. (2011), Markerless motion capture using multiple color-depth sensors, in 'Proc. Vision, Modeling and Visualization (VMV)', Vol. 2011, p. 3.

Blackwell, S., Robinette, K., Daanen, H., Boehmer, M., Fleming, S., Kelly, S., Brill, T., Hoeferlin, D. and Burnsides, D. (2002), Civilian american and european surface anthropometry resource (caesar), final report, vol. ii, Technical report, AFRL-HE-WP-TR-2002-0173, US Air Force Research Laboratory, Human Effectiveness Directorate, Crew System Interface Division, Wright-Patterson AFB, OH.

Clark, R. A., Pua, Y.-H., Fortin, K., Ritchie, C., Webster, K. E., Denehy, L. and Bryant, A. L. (2012), 'Validity of the microsoft kinect for assessment of postural control', *Gait & Posture* .

Drillis, R. and Contini, R. (1966), *Body segment parameters*, New York University, School of Engineering and Science.

Khoshelham, K. (2011), Accuracy analysis of kinect depth data, in 'ISPRS workshop laser scanning', Vol. 38, p. 1.

Maier, M. J. (2011), Estimating anthropometric marker locations from 3-d ladar point clouds, Technical report, DTIC Document.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. and Blake, A. (2011), Real-time human pose recognition in parts from single depth images, in 'Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on', IEEE, pp. 1297–1304.

Ye, M., Wang, X., Yang, R., Ren, L. and Pollefeys, M. (2011), Accurate 3d pose estimation from a single depth image, *in* 'Computer Vision (ICCV), 2011 IEEE International Conference on', IEEE, pp. 731–738.