# Creating Custom Human Avatars for Ergonomic Analysis using Depth Cameras

Matthew P. Reed, Byoung-Keon Park, K. Han Kim
University of Michigan Transportation Research Institute

Ulrich Raschke
Siemens PLM

Ergonomic analysis of industrial tasks is often conducted through the use of human figure modeling software, such as the Jack software from Siemens. Typically, figures are scaled by inputting overall anthropometric dimensions, such as stature, body weight, and erect sitting height. This paper presents a method for rapidly generating a figure that matches the body dimensions and shape of an individual. A system with two Microsoft Kinect depth cameras is used to gather shape data. A statistical body shape model (SBSM) generated from analysis of over 250 men and women is used to fit the data. The output of the shape fitting, expressed as a set of 20 principal component scores, is input to the same male or female SBSM implemented in the Jack software. The result is a figure model that closely matches both the size and shape of the scanned individual. This methodology will be useful for a range of applications for which having a customized manikin is advantageous.

## INTRODUCTION

Recent advances in camera technology have revolutionized the gathering of 3D information from objects and scenes. One promising application is the capture of anthropometric data from human subjects for ergonomic analysis. In recent years, human figure model software has been widely used to assess capability and injury risk in industrial tasks. The software avatars are usually scaled to match an actual or prospective worker using standard anthropometric data, such as stature and body weight. In some applications, a photograph of a worker performing the task is used to obtain both posture and scaling information. For laboratory investigations, marker-based motion-capture systems are frequently used.

The Microsoft Kinect depth camera, originally developed for the Xbox gaming system, has been widely adopted for capturing motions in research settings, including in ergonomics applications (Anderson et al. 2012, Tong et al. 2012, Smisek et al. 2013). A growing body of research has addressed the accuracy of the anthropometric information in the kinematic linkage that is exported by the software.

The current effort is focused on capturing body shape to enable the rapid creation of manikins that match the body dimensions of a particular individual without the need for direct measurement. The new capability has the potential to improve the speed and accuracy of ergonomics research as well as product and workspace evaluations.

## METHODS

Figure 1 shows the hardware system. Kinect sensors are located on pedestals behind and in front of the standing subject. To overcome limitations in the sensor field of view, three images from each sensor are obtained, using the integrated motor to drive the sensor to 3 angles (horizontal and 25 degrees above and below horizontal) with respect to horizontal. A custom calibration of the sensors obtained by scanning an object with known geometry is applied to the sensor depth image (Reed and Park, 2013). In brief, a 500-mm-square planar object is moved through the scanning volume. At each sample location, deviations from a least-squares planar fit to the data are modeled using a radial-basis-function interpolator. The resulting deviation model is applied to all subsequent data. During scanning, data from the vertical poles in the scene are used to register the images. The overall scanning time is 12 seconds.

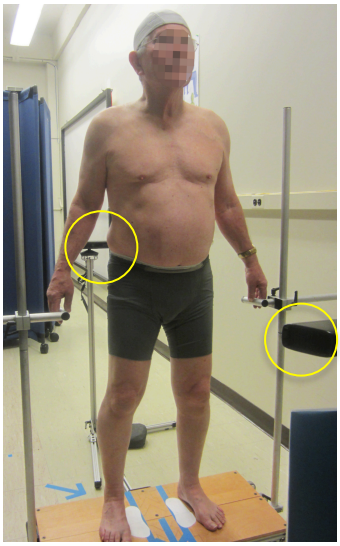

Figure 1. Hardware system with Kinect sensors to record the front and back of the subject. Circles show the sensor locations.

Figure 2 shows sample data from the sensor, along with data from the same subject obtained with a VITUS XXL (Human Solutions GmbH) high-resolution laser scanner (approximately 300k points). The data obtained from the sensors have holes along the sides and exhibit noise.

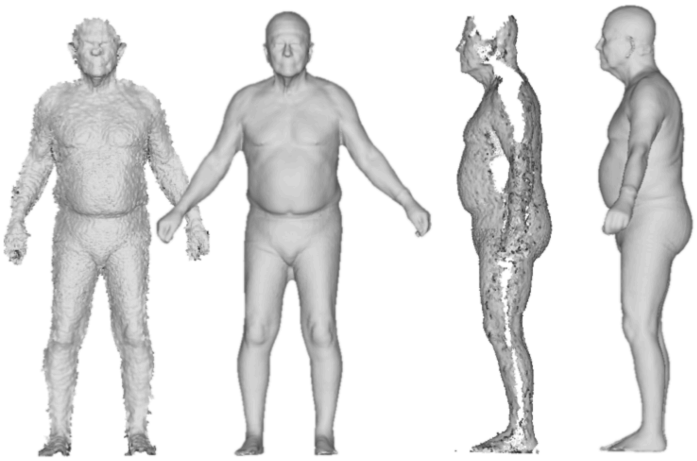

Figure 2. Sample data from the Kinect-based scanner compared with data from the same subject in the VITUS XXL laser scanner.

Separate male and female statistical body shape models (SBSM) were created using whole-body laser scan data from 311 men and 127 women (Reed et al. 2014). In brief, a standard template with 30k vertices was fit to scan data from a single standing posture for each subject. A principal component analysis (Joliffe, 2002) was conducted using the methods described in Reed and Parkinson (2008). Sixty principal components (PCs), accounting for over 99 percent of the variance in the mesh vertex locations, were retained. A least-squares regression analysis was conducted to predict PC scores from standard anthropometric variables, such as stature and body weight. The adult male and female models were implemented in the Jack human modeling software (Siemens PLM Software Inc.), enabling realistic figures to be generated across a wide range of body sizes (Figure 3).

Figure 3. Male and female figures in the Jack human modeling software system generated from the statistical body shape models based on laser-scan data.

The method for analyzing the Kinect data is similar in some respects to Weiss et al (2011), but the task is simplified by requiring a standardized posture that matches the posture used to generate the SBSM. The Kinect data are interpreted by finding the body shape, as defined by the gender-specific SBSM, that best matches the data. This process is posed as an optimization problem: find the set of PC scores that results in a minimum distance error between the data and 3000 points distributed across the surface of the body. For the current work, the first 20 principal components were used. By exploiting the structure of SBSM, this optimization converges in less than one second, yielding a vector of 20 numbers that describe the scanned individual's body size and shape.

This PC score vector is then input to the SBSM implemented in the Jack software. The Jack implementation (Reed et al. 2014) maps the resulting body shape onto the Jack kinematic linkage, merges realistic hands and feet onto the models, and integrates the Jack head and face. The entire process can be executed in less than one minute, yielding a fully functioning Jack manikin that represents the body size and shape of the scanned individual.

## RESULTS

To demonstrate this system, Jack manikins were created for four women scanned using the Kinect system. Figure 4 shows the Jack manikins in the default posture and in a range of alternative postures. The accuracy of the system was assessed by calculating the distances between the Kinect data, at SBSM fit to the data, and the resulting Jack manikin. Figure 5 shows the results of the comparison. In each panel, a "heat map" shows the distribution of distance discrepancies across each figure.

In all cases, the mean distance error was less than 2 mm in the combined torso, arm, and thigh areas. The 95$^{th}$ percentile error in these body regions were generally less than 15 mm. The largest discrepancies were observed in the lower abdomen and hip areas.

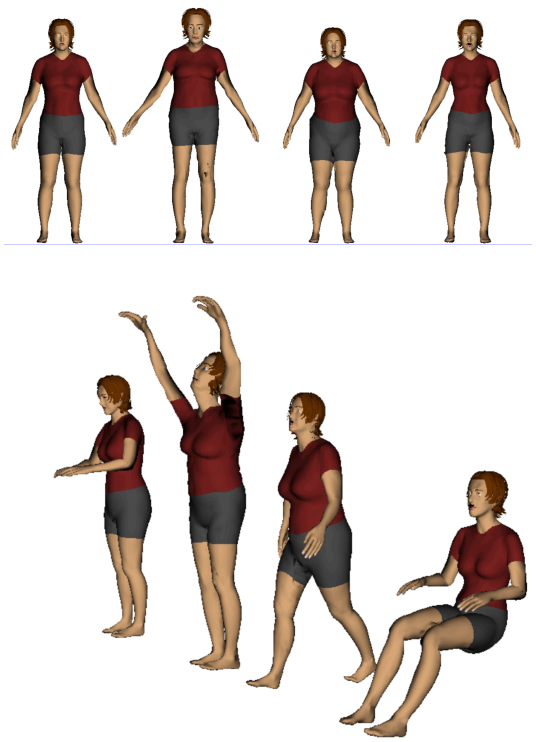

Figure 4. Examples of Jack manikins generated for 4 women from Kinect scans.

## DISCUSSION

This paper presents an integration of depth camera technology with a commercial ergonomics tool to generate custom manikins. The work is conceptually similar to Weiss et al. (2011) in that the limitations of the Kinect sensor are overcome through use of an SBSM. However, the current system provides strong advantages for ergonomics analysis through integration with a widely used human modeling software package.

Like all statistical models, the male and female SBSMs used in this work are limited by the scope of data on which they are based. Both datasets span more than 98% of the U.S. adult population with respect to stature, body weight, and other variables. The age range spans 18 to 90 years, with substantial numbers of adults over age 45. The PCA method enables generation of body shapes not present in the original database, but some body shapes cannot yet be modeled accurately. In general, those with substantial asymmetries or skeletal deformities are most problematic.

This preliminary evaluation shows good accuracy, with 95$^{th}$-percentile errors on the order of the magnitude of clothing effects. The accuracy of the Kinect system has been evaluated using geometric objects with known dimensions. Root-mean-square error for is typically less than 3 mm, with some variation across the scanned volume. Comparisons of scans of the same individual measured with the Kinect system and the VITUS XXL laser scanner reveal discrepancies typically less than 5 mm, with most differences due to posture. One approach to evaluating the overall system would be to compare standard anthropometric dimensions measured on individuals to the approximations of the same dimensions measured on the resulting Jack manikins. However, this approach will be confounded by the fact that the measurement postures for standard anthropometric dimensions differ from the scanned posture.

The largest discrepancies are likely to occur in limb segment lengths, because the scanned posture does not provide strong differentiation for elbow and knee locations. Consequently, the fitting algorithms are likely to produce limb segment proportions closer to the mean values than is accurate for those with unusual limb proportions.

A significant limitation of the current system is that the SBSM is based on scans of minimally clad individuals. This ensures that the data provide the best available information on body shape, but means that the results will be less accurate when fitting scans of clothed individuals. Work is underway to create a weighting and scaling function that emphasizes model fit in areas where clothing is likely to fit closely (e.g., shoulders) and de-emphasizes areas where loose clothing is expected (e.g., under the arms).

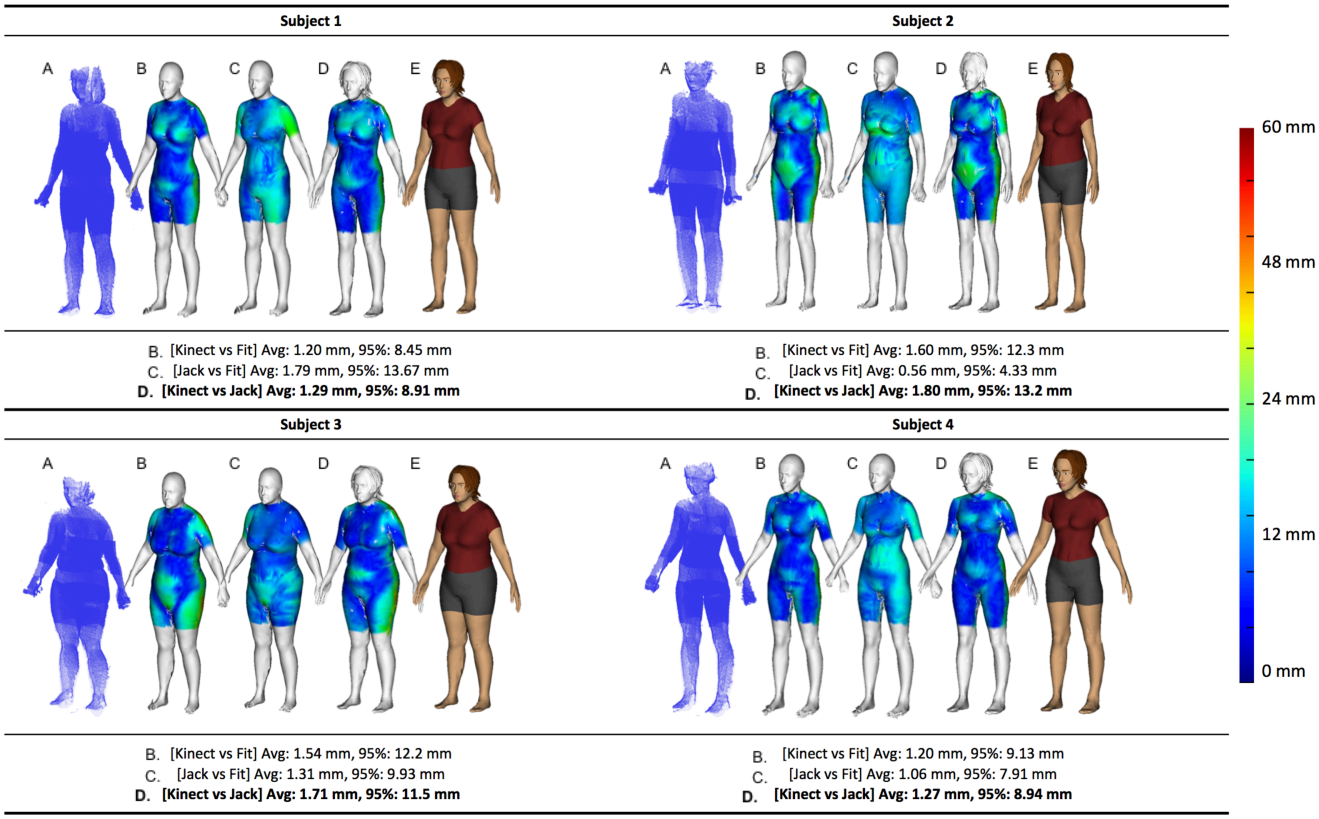


Figure 5. Fitting error evaluation. In each panel, the Kinect data are shown at the left (A) and the Jack figure at the right (E). The second figure (B) compares the Kinect data with the SBSM fit. The center figure (C) compares the Jack manikin with the fit, and the fourth figure (D) compares the Kinect data with the Jack figure.

Rapid improvements in these methods are expected. The Kinect sensors scan at 30 frames per second, suggesting that a much more rapid capture is feasible with a larger field of view. The current system is relatively slow due to a decision to maximize accuracy by bringing the sensor closer to the subject, which necessitates imaging from 3 angles. Moving the sensor to each angle with the built-in motor takes several seconds. An alternative would be to use a moving sensor, either on a mechanism or hand-held. For industrial use, a system even more portable than the current system, which takes five minutes to set up, would be desirable. Higher-accuracy, higher-resolution cameras can be expected. The system can use 3D human surface data from any source (laser, structured light, texture-based, etc.), so other options will become available.

The SBSMs will also be improved through the incorporation of additional data. Hundreds of additional scans are available and will be incorporated into the next generation of the model. We will also explore the creation of models that are specific to particular populations differentiated on the basis of national or regional origin. For example, a model based on Chinese population data might provide better fitting for applications in that country.

This approach can be readily applied to other human figure model software. A critical step is the regeneration of the SBSM using a template mesh specific to the target software. A robust implementation also requires smooth blending of the body shape predictions with model-specific hands, feet, and head.

This tool will be most useful for performing direct evaluations of workers performing tasks. Currently, many Jack analyses are based on photographs or video of workers performing a task. The analyst attempts to scale the figure to match the worker's body dimensions, then manually manipulates the figure to approximate the working posture. Biomechanical analyses are then performed using the Jack manikin to assess, for example, low-back loading.

The work presented here is a first step toward automating this procedure, enabling rapid generation of an accurate, subject-specific manikin. In future work, we will use Kinect data along with the subject-specific manikin to rapidly and accurately capture working postures and motions, enabling faster and accurate ergonomic assessments.

## ACKNOWLEDGMENTS

This work was supported in part by the Automotive Research Center, a U.S. Army Center of Excellence for Modeling and Simulation of Ground Vehicles led by the University of Michigan. Additional support was received from the MCubed program at the University of Michigan. The authors acknowledge the important contributions of Rishi Tirumali of Siemens PLM.

## REFERENCES

Andersen, M. R., Jensen, T., Lisouski, P., Mortensen, A. K., Hansen, M. K., Gregersen, T., & Ahrendt, P. (2012). Kinect depth sensor evaluation for computer vision applications. Århus Universitet.

Jolliffe, I. (2002). Principal Component Analysis, 2 ed. Springer.

Reed, M.P. and Parkinson, M.B. (2008). Modeling Variability in Torso Shape for Chair and Seat Design. Proceedings of the ASME Design Engineering Technical Conferences.

Reed, M.P. (2012). A pilot study of three-dimensional child anthropometry for vehicle safety analysis. Proceedings of the 2012 Human Factors and Ergonomics Society Annual Meeting. HFES, Santa Monica, CA.

Reed, M.P., and Park, B-K (2014). Rapid generation of custom avatars using depth cameras. *Proc. 3rd International Digital Human Modeling Conference*, Tokyo, Japan.

Reed, M.P., Raschke, U., Tirumali, R., and Parkinson, M. B. (2014). Developing and implementing parametric human body models in ergonomics software. *Proc. 3rd International Digital Human Modeling Conference*, Tokyo, Japan.

Smisek, J., Jancosek, M., & Pajdla, T. (2013). 3D with Kinect. In *Consumer Depth Cameras for Computer Vision* (pp. 3-25). Springer London.

Tong, J., Zhou, J., Liu, L., Pan, Z., & Yan, H. (2012). Scanning 3D full human bodies using Kinects. *IEEE Transactions on Visualization and Computer Graphics*, 18(4), 643-650.

Weiss, A., Hirshberg, D., & Black, M. J. (2011). Home 3D body scans from noisy image and range data. *In IEEE International Conference on Computer Vision (ICCV)*, 2011 (pp. 1951-1958). IEEE.