

A Model-based Approach to Rapid Estimation of Body Shape and Postures Using Low-Cost Depth Cameras

Byoung-Keon D. PARK*, Matthew P. REED

University of Michigan, Transportation Research Institute, Ann Arbor MI, USA

DOI: 10.15221/17.281 <http://dx.doi.org/10.15221/17.281>

Abstract

Depth cameras have revolutionized anthropometry providing efficient ways to gather 3-dimensional information of a human body in varied settings. However, the accuracy and resolution of the current depth camera systems limits their applicability for many applications. Commercial markerless motion-capture systems based on depth camera technology are similarly limited, particularly by the lack of a model-based tracking mechanism and a high dependence on unobstructed camera views. This paper presents a series of model-based methods for estimating body configuration and postures based on depth and posture data obtained from a single Kinect v2 sensor. The software system records and processes multiple depth images of a person from different point of views to capture the whole-body shape. A statistical body shape model that can represent a wide variety of human body shapes and poses was generated by analyzing template-fit whole-body laser scans and measured anatomical landmark data using a principal component analysis (PCA). PCA reduces the high dimensionality of the original data source by projecting the dataset to a low dimensional principal component (PC) space. In the PC space, only realistic body shapes and landmark data can be generated, and this space allows for efficient body shape search due to the low dimensionality. Using this model, a rapid fitting method for generating a subject-specific manikin from Kinect depth data was developed that can estimate a minimally-clad body shape under normally clothing. Posture data from subsequent movements estimated by the built-in skeleton tracker in the Kinect system were further improved by fitting each body segment of the manikin to a corresponding partial depth dataset while the segment lengths were preserved as defined in the manikin. This study demonstrated how a model-based approach can improve the accuracy and feasibility of a depth camera system so that the system can be readily applied for various applications, including in-vehicle occupant dynamic analysis, occupant classification, and markerless motion analysis.

Keywords: Depth Camera, Kinect, Model-based, Statistical Body Shape Model, Markerless Motion Capture, Clothing

1. Introduction

Recently, depth cameras that return 3D depth information of surroundings have been extensively used in a wide range of applications due to its low cost, low profile, and portability. The advance of the related technologies has been improved the accuracy and reliability of these cameras. The Microsoft Kinect system, for example, switched its depth sensing technique from a structured light method to a time-of-flight, which is less sensitive to interference with other sensors, for the latest version (V2) to provide color and depth information with a higher resolution and a wider field of view [1]. Most importantly, many validation studies reported that, while the accuracy decreases exponentially with increasing distance for Kinect v1, Kinect v2 has a constant accuracy about 10 mm over distance [2].

Due to the improved depth sensing capability, Kinect software development kit (SDK) provides more reliable face tracking and skeleton (i.e. posture) estimation up to 6 people, simultaneously. This benefits many applications where portable markerless human posture tracking systems are required. Since the Kinect system requires only a single sensor and a compatible computer to track the postures, motion data can be gathered in a wide range of environments, such as workplaces [3], schools [4], even moving vehicles [5]. However, the accuracy of the original depth data what the current depth camera systems provide limits their applicability for engineering applications.

In recent studies, model-based approaches have been a solution for these limits of the depth data quality, especially in body shape and posture modeling area. For example, Weiss et al. [6] presented a model-based body shape and posture estimation method from multiple depth images based on SCAPE model [7]. Pishchulin et al. [8] proposed S-SCAPE that is a simplified and efficient variant of the SCAPE model to model a subject-specific body shape from depth images. Baak et al. [9] and Ye et al. [10] utilized a pose database that has motion capture data to improve posture estimation based on depth images.

In this paper, we present a series of model-based methods for rapidly estimating accurate body shape and posture of a person using a single Kinect v2 sensor. While the Kinect SDK provides a posture estimating functionality via a built-in skeleton tracker, the accuracy was further improved by utilizing a subject-specific manikin that generated from a body shape model. We applied the Kinect-estimated posture data as an initial guess to the articulated manikin per frame, and each segment of the manikin was fitted to corresponding depth data to capture the details in actual postures. Clothing effects were learned from the manikin generation step using the inscribed-fitting method [11] and applied in the posture fitting step.

2. Method

2.1. Statistical Body Shape Model

A new human figure framework, BioHuman, developed at UMTRI (<http://umtri.umich.edu>) was used in this study. This framework incorporates an efficient, quad-based, symmetrical surface mesh along with anatomically defined surface landmarks and internal joint centers. BioHuman body shape model currently includes 14k vertices, 96 landmarks including joint centers, and 16 standard anthropometric dimensions. Figure 1 shows the mesh structure of BioHuman and the surface landmarks on the mean shape.

A total of 155 male and female subjects were recruited, with stature range 145 to 196 cm, BMI 16 to 42 kg/m², and age range of 20 to 95 years old. Whole body laser scan data of a standard standing pose were homologized by fitting BioHuman model to each scan as a template using a two-level fitting method developed in our previous study [12]. The fitted scans were analyzed using principal component analysis (PCA) along with the 6 standard anthropometric measures, and 155 body landmarks and joint locations. 60 principal component (PC) scores were retained from the PCA to describe the 99%ile of the data variations. A multivariate regression model was built to associate the PC scores to a set of predictors such as stature, sitting height to stature ratio, BMI, gender and age, so that a subject-specific manikin can be generated from these few predictors. The resulting statistical body shape model is publicly available at <http://humanshape.org>.

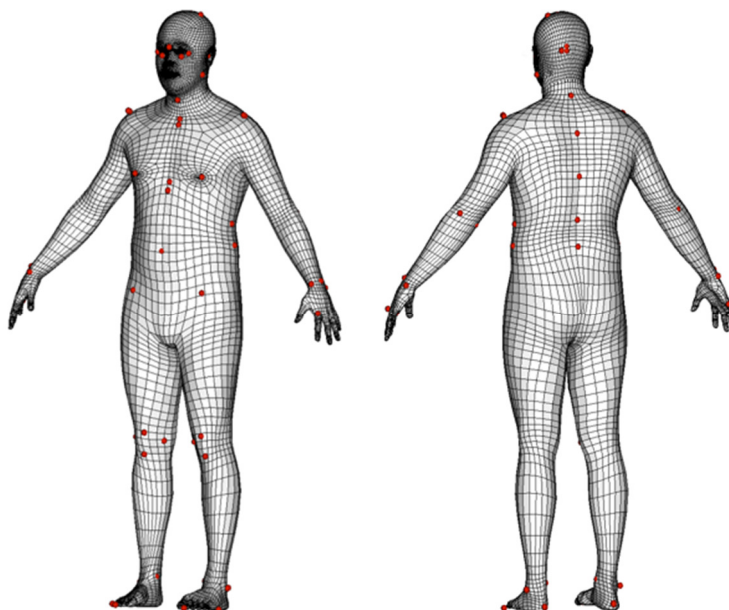


Figure 1. Mean shape of the BioHuman standing male model and surface landmark locations.

2.2. Subject-specific Manikin Generation

The body shape space of the BioHuman model is a low dimensional space defined by a small number of PC scores, which allows an efficient body shape search in that space. Also, any combinations of the PC scores found within this body shape space generates only a realistic body shape. Based on these characteristics, a rapid subject-specific body shape fitting method was developed [13]. This method generates a water-tight subject-specific body shape manikin even from a given incomplete information of a target subject. For example, if a scan of a partial body shape is given, a whole body shape can be estimated by finding a set of PC scores of the body shape model that meet the given input data using this fitting method.

In the current study, a set of depth images obtained from a single Kinect V2 sensor were used for generating a subject-specific manikin using the fitting method. The software system takes shots of the front, side, and back side of a person in about 10 seconds. The depth data are then post-processed to improve the data accuracy and quality. Multiple images per shot are averaged to reduce the noise level, and background points were removed from the images based on the skeleton data estimated by the SDK. The system automatically aligns and merges the images based on detected geometrical correspondences between the processed images. Figure 2 shows the hardware setup and the graphic user interface of the developed scanning system.

The inscribed fitting method developed in a previous study [11] was employed to estimate the minimally clad body shape from clothed scan data. This fitting method was motivated by the fact that the body shape should lie inside the surface of a clothed body, and at the same time, some parts like face, ankles, and hands of the scan should be close to the actual body shape. This yielded an assumption that an actual body shape is the maximum body volume within the scan surface that is available in the body shape space of the model. Conceptually, this is like finding an inscribed sphere in a polygon. Figure 3 shows an example of clothed scan data captured from a single Kinect sensor, and an inscribed-fitted BioHuman manikin of the scan. In this case, 10 iterations were performed for fitting the model and the known stature of the subject was fixed during the fitting.

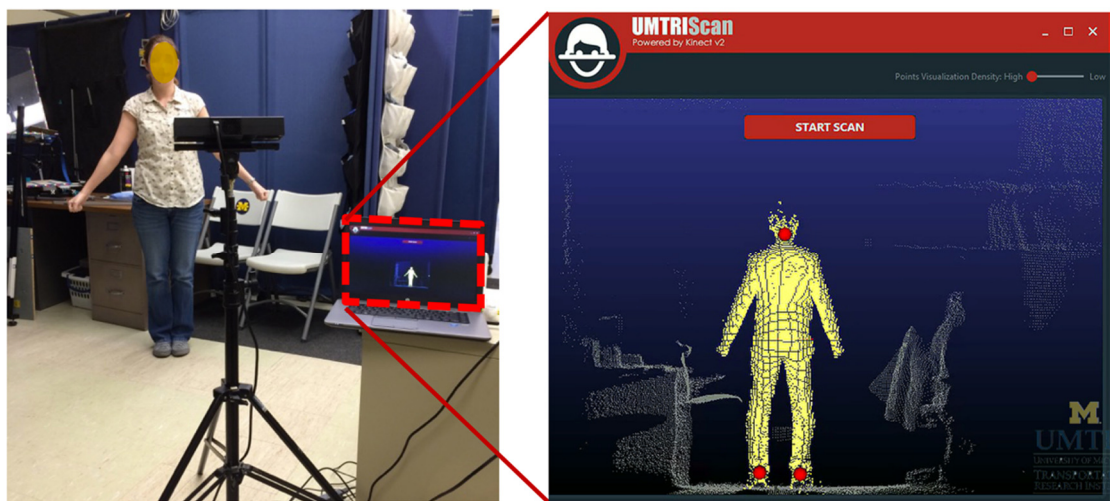


Figure 2. A portable whole body scanning system: Hardware setup (left) and software (right).

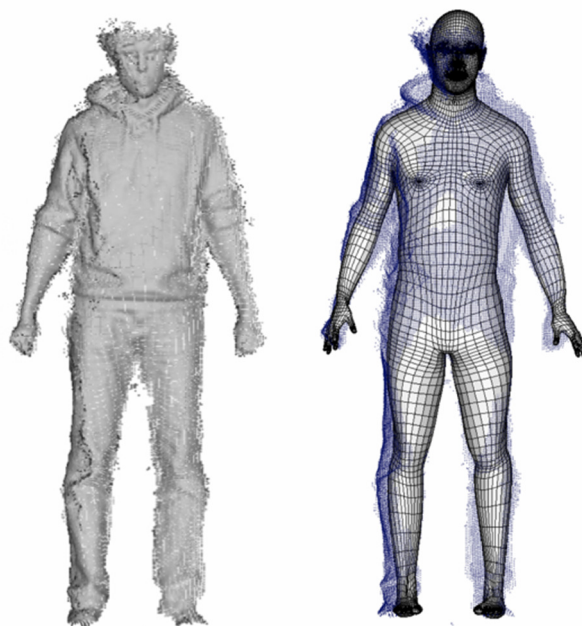


Figure 3. An example of input Kinect clothed scan (left) and an inscribed-fitted BioHuman manikin (right)

2.3. Model-based Posture Estimation

For the current study, BioHuman was articulated based on estimated joint locations to represent various postures of a specific person. For the articulation, the model was segmented into 17 body parts. An arbitrary posture can be applied to a BioHuman manikin. To improve computational efficiency, blending functions are not applied at joints in this study.

The accuracy of the original joint locations estimated from the SDK is improved by fitting a subject-specific manikin to depth data frame by frame. This fitting step is performed in post-processing, so the Kinect depth data and segment coordinate systems are stored prior to these calculations. The overall approach to fit a model to depth data consists of three steps: (1) segment the depth per the body segments of the manikin, (2) transform each manikin segment to the corresponding segment using an articulated iterative closest point (ICP) algorithm, and (3) update the segment coordinate systems based on the estimated transformations. As described above, Kinect joint orientation data were applied to the manikin for each frame as an initial guess.

When each segment is fitted to the depth data using ICP, the joint connectivity between the sharing segments is considered as a boundary condition. The joint connectivity is considered in the ICP process by including the joint location in the points to be aligned and applying a weight (between 10 and 100) to ensure that the joint constraint is met. Figure 3 shows sample fitting results with the joint constraints. Figure 4 shows an example of a posture fit with and without joint connectivity consideration.

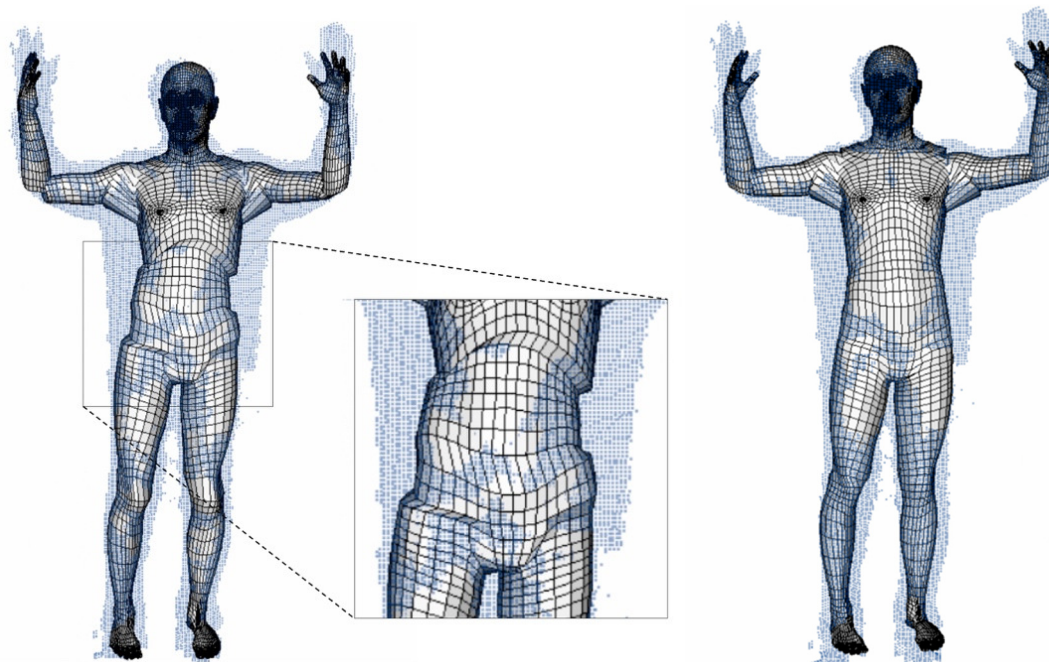


Figure 4. Model fitting without joint constraints (top) and with joint constraints (bottom, weight per joint = 10)

3. Results

The fitted body shape manikins estimated from clothed depth data were compared to the minimally-clad laser scans of the same subjects. The comparison showed that the absolute mean error was 11.1 mm, 95th %tile error was 17.1 mm, and the root-mean-square-error (RMSE) was 11.7 mm. In Figure 5, the mean errors at each vertex across all the scans were color-coded on a mean body shape. The fitting time was under two seconds per scan on average on a typical laptop computer (I7 3.4GHz CPU with 16 GB RAM).

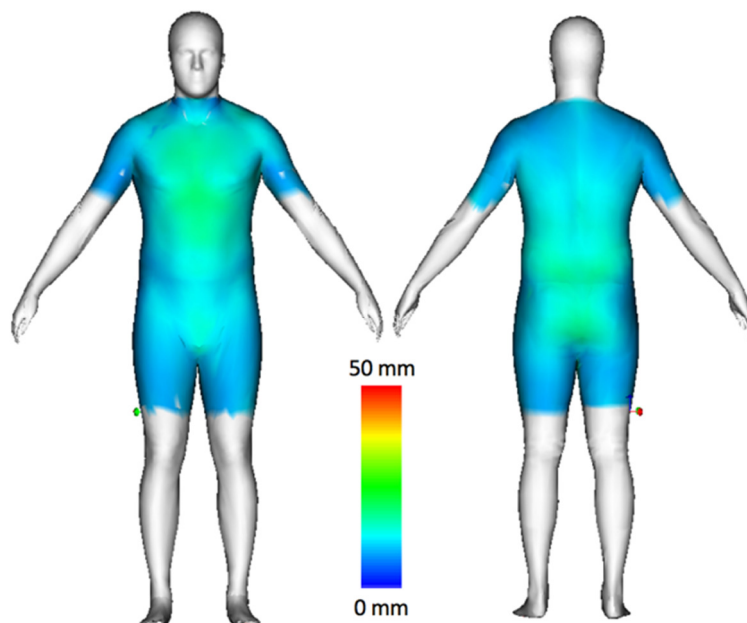


Figure 5: Mean error distribution was color-coded on a mean body shape.

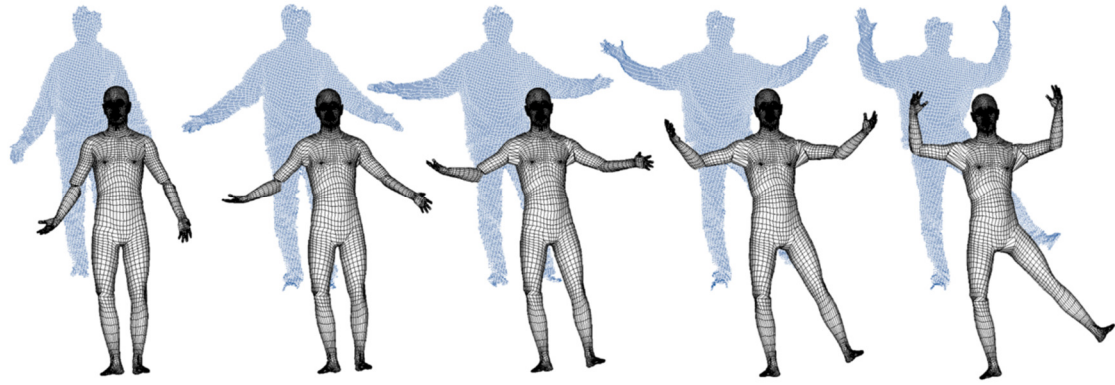
Figure 6 shows comparisons of the series of depth data and the posture estimating results using the fitted subject-specific manikin. Within two to four iterations, all the segments were well fitted to the depth data without resulting in unrealistic postures. The computation time of a single fitting iteration, including the depth segmentation time, varied around 100 milliseconds (10 Hz) when 1/10 of depth points were used. Torso flexions and extensions were well estimated from the fitting, improving on the original Kinect skeleton data, which do not represent any degrees of freedom within the torso. Note that these motions were taken without any obstacles between the subject and the sensor, and wrist and ankles were not fit to depth data due to poor Kinect data quality for these joints. The results showed that the system can reliably track motions even a subject is moving fast (Figure 6 (c)) or some parts of the body are self-occluded by the other parts (Figure 6. (b)).

4. Discussion

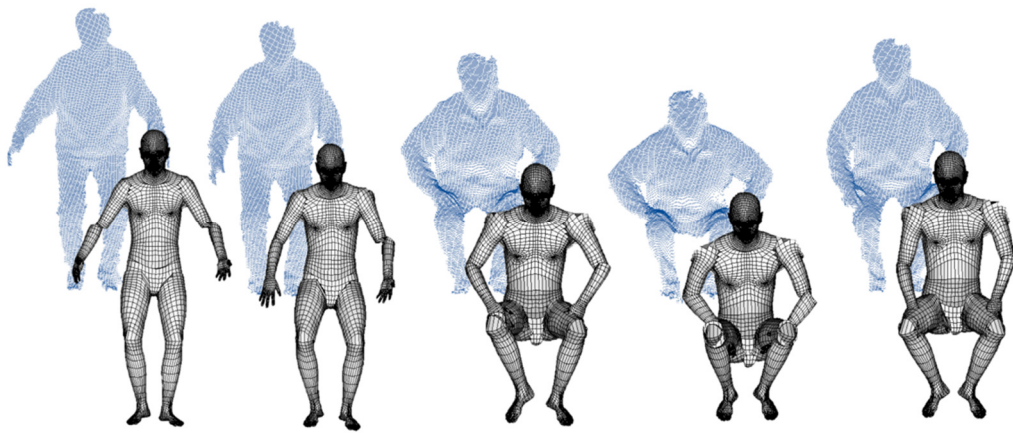
This study investigated the utility of Kinect V2 for subject-specific manikin generation and markerless posture estimation. A preliminary evaluation demonstrated that although the version-2 sensor is considerably improved relative to the first version, the posture tracking accuracy of software supplied with the sensor is not adequate for research purposes. To address these limitations, a custom software system was developed that leveraged previously developed statistical models and methods.

The system applied various novel developments. A new articulated body shape model, BioHuman, was utilized. The model incorporates 17 articulated body segments based on estimated joint locations in order to represent various postures. The inscribed fitting method finds a body shape from a normally clothed scan data, so that the system can be applied to broader situations. The original Kinect skeleton data representing postures of a person were successfully applied to subject-specific BioHuman manikins. For improving the accuracy of this initial guess, we developed a model-based posture tracking method based on Kinect depth data. The results demonstrated that this approach has good potential to track postures in daily living situations.

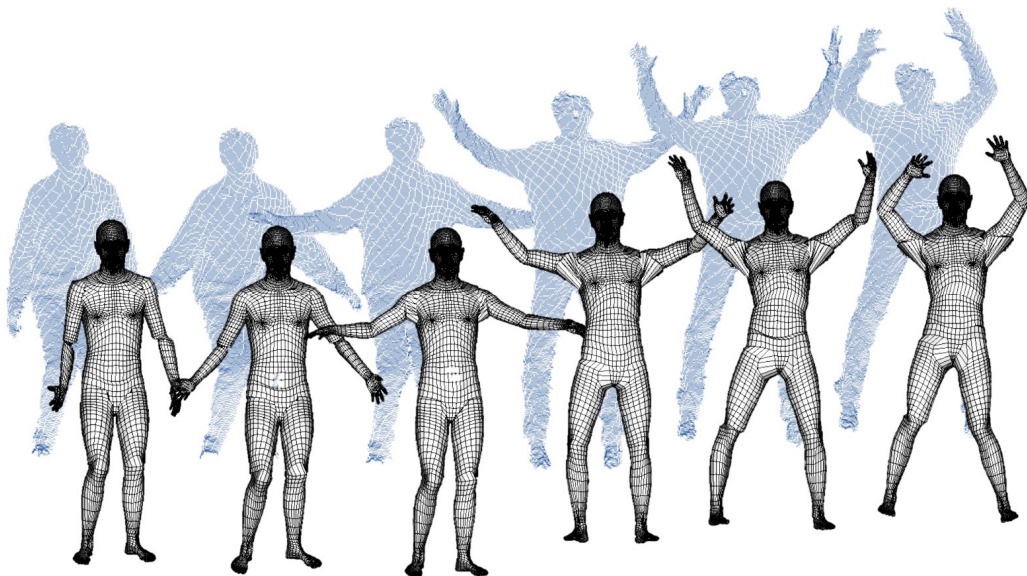
This study has substantial limitations that will be addressed in subsequent work. Many of the limitations are due to the Kinect sensor itself. As described in this report, the built-in skeleton tracking is of limited utility for posture tracking, due to large errors and generally poor performance, including substantial dropouts. Since the model-based approach developed in this study uses the original Kinect skeleton data as initial guess, fitting results might be degraded if the original data has poor quality. For example, if a lower leg is hidden to the sensor for a while, the skeleton tracker cannot estimate joints locations related to the leg, and the new system also cannot estimate the good quality of leg trajectories. This could be resolved by using a statistical posture model that predicts series of joint locations from incomplete skeleton data.



(a) Raising arms and a leg



(b) Sitting and standing



(c) Jumping

Figure 6. Tracking results of short motions by fitting a BioHuman manikin (white) to Kinect depth data (blue)

Although the results of this study are promising, more work will be needed to develop a robust system that can be applied to vehicle occupant positioning research or other applications. Quantitative validation of the developed system is needed. From simple motions to complex motions interacting with surrounding objects, various motion data are needed to collect for in-depth validation of the system. Also, in-depth clothing effect analysis is needed not only for estimating the body shape, but for the posture tracking. Although the new system can estimate a plausible body shape and a kinematic linkage from a clothed scan, the resulting posture data may not be accurate as posture data obtained from a minimally clad person. More work is needed to determine the likely bounds of clothing effects and how they can be mitigated.

References

- [1] Yang, L., Zhang, L., Dong, H., Alelaiwi, A., & El Saddik, A. (2015). Evaluating and improving the depth accuracy of Kinect for Windows v2. *IEEE Sensors Journal*, 15(8), 4275-4285.
- [2] Wasenmüller, O., & Stricker, D. (2016, November). Comparison of kinect v1 and v2 depth images in terms of accuracy and precision. In *Asian Conference on Computer Vision* (pp. 34-45). Springer, Cham.
- [3] Dutta, T. (2012). Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics*, 43(4), 645-649.
- [4] Hsu, H. M. J. (2011). The potential of Kinect in education. *International Journal of Information and Education Technology*, 1(5), 365-370.
- [5] Rahman, A. S. M., Saboune, J., & El Saddik, A. (2011). Motion-path based in car gesture control of the multimedia devices. In *Proceedings of the first ACM international symposium on Design and analysis of intelligent vehicular networks and applications* (pp. 69-76). ACM.
- [6] Weiss, A., Hirshberg, D., & Black, M. J. (2011, November). Home 3D body scans from noisy image and range data. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 1951-1958). IEEE.
- [7] Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., & Davis, J. (2005, July). SCAPE: shape completion and animation of people. In *ACM Transactions on Graphics (TOG)* (Vol. 24, No. 3, pp. 408-416). ACM.
- [8] Pishchulin, L., Wuhrer, S., Helten, T., Theobalt, C., & Schiele, B. (2017). Building statistical shape spaces for 3D human modeling. *Pattern Recognition*, 67, 276-286.
- [9] Baak, A., Müller, M., Bharaj, G., Seidel, H. P., & Theobalt, C. (2013). A data-driven approach for real-time full body pose reconstruction from a depth camera. In *Consumer Depth Cameras for Computer Vision* (pp. 71-98). Springer London.
- [10] Ye, M., Wang, X., Yang, R., Ren, L., & Pollefeys, M. (2011, November). Accurate 3d pose estimation from a single depth image. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 731-738). IEEE.
- [11] Park, B-K., Corner, B.D., Kearney, M., and Reed, M.P. (2016). Estimating human body characteristics under clothing using a statistical body shape model. *Proc. 4th International Digital Human Modeling Conference*. Montreal, Canada.
- [12] Park, B. K., & Reed, M. P. (2015). Parametric body shape model of standing children aged 3–11 years. *Ergonomics*, 58(10), 1714-1725. DOI: 10.1080/00140139.2015.1033480
- [13] Park, B. K., Lumeng, J. C., Lumeng, C. N., Ebert, S. M., & Reed, M. P. (2015). Child body shape measurement using depth cameras and a statistical body shape model. *Ergonomics*, 58(2), 301-309.