Validity and Repeatability of the Sizestream 3D Scanner and Poikos Modeling System

Teddie E. VONK*[1], Hein A.M. DAANEN[1,2,3]

[1] Create-it applied research & AMFI, University of applied sciences Amsterdam, The Netherlands;
[2] Faculty of Behaviour and Movement Sciences, VU University Amsterdam, The Netherlands;
[3] TNO, Soesterberg, The Netherlands

Abstract

Three-dimensional (3D) body scanning becomes increasingly important in the medical, ergonomical and apparel industry. The SizeStream 3D body scanner is a 3D body scanner in the shape of a fitting room that can generate a 3D copy of the human body in a few seconds. The Poikos modeling system generates a 3D image of a person using a front- and side photo. This study evaluates the repeatability and validity of both systems with human subjects.

Hundred fifty-six participants were included in this study, of whom 85 were scanned twice by the SizeStream Scanner and 139 by the Poikos modeling system. The repeatability is assessed by calculating the intra-class correlation coefficients (ICC) and standard error of measurement (SEM), and the validity of 6 Sizestream and 4 Poikos measurements is evaluated by comparing these measurements with collected tape measurements.

The ICC and the SEM results indicate that 79 of the 163 SizeStream measurements are repeatable enough to use for fashion purposes, since they had an ICC above 0.80 and a SEM below 10mm. Fifty-one measurements give a good indication but are not accurate enough for pattern making. The waist, chest and hip circumferences are valid after a correction of the over- or underestimation of the measurements. The Poikos modeling system is a promising, but is as expected, less repeatable and valid than the SizeStream scanner. Although the Poikos modeling system can give a good estimation of the body shape, the measurements are not accurate enough (SEM > 10mm) to use in the fashion industry. Future studies have to be performed to validate more Poikos and SizeStream measurements and to assess the usability of these measurements for the fashion industry.

1. Introduction

Three-dimensional (3D) body scanning becomes an increasingly important tool in the medical, ergonomical and apparel industry to determine human body dimensions. The first scanners, which were generally expensive and rather large, arrived on the market at the end of the previous century. Because of the technological advances made the last decades, scanning becomes cheaper without compromising accuracy [1]. Especially the arrival of depth cameras, such as the Microsoft Kinect, stimulated the 3D body scan industry. These cameras are not only inexpensive, compact and easy to use, they can also capture data at video rate [2]. The SizeStream 3D body scanner is a scanner based upon 14 Kinect cameras and generates not only a digital copy of the full body in only a couple of seconds, it also derives over a hundred one-dimensional body measurements from this 'scanatar'. These qualities make the SizeStream body scanner attractive for the apparel industry, when the measurements are accurate.

Even less expensive and more accessible are systems that are based on two-dimensional photos taken with a digital camera (eventually of a smart phone or tablet). These systems combine shape information of the photos taken from multiple angles to build up a 3D image. The last years some attempts are made to create a 3D avatar from a single photo [3], and although the avatars are visually pleasing, they are not accurate enough to extract body dimensions. The solution is to increase the number of photos, so shape information of the silhouettes and its relation to the human body can be obtained [4]. The Poikos modeling system aims to serve, among others, the mass-consumption, health care and retail market, by generating 3D avatars as well as more than 100 one-dimensional derived body dimensions from an image taken with a smart phone, tablet or laptop from the front and side.

---

* t.e.vonk@hva.nl; +31621157664; http://amsterdamcreativeindustries.nl/lab/fashion-technology; www.amfi.nl

Although several high end 3D body scan systems are evaluated [1], [5], no studies to our knowledge are conducted to assess the accuracy and validity of the SizeStream body scanner in real life, nor of the Poikos modeling system. SizeStream reports that a cylinder test resulted in a circumferential accuracy of 5mm, but no information is given about the validity or repeatability of the scanner when human subjects are getting scanned. Since the shape information of a human body is much more complex than a simple cylinder, and the accuracy of the scan is affected by respiration or foot placement [6], establishment of the repeatability and accuracy in real life is important. Therefore, this study evaluated the repeatability and validity of the SizeStream scanner and Poikos modeling system by scanning a large amount of subjects multiple times. The repeatability and accuracy was captured by comparing derived measurements of the sequential scans of the same system, while the validity was measured by comparing the two systems and collected tape measurements.

2. Methods

2.1. Participants

Hundred sixty-six subjects were recruited to participate in the study, from whom 156 gave their informed consent. The participants were mainly recruited from the Amsterdam Fashion Institute (AMFI), were the SizeStream scanner is located.

2.2. Material

The SizeStream body scanner is used to collect 3D body scans of the subjects. The scanner consists out of 14 infrared depth sensors, located under 6 angles at 7 different heights. The sensors are fixed in a welded steel construction. The sensors are connected through UBS ports to a Windows 7 –64 bit computer. SizeStream scanner software is used to control the scanner and load in the data. While the study was running, the software was updated 4 times and new measurements were added. This causes that some measurements are not calculated for all subjects. Subject 1-6 are scanned with version 4.0.0, subject 7-17 with version 4.3.0, subject 18-23 with version 4.4.1, and subject 24-166 with version 4.4.3. For this experiment, the scanner was set at the highest accuracy. The SizeStream software averages the results of 2/3/4 or 5 scans per trial, to reduce the effect of small (breathing) motions. Since a pilot test showed no difference between 3,4 or 5 scans per trial, every subject was scanned 3 times per trial. One trial took roughly 30 seconds, and loading the data and calculating the derived measurements 60 seconds.

An offline version of the Poikos modeling system was used to collect a front and side photo of the participants. A MacBook Pro (OS X 10.9.?, Intel HD Graphics 4000, 1024 MB) was used to run the application. An outline of a human body were subjects had to fit into was projected on the computer screen. The software was counting down when the desired stature was adopted. Both collected photos had to be send to Poikos to process them and collect the measurements. Currently there are two methods to process the scans, namely using the raw photos (RP) and processing the photos through manual segmentation in Photoshop (MS), both in combination with the Poikos software (v0.61). Except from the photos, subjects' body height was needed to generate a correct 3D avatar and 1D body measurements. Stature (body height) was measured according to ISO 7250.

Tape measurements were carried out by multiple first and second year design students of AMFI. Most measurements were collected by fellow students during classes the participants' were attending. All measurers were instructed how to take correct tape measurements.

2.3. Protocol

The SizeStream scanner was calibrated once a day, before the start of the measurements. All recruited participants came in on a set timeframe and were instructed about the study. After signing the consent form, each participant was asked to fill in their characteristics (age, length, weight, sex) and change out of their cloth except from their underwear. If subjects were uncomfortable standing in the scanner in their underwear, they were allowed to wear a special, not body deforming, close fitting short and top over their underwear. To make sure the neckline could be calculated correctly, participants with long hair were asked to tie their hair at the back of their head.

Each participant was scanned with the SizeStream 3D scanner. Recommendations of the company regarding the posture were followed: participants had to stand with the leg slightly spread and the elbows slightly bended while holding a handle on each side. The researcher checked the posture before measurements started. Afterwards, subjects were asked to step out and stand in front of the laptop to take a front- and side picture needed for Poikos. Again, the posture was set by the company and was similar to the posture recommended by SizeStream. This protocol was repeated in total for roughly half of the participants, while the other half only took the photos twice.

2.4. Data analysis

A list with all calculated body measurements by SizeStream was saved as a text file. Matlab (Mathworks, R2014b) was used to order and transport these measurements to a spreadsheet. Statistical analyses were carried out by SpSS Statistics (IBM, 22.0.0.0). A two-way mixed reliability analysis was done to compute the intra-class correlation coefficients (ICC) and standard error of measurement (SEM) [7] and paired t-tests were carried out to compare the results of the both scanning systems and the collected tape measurements.

	SizeStream				Poikos			
	Body height (mm)	Age (years)	Weight (kg)	Sex (m/f)	Length (mm)	Age (years)	Weight (kg)	Sex (m/f)
Mean	171	22	63		172	21	63	
SD	82	5	10	10/70	85	4	10	20/119
Min	155	16	46		155	16	43	
Max	188	59	92		195	59	95	

Table 1: Participants' characteristics per group (SD=standard deviation)

3. Results

3.1. Participants

Hundred thirty-nine participants were scanned twice with the Poikos modeling systems. Eighty-five of them were also scanned twice with the SizeStream scanner, of whom 5 had to be excluded because of a software error. Three participants' measurements above the chest circumference line were excluded because they forgot to tie their hair up, and only the neck/collar measurements were excluded for one subject. Tape measurements of 59 participants were collected; the waist and hip circumference of 52, the chest circumference of 33, the neck-bustpoint distance of 13, arm length of 24 and wrist circumference of 25 participants. An overview of the participants' characteristic can be found in table 1. No significant differences were found between the groups scanned twice with the SizeStream scanner and the Poikos modeling system.

3.2 SizeStream Scanner

Of the 161 automatic calculated measurements, 120 had an ICC above 0.90, while 20 had one below 0.80. Since the ICC represents the amount of variance in a variable that is caused by the true value (1-ICC represents the variance attributable to error) [7], ICCs below 0.80 are treated as measurements with low repeatability. The 20 measurements with low repeatability can be found in table 2.

Wrist circumference right	0.29
Across axilla chest front length	0.45
Shoulder arc right	0.50
Across axilla chest back length	0.53
Shoulder arc left	0.59
Shoulder length left	0.60
Back vertical rise	0.61
Wrist circumference left	0.66
Front vertical rise	0.70
Shoulder slope left	0.70
Shoulder length right	0.70
Back shoulder width horizontal	0.71
Arm volume left	0.73
Abdomen rise	0.74
Halter	0.75
Shoulder slope right	0.76
Back shoulder width at 45 degree angle	0.77
Arm hole circumference left	0.77
Back shoulder width through back of neck	0.77
Neck circumference	0.79

Table 2: SizeStream derived measurements with ICC values below 0.80

The accuracy of the scanner is measured with the SEM. The range in which the true value lies with 95% certainty can be computed with:

$$T = OM \pm 1.96(SEM)$$

where T is the true value and OM is participant's observed (scan-derived) measurement. Knowing this, measurements with SEMs above 10mm are treated as not accurate enough for the fashion industry, since it leads to ranges of almost 4 centimeters.

The mean SEM of the 130 length or circumference measurements (volume and surface measurements are not analyzed) with ICC values above 0.80 was 10.1mm (SD = 6.4mm). Seventy-nine measurements had SEMs below and 51 above 10.0mm.
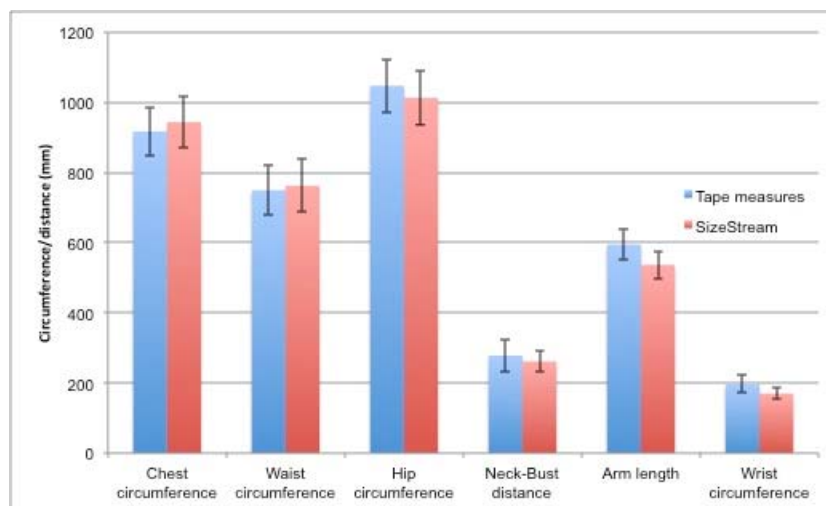

Fig 2: SizeStream versus tape measures

The validity of the scan-derived measurements can be tested by comparing these measurements with tape measurements (Fig. 2). Eleven different tape measurements were collected, of which 6 were also derived by the scanner: chest, waist, hip, wrist circumference, neck-bustpoint distance and arm length. A paired t-test showed a significant difference (p<0.001) between the all scan derived measurements and their corresponding tape measures except from the neck-bustpoint distance (figure 4). Although the measurements were significantly different, strong and significant correlations were found the chest (r^2 = 0.95, p<0.001), waist (r^2 = 0.92, p<0.001) and the hip circumference (r^2 = 0.96, p<0.001).

3.3 Poikos modeling system

Fourteen of the 56 derived measurements by the Poikos modeling system through MS had an ICC below 0.80 and were judged as low repeatable. Of the other 42 measurements, only two had an ICC above 0.90, namely the bizygomatic breadth and the head breadth. The mean SEM of the measurements with an ICC above 0.80 was 54.5mm (SD = 30.4). Except from the two breadth measurements with an ICC above 0.80, all measurements had a SEM above 10.0mm.

Only 29 participants could be included in the repeatability study when RP were used, because of the quality of the build-up 3D models. Of the 110 measurements derived in this process, 103 had an ICC above 0.90 and 84 had one above 0.90.

The chest, waist, hip circumference, and the arm length were measured with a tape measure as well as derived with the Poikos modeling system. Since only two of the participants with high quality RP scans were also measured with a tape measure, also the RP scans with moderate quality were added to be able to compare the different measurement methods. Figure 5 shows clearly that the spread in the MS derived measurements is much larger than in the tape and RP derived measurements. Significant differences were found between the tape and the Poikos waist circumferences (p < 0.001). Although no significant differences were found between the other tape and MS measurements, the correlations between the measurements derived by Poikos and collected with tape measures were weak (r^2 < 0.60). This in contrast with the correlations found between the tape and the RP measurements; the chest and waist circumferences were moderately correlated (r^2 = 0.76 & 0.68 resp.) and the hip circumferences were strongly correlated (r^2=0.86). No correlation was found between the arm length measurements.
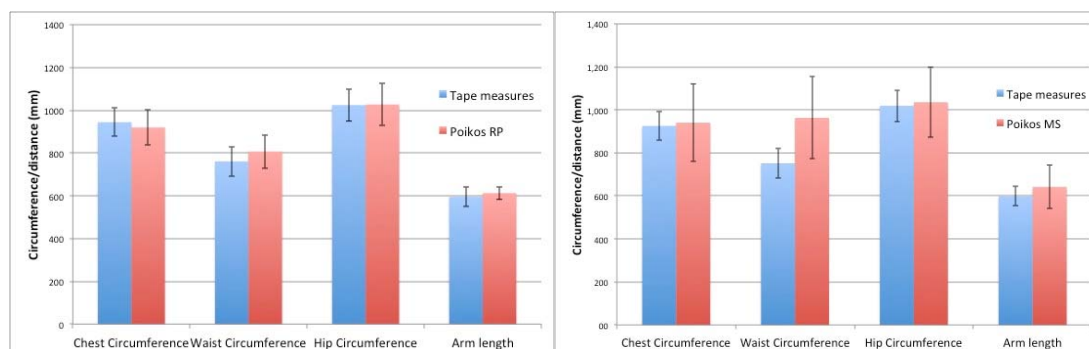
Fig 3: Poikos derived versus tape measures

To validate more measurements derived by Poikos, the measurements are compared with the matching SizeStream measurements. Twenty-two Poikos MS and twenty-five RP measurements were compared with matching SizeStream measurements. Only 10 MS measurements had significant, but weak, correlations with the SizeStream measurements ($p < 0.05$, $r^2 < 0.4$). In contrast, all RP measurements except from the head circumference were significantly correlated with their SizeStream counterpart ($p < 0.05$). Eleven couples were only very weakly correlated ($r < 0.5$), while five had moderate to strong correlations ($r^2 > 0.7 < 0.9$), namely right elbow height, waist height, crotch height, vertical trunk circumference and thigh circumference. The mean difference between the RP and SizeStream derived measurements with correlations above $r^2 = 0.5$ can be found in figure 6.
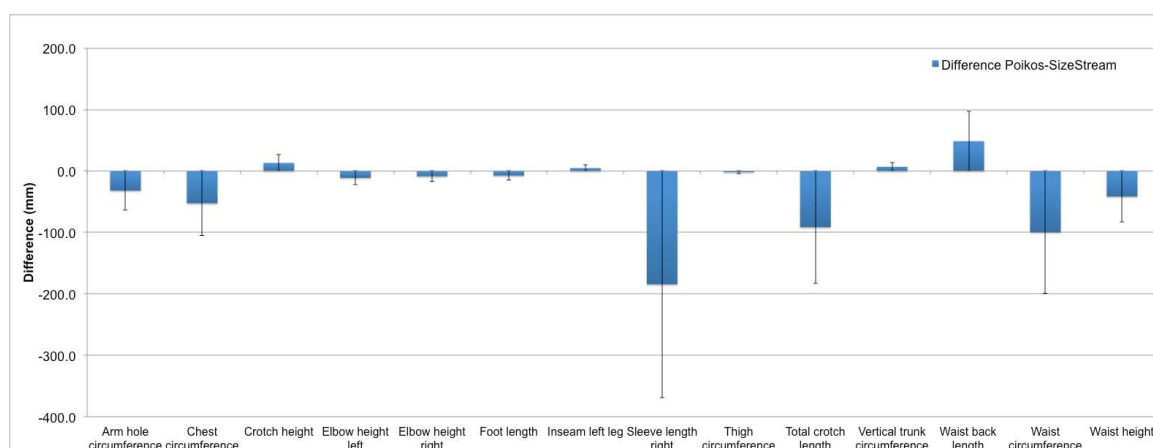

Fig 6: Difference between Poikos RP and SizeStream measurements

4. Discussion

The high ICC scores for more than 85% of the calculated SizeStream variables suggest that the scanner has good repeatability. Although the high ICC scores indicate high repeatability, only 79 measurements had SEM values under 10.0mm. Since this means that the other measurements have a 95% confidence interval of more than 40.0mm, they are not accurate enough to use in the fashion industry. It is striking that most of the 20 measurements with low ICC scores are located around the shoulder and arm and none are located in the lower body. Occlusion errors in the armpit of the scan may explain the low repeatability of some of the measurements. Since the sensors are only located at the front and back of the participant, the lights are barely or not reaching the armpit, which means that the shape has to be interpolated from information of the surrounded areas. This, and the fact that significant differences but no correlations were found between the arm length and wrist circumference of the SizeStream and tape measures, suggest that it may be better to use tape measures to collect measurements of the arm and shoulder.

The chest, waist and hip circumferences are important measurements in the fashion industry. Although the chest circumference had an SEM value just above 10.0mm (namely 10.1mm), all three measurements were rated as highly repeatable based upon the ICC scores. The strong correlations, but significant differences, between the tape measures and the SizeStream measurements indicate that the SizeStream scanner gives a good indication but not absolutely correct value. This is in accordance with [8], who found also differences between tape and scan measurements. However, their scan measurements were generally larger than the tape measurements, while in this study scan

derived hip circumferences were significant smaller. This may be due to the difficulty to measure the hip circumference by tape measure. While the scanner finds the highest circumference measured perfectly horizontal to the floor, it is hard to hold a tape measure horizontal over the buttocks. It may have happened that the tape measures were slightly tilted which could have increased the circumference.

It is not surprising that no correlations were found between the tape and SizeStream wrist circumferences, because of the low ICC scores of this SizeStream measurement. This may be caused by the errors occurring in the mesh around the hand area. It is not clear why these errors occur, but it may be due to the fact that participants have to hold handles during the scan. This and the fact that arm length are normally measured with a 90° flexed elbow, may also explain the difference in arm length. [9] found a significant effect of arm posture on the validity of scan derived arm length measurements.

Also no significant correlation is found between the neck-bust distance of the SizeStream and tape measurements. Multiple measurers stated that they always asked the participant to point out the location of their nipple, since it is very hard to locate based on the shape of the breast. Except from this, female participants were wearing bras that influenced the shape of their breast. Since the SizeStream scanner has to locate the nipple using only shape information, it is not surprising that the neck-bust distance is not similar to manual measurements. To confirm this theory, the landmark location used by the SizeStream scanner has to be compared with the real location of the nipple.

Hip, chest and waist SizeStream and tape measurements are strongly correlated, which suggests that they are valid after a correction for the over- or underestimation of the measurement by the SizeStream software. It is important to keep in mind that tape measurements are also not free of errors [10], [11], which could have led to an overestimation of the error caused by the scanner. To make sure the right correction is used, a study has to be done where clothing based on the scan-derived measurements is actually made, fitted and evaluated.

The Poikos modeling system is, as expected, less repeatable than the SizeStream scanner. However, when the RP method is used, the system is repeatable for over hundred derived measurements. Besides the fact that only ~20% of the scans were of good quality, the moderate correlations between the chest and waist circumferences suggests that the Poikos RP measurements give a good indication, but not the real value for most people. The significant difference between the waist circumferences of Poikos and manual measurements indicates an overestimation of this measurement by the Poikos modeling system. The lack of a significant difference and the strong correlation between the hip circumferences of Poikos and manual measurements indicates that this RP measurement is valid. This, and the correlations between the SizeStream and Poikos derived measurements, suggest that the avatars produced with the RP method give a good indication of the real body shape. This is in line with the results found by [4], who also compared their constructed 3D avatar based on 2D images with an avatar produced by a 3D scanner. They found a mean dimensional difference of 12.8mm(SD = 10.3) by comparing the shapes of both avatars.

Since arm length derived by the SizeStream scanner is invalid, it is likely that the sleeve length is also invalid. The big difference found between the Poikos and SizeStream sleeve length does not have to mean that the Poikos measurements is invalid. This also holds for the armhole circumference. To validate these measurements, tape measures have to be carried out.

5. Limitations & future implications

Tape measurements were collected by multiple students, without supervision of the researchers. While all students were thoroughly instructed by a teacher, it cannot be guaranteed that all collected measurements were carried out correctly. Except from that, [12] showed inter- and intra-observer errors when measures had to landmark different points at the body. To validate the tape and scan measurements, a study were patterns are made based on both measurements has to be carried out to be able to evaluate the difference in fit.

Although the computed correlations give some insight in the relationships between different measurements, it is not the best way to assess the agreement of the different measuring methods. Future analysis using Bland Altman plots has to establish if the measurements that are correlated are indeed in good agreement with each other.

The SEM of 10 mm used in this study is not a proven borderline for the fashion industry. Following research has to be done to figure out if this threshold is indeed correct, or if measurements with a good ICC and a higher SEM are still repeatable enough for pattern making. It is important that guidelines regarding variability in measurements are made, since body scanning will be more and more implemented in the fashion industry.

Most participants were students at AMFI, which means that the subject pool was not a representative part of the (Dutch) population. Although it is not expected that ICC or SEM values are influenced by this, it may be that validity of measurements are since [8] showed a significant effect of age and BMI on the mean difference between scan- and tape measurements. To validate the scanner for the whole population, some extra people have to be included.

6. Conclusion

The ICC and the SEM results indicate that 79 of the 163 SizeStream measurements are repeatable enough to use for fashion purposes and 51 give a good indication but are not accurate enough for pattern making. The 20 measurements with low repeatability have to be ignored till their repeatability is improved. Validity of the measurements has to be tested more thorough, since this study could only validate 3 of the 6 compared measurements. Since three of the six validated SizeStream measurements are rated as not valid, it is important that the other measurements are validated as well before they are used in the fashion industry.

The Poikos modeling system is a promising, but is as expected, less repeatable and valid than the SizeStream scanner. The RP method is much better than the MS method, except that only 29 scans could be used. When the company solves this issue, or at least gives immediately feedback to the user about the quality of the build scan, is the Poikos modeling system a useful manner to give an indication of the shape of the human body. It is not accurate enough to use in fashion though.

References

[1]   H. a M. Daanen and F. B. Ter Haar, "3D whole body scanners revisited," *Displays*, vol. 34, no. 4, pp. 270–275, 2013.

[2]   J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using kinects," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 4, pp. 643–650, 2012.

[3]   S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, "Parametric reshaping of human bodies in images," *ACM Trans. Graph.*, vol. 29, no. 4, p. 1, 2010.

[4]   Y. L. Lin and M. J. J. Wang, "Constructing 3D human model from front and side images," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5012–5018, 2012.

[5]   M. Kouchi, M. Mochimaru, B. Bradtmiller, H. Daanen, P. Li, B. Nacher, and Y. Nam, "A protocol for evaluating the accuracy of 3D body scanners," *Work*, vol. 41, no. SUPPL.1, pp. 4010–4017, 2012.

[6]   L. Mckinnon and C. L. Istook, "Body scanning: The effects of subject respiration and foot positioning on the data integrity of scanned measurements," *Journal of Fashion Marketing and Management*, vol. 6, no. 2. pp. 103–121, 2002.

[7]   J. P. J. Weir, "Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM.," *J. Strength Cond. Res.*, vol. 19, no. 1, pp. 231–40, 2005.

[8]   H. Han, Y. Nam, and K. Choi, "Comparative analysis of 3D body scan measurements and manual measurements of size Korea adult females," *Int. J. Ind. Ergon.*, vol. 40, no. 5, pp. 530–540, 2010.

[9]   J. M. Lu, M. J. J. Wang, and R. Mollard, "The effect of arm posture on the scan-derived measurements," *Appl. Ergon.*, vol. 41, no. 2, pp. 236–241, 2010.

[10]   C. C. Gordon and B. Bradtmiller, "Interobserver Error in a Large Scale Anthropometric Survey," *American Journal of Human Biology*, vol. 4. pp. 253–263, 1992.

[11]   M. Kouchi, M. Mochimaru, K. Tsuzuki, and T. Yokoi, "Random errors in anthropometry.," *J. Hum. Ergol. (Tokyo).*, vol. 25, no. 2, pp. 155–166, 1996.

[12]   M. Kouchi and M. Mochimaru, "Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry," *Appl. Ergon.*, vol. 42, no. 3, pp. 518–527, 2011.