Published in final edited form as:

Med Eng Phys. 2009 December ; 31(10): 1337–1342. doi:10.1016/j.medengphy.2009.08.011.

Reliability, validity, and precision of an active stereophotogrammetry system for three-dimensional evaluation of the human torso

Scott M. Paul, MD¹, Andrew P. Chamberlin, BS¹, Charles Hatt, MS², Amritha V. Nayak, MS³, and Jerome V. Danoff, PT, PhD¹

¹ Rehabilitation Medicine Department, Clinical Center, National Institutes of Health, Bethesda, Maryland ² Michigan Tech Research Institute, Ann Arbor, MI ³ Department of Biomedical Engineering, Catholic University of America, D.C

Stereophotogrammetry (SP) dates back to the 1800's, around the dawn of photography. It uses the differences in a set of two-dimensional photographs taken simultaneously from slightly different viewpoints to create a three-dimensional (3D) image. The first suggested medical application of SP was by Dr. Oliver Wendell Holmes, Sr. (the father of the Supreme Court Justice). He proposed to use SP to design better fitting prostheses for wounded soldiers during the U.S. Civil War.¹ Reliable and precise measurement of the superficial human anatomy in three-dimensions is important in the diagnosis, monitoring, and (in instances such as the one suggested by Dr. Holmes) treatment of medical conditions that have dysmorphologies due to skeletal, soft tissue and skin pathologies. Many medical conditions result in asymmetry of body parts, and changes in surface area and/or volume may be predictive of progression in conditions such as scoliosis, subcutaneous tumor growth, or lymphedema. With the development of appropriate algorithms, SP could be used to monitor these and other conditions, as well as to determine efficacy of intervention treatments.

However, until the advent of computerized image processing, technological limits prevented SP from becoming a viable medical tool. With modern computers, digital SP can non-invasively create a detailed 3D computer model of the body to assess and quantify a patient's morphology. For quantification of complex 3D geometry, modern digital SP offers many advantages over other current technologies, such as CT and MRI, including instantaneous image capture, reduced cost, no patient exposure to ionizing radiation, and no need for sedation for the reduction of movement artifact. It also offers advantages over simple manual techniques, such as tape measures and calipers, including image storage to allow quantitative comparisons over time and the ability to standardize measurement techniques. Additionally, SP may offer greater precision than manual techniques.²

Many conditions of interest to rehabilitation professionals, such as osteogenesis imperfecta, neurofibromatosis, truncal obesity, and lymphedema associated with mastectomy could be examined with torso-oriented SP. However, reliability, validity and precision of both the

Conflict(s) of interest: None

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

image capture and the image analysis processes for SP must be established before SP of the torso can become an accepted clinical or research outcome measure. Most of the previous studies of the technical capabilities of SP have been limited to craniofacial research.^{3–8} A few investigators have looked at applications involving the torso, but they have limited their work to imaging and analysis of the female breast.^{9–11} In addition, these studies were limited to a 180 degree capture field. There are very limited studies of the technical capabilities of 360 degree torso SP.^{12–14} This is a necessary precursor to applying SP to assessment of scoliosis, rib deformities and other truncal dysmorphologies. Therefore, the purpose of this study was to determine the reliability, validity and precision of a commercially available 360 degree SP system.

METHODS

A. Device

The 3dMDtorso^a is a commercially available, Windows PC based, active SP system including hardware and software. The system employs “pods” consisting of one flash, one speckle pattern projector, two black and white digital cameras, and one color digital camera. The projection of the speckle pattern on the subject’s surface is captured stereoscopically by the digital cameras in each of the pods. The color cameras are used to capture the surface texture of the subject. The system utilizes a proprietary active image capturing technique which can acquire all requested images in ~1.5 ms. The images from the four pods are combined using proprietary software based on the Tsai algorithm¹⁵ to produce a 3D image (Figure 1). The pods can be set up to capture either a field of 180 or 360 degrees (Figure 2a and 2b). The analysis of images was carried out using the 3dMD Acquisition Version 5.2.2.^a

B. Items Measured

Solid half-spheres, cones, rectangular boxes and cylinders were obtained (Table 1). The linear dimensions were manually measured with a caliper, and mathematical equations were then used to calculate the surface areas and volumes. Water displacement was used to confirm the volumetric calculations.

A “Lady-Swing” female human-form mannequin^b, constructed of polyurethane foam over bendable metal armatures and covered with a plastic skin-like material, was used to simulate a human form. A female model was chosen to provide increased anatomical complexity. Small colored pins were used to mark specific anatomical points including the acromion processes, nipples, scapulae and xiphoid process.

Prior to image capture, two raters familiarized themselves with the system until they were proficient at performing the required operations. A series of tests were then performed to determine the stability, reliability, validity, and precision of both the image capture and the image analysis processes.

C. Experiment 1: Simple surface with geometric solids

The goal was to determine the reliability, validity and precision of measurements obtained by the system in a simplified environment. The camera pods were set up in 180 degree mode. The geometric solids were attached to a flat board using hook and loop fastener tape. The board, with the attached solid, was positioned in front of the pods using a multi-axial stand. After capturing all of the images, the two raters took turns taking measurements using the system’s software to determine the linear dimensions of the geometric solids. Intrarater

^a3dMD, LLC, Atlanta, GA

^bBendies Forms, Inc., South Granby, QC, Canada

reliability was tested by having one rater repeat all measurement after 48 hours and again after 96 hours; thereby resulting in three sets of measurements. Intra-class correlation coefficients (ICC) were calculated to determine intrarater reliability. All ICC calculations were of the form ICC (1,K) where our raters were considered representative of all raters who might use this system, and K measurements were taken at each point. Interrater reliability was determined by comparing the measurements taken by the two raters using ICC. Criterion validity of linear, surface area, and volume measurements of the shapes calculated by the SP system were tested by comparing SP measurements to caliper-derived values and, for volume, to water displacement measurements. The caliper has an error of ± 0.5 mm and the vessel used for water displacement had an error of ± 0.5 mL. Two-way repeated measures analysis of variance (ANOVA) was used to compare the three different methods and the two raters. Images of half-spheres of decreasing size were also analyzed to test the limits of system precision.

D. Experiment 2: Human-form mannequin

The stability of system image capture was tested by taking repeated images of the human-form mannequin over time. The mannequin was placed in a standing upright position, facing pod 1, and images were obtained every 20 minutes for 5 and 2/3 hours to obtain a total of 17 images. This process was performed in 180 degree mode and then repeated in 360 degree mode. A one-way repeated measures ANOVA was used to determine stability in the image capture. In addition, two raters took turns using the analysis software to calculate the surface distances between the pre-placed pins. Interrater reliability was tested by calculating ICC (1,K) values between these two sets of measurements. The mannequin was measured manually using an Anthropometer^c and Gulick Anthropometric Tape^d and these results were compared to the SP results using two-way repeated measures ANOVA to determine concurrent validity. ICC was also calculated to compare results between the two raters.

E. Experiment 3: Human-form mannequin with geometric solids

The system's ability to measure clinically relevant dysmorphology was modeled by placing geometric solids onto the mannequin. A series of images, in both 180 and 360 degree modes, were acquired with each of the geometric solids attached to the anterior torso of the mannequin. The two raters took turns using the software to calculate the surface area and volume of the geometric solids. The interrater reliability was tested by calculating ICC (1,K) values between SP calculations of the two raters. Concurrent validity was determined by comparing the SP calculated measurements to the caliper-derived and geometrically calculated measurements using repeated measures ANOVA. The volumes calculated by the system were also compared to water displacement results. As in Experiment 1, images of half-spheres of decreasing diameter were analyzed to test the limits of precision of the system.

To further test the validity of the system to measure the surface area of specific body parts, the surface area of the mannequin's scapula and breast were calculated using the system's software package. The area to be measured was kept consistent by marking the mannequin using pins to delineate the margins of any structures distinct from the surface. Manual measurements of the scapula and breast were taken using alginate gel^e and the method outlined by Lee and Choi.² The alginate gel was scanned into a computer using TWAIN software included with a PSC 2510 scanner^f and measured using ImageJ software^g. A 95%

^cLafayette Instrument Company, Lafayette, IN

^dPatterson Medical Products, Inc, Bolingbrook, IL

^eRite-Dent Manufacturing Corporation, Hialeah, FL

^fHewlett-Packard, Palo Alto, CA

^gResearch Services Branch, National Institutes of Health, Bethesda, MD

confidence interval, using the t-statistic, was calculated to determine the concurrent validity of SP when compared to the alginate gel method.

RESULTS

A. Experiment 1

Individual ICC values were calculated for linear measures, surface area and volume. Intrarater reliability was strong for all measures with ICC's of $r > 0.99$ ($P < 0.001$). There were also strong interrater correlations ($r > 0.99$, $P < 0.001$) (Table 2).

There was strong correlation on ICC between manual and SP measures (see Table 2). In addition, results from the two-way repeated measures ANOVA showed no difference between SP and manual techniques for linear ($P > 0.28$), surface area ($P > 0.06$) and volume ($P > 0.17$) measurements. There were no significant differences between SP calculated volumes and the volume determined by fluid displacement (Table 3). There were also no differences between the two raters for any of the measurements.

The percent error in the surface area and volume measurements for the different size spheres were compared to examine the limits of precision. A logarithmic regression provided an excellent fit for the surface area ($r = 0.97$) and volume ($r = 0.97$) measurements (Figure 3).

B. Experiment 2

ICC was used to compare measurements taken from the 17 images by two raters of the mannequin in a fixed position over time. Collected data were very stable with negligible detectable change over several hours of collection (Figure 4). Good stability was established by a high ICC correlation ($r > 0.99$). There were no statistical differences between manual and SP caliper measurements ($P > 0.73$). However, there was a significant difference between manual and SP methods when taking circumferential measurements ($\Delta\bar{x} = 0.79$ mm, $P < 0.02$). Despite this significant difference, there was a strong correlation between the manual and SP measurements, for both caliper ($r > 0.98$) and surface distance ($r > 0.99$) measurements in both 180 and 360 degree modes. Results for Experiment 2 can be seen in Table 4.

C. Experiment 3

In this experiment, geometric solids were placed on the mannequin, and calculations of surface area and volume were made for these objects. The interrater reliability was strong for surface area ($r = 0.99$) and volume ($r = 0.99$). There was no statistically significant difference between SP-calculated and geometrically derived surface areas ($P > 0.07$) or volumes ($P > 0.15$). In addition, there were strong correlations between SP and geometrically calculated surface area ($r = 0.99$) and volumes ($r = 0.99$, Table 5). There was no significant difference between SP calculated volumes and fluid displacement measurements ($P > 0.68$).

The mean surface area of the scapula was 43.1 cm^2 when measured with the alginate gel method. The 95% confidence interval for the SP calculated surface area of the scapula was $37.1 \pm 0.4 \text{ cm}^2$. Using alginate gel, the mean surface area of the breast was 301.4 cm^2 . The 95% confidence interval for SP-calculated surface area of the breast was $312 \pm 2 \text{ cm}^2$. Based on these two measurements, there is a modest, but statistically significant difference between the SP and alginate gel methods in determining surface areas.

A logarithmic regression was the best fit when comparing percent error and object size for the surface area ($r = 0.97$) and volume ($r = 0.94$) measurements of different sized half-spheres when placed on the human-form mannequin (Figure 3).

DISCUSSION

In this study, we have assessed the capabilities of a clinical SP system beginning with simple geometric shapes and progressing to a more complex model using a human-form mannequin. In each of these experiments, excellent intra- and interrater reliability was demonstrated by strong intraclass correlations within rater and between raters. The SP system also demonstrated excellent stability over repeated measurements. In general, very good validity with reference to manual techniques was demonstrated.

There were statistically significant differences between SP calculations and the manual measurements of linear distances along a surface by tape measure and of surface areas by alginate gel method. These findings may be attributable to technical differences in these methods. The pressure of the Gulick Anthropometric tape on the surface of the compressible polyurethane foam mannequin causes some deformation, whereas the SP system makes no physical contact with the model. The difference, although statistically significant, was sub-millimeter and will not be clinically important for most applications. Similarly, the placement of the alginate gel onto the surface of the mannequin may cause deformation to the body of the mannequin. There may also have been a change in size of the alginate sheets between removal from the mannequin and placement in the scanner, because alginate gel contracts while curing. Although we transferred the gel to the scanner as quickly as possible, some contraction could have occurred. Lee and Choi² also noted that when alginate gel was compared to other measurement methods, the greatest difference was between alginate gel and a 3D scanning device.² In addition, the greatest error in alginate gel measurements of surface areas occurred when measuring smooth surfaces with large inclines.² The smooth skin-like material and the upright standing position (incline ~ 90 degrees) of the mannequin may have also contributed to differences between our alginate gel and SP measurements. Another geometric issue may be related to the specific shape of the surface being scanned. A straight line or flat surface would be amenable to more precise calculations for distance, area, or volume by the system software. When curved surfaces are encountered, software algorithms would need to utilize limit-based calculations. The accuracy of these types of calculations may be processor dependent. Our results indicated excellent accuracy for most curved surfaces, and we would expect these to become even better with any future improvements in software or hardware.

Since the anatomical accuracy of the mannequin was somewhat limited, pins were placed to highlight anatomical points. It is possible that this gave the system an “unfair” advantage. We have not yet assessed the ability to accurately identify landmarks for measurements directly from an image, without marker assistance, in comparison to direct observation and/or palpation of the subject. Thus, until that validity and accuracy is established in future studies, marking of landmarks on the subjects prior to image capture is recommended.

Although only a limited number of values were obtained, our results suggested that this SP system has a limit of precision for measuring smaller objects. The regression curves obtained show that the error increases exponentially as the object size decreases. We demonstrated that a 5% error in surface area calculations may occur in objects smaller than 23.5 cm^2 . Similar results were obtained for the percent error in volume measurement, which suggests that a 5% error may occur in volumes less than 32 mL. We did include a wide range of sizes; however, additional studies of different sized spheres and other shapes may be desirable. In terms of human measurements, our data suggest acceptable accuracy for

clinical studies involving structures such as breasts, upper and lower limbs and shoulders. Smaller structures such as individual finger joints or small surface deformities (e.g. paracutaneous tumors) may be at the limits of accuracy of the current system.

Our study produced results similar to others in the craniofacial SP literature with respect to validity, reliability and precision of linear measurements.^{3, 4, 7} Our approach was similar to Weinberg and colleagues, who used eighteen mannequin heads to compare linear craniofacial SP measurements.⁴ Their analysis produced results comparable to ours for both interrater reliability and validity. However, our search of the literature revealed very few studies concerned with volume calculations^{10, 16} and no studies which have analyzed reliability or precision for surface area measurements. A study by Losken and colleagues,¹⁰ attempted to validate preoperative breast volumes calculated by SP against water displacement measurements of the breast after excision by mastectomy. They reported similar interrater reliability to ours, but they had greater variability in their volume measurements. The greater variability in measurements may have been due to the uncertainty of whether the excised breast tissue represented the entire volume measured on the woman pre-operatively by SP. By using removable geometric solids and a human-form mannequin, we were able to look systematically at surface areas and volumes without similar concerns.

CONCLUSION

The results of this study support the use of SP as a tool to quantify and study complex 3D abnormalities of the human body. Validity, reliability and precision of the system can now be further examined in studies of human subjects. We plan to explore the capability of SP to serve as a clinical research tool to detect changes over time in persons with conditions including spinal and rib deformities in osteogenesis imperfecta, truncal obesity, post mastectomy edema, and dermal and subdermal lesions with neurofibromatosis Type 1.

Acknowledgments

Thanks to Ching-yi Shieh, PhD, for assistance in statistical analysis.

References

1. Herron R. A biomedical perspective on stereographic anthropometry. *Biomedical Sciences Instrumentation*. 1969; 1969:53–56.
2. Lee J-Y, Choi J-W. Validity and Reliability of an Alginate Method to Measure Body Surface Area. *Journal of Physiological Anthropology*. 2006; 25(3):247–255. [PubMed: 16763367]
3. Aldridge K, Boyadjiev SA, Capone GT, DeLeon VB, Richtsmeier JT. Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *American Journal of Medical Genetics Part A*. 2005; 138A(3):247–253. [PubMed: 16158436]
4. Weinberg SM, Naidoo S, Govier DP, Martin RA, Kane AA, Marazita M. Anthropometric Precision and Accuracy of Digital Three-Dimensional Photogrammetry: Comparing the Genex and 3dMD Imaging Systems with One Another and with Direct Anthropometry. *Journal of Craniofacial Surgery*. May; 2006 17(3):477–483. [PubMed: 16770184]
5. Ghoddousi H, Edler R, Haers P, Wertheim D, Greenhill D. Comparison of three methods of facial measurement. *International Journal of Oral and Maxillofacial Surgery*. 2007; 36(3):250–258. [PubMed: 17113754]
6. Grobbelaar R, Douglas TS. Stereo image matching for facial feature measurement to aid in fetal alcohol syndrome screening. *Medical Engineering & Physics*. 2007; 29(4):459–464. [PubMed: 16904364]

7. Meintjes EM, Douglas TS, Martinez F, et al. A stereo-photogrammetric method to measure the facial dysmorphology of children in the diagnosis of fetal alcohol syndrome. *Medical Engineering & Physics*. 2002; 24(10):683–689. [PubMed: 12460727]
8. Gregory A, Lipczynski RT. The three dimensional reconstruction and monitoring of facial surfaces. *Medical Engineering & Physics*. 1994; 16(3):249–252. [PubMed: 8061912]
9. Galdino GM, Nahabedian M, Chiaramonte M, Geng JZ, Klatsky S, Manson P. Clinical Applications of Three-Dimensional Photography in Breast Surgery. *Plastic & Reconstructive Surgery*. Jul; 2002 110(1):58–70. [PubMed: 12087232]
10. Losken A, Seify H, Denson DD, Paredes AA, Carlson G. Validating Three-Dimensional Imaging of the Breast. *Annals of Plastic Surgery*. 2005; 54(5):471–476. [PubMed: 15838205]
11. Nahabedian M, Galdino G. Symmetrical Breast Reconstruction: Is There a Role for Three-Dimensional Digital Photography? *Plastic & Reconstructive Surgery*. Nov; 2003 12(6):1582–1590. [PubMed: 14578788]
12. Jaremko, JL.; Poncet, P.; Ronsky, J., et al. *Computer Methods in Biology and Biomedical Engineering*. Vol. 5. Taylor & Francis; 2002. Comparison of Cobb Angles Measured Manually, Calculated from 3-D Spinal Reconstruction, and Estimated from Torso Asymmetry; p. 277-281.
13. Pazos V, Cheriet F, Danserau J, Ronsky J, Zernicke R, Labelle H. Reliability of trunk shape measurements based on 3-D surface reconstructions. *European Spine Journal*. 2007; 16(11):1882–1891. [PubMed: 17701228]
14. Poncet P, Kravarusic D, Richart T, et al. Clinical impact of optical imaging with 3-D reconstruction of torso topography in common anterior chest wall anomalies. *Journal of Pediatric Surgery*. 2007; 42(5):898–903. [PubMed: 17502208]
15. Tsai R. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*. 1987; 3(4): 323–344.
16. Losken A, Fishman I, Denson DD, Moyer HR, Carlson GW. An Objective Evaluation of Breast Symmetry and Shape Difference Using 3-Dimensional Images. *Annals of Plastic Surgery*. Dec; 2005 55(6):571–575. [PubMed: 16327452]

Figure 1.

Series of sample images of a 4" sphere taken from one of the imaging pods. The two black and white images (A and B) show the speckle pattern applied by the projector. The color image (C) provides surface texture which is applied to the 3D model. The wire frame model (D) is created by compiling the images from all four pods, a total of eight images (plus four color images to apply the surface texture).

Figure 2.
Diagrams of the camera pod arrangements for the two capture modes: 180 (A) and 360 (B) degrees.

Figure 3.

System precision as represented by percent error in SP measurement of half-sphere size: Percent error in the SP calculated surface area (A) and volume (B) of different sized half-spheres imaged on a flat surface and on a human-form mannequin compared to the manual measurements.

Figure 4.

The stability of the system image capture was tested by taking repeated images of the human-form mannequin over time in both 180 (A) and 360 (B) degree modes. Images were obtained every 20 minutes for 5 and 2/3 hours to obtain a total of 17 images. The four measurements that were taken on each image were: The caliper distance between nipples (NC), the surface distance between nipples (NSD), the caliper distance between acromia (AC) and the surface distance between acromia (ASD).

Table 1

The manually and SP measured dimensions of all of the geometric solids used in Experiments 1 and 3.

Object	Manual Linear (cm) ^a	SP Linear (cm) ^a	Manual Surface Area (cm ²)	SP Surface Area (cm ²)	Manual Volume (cm ³)	SP Volume (cm ³)
Truncated Cones						
Narrow	r ₁ =1.8, r ₂ =3.6, h=11.8	r ₁ =1.7, r ₂ =3.6, h=12.0	214.4	242.8	284.7	287.1
Wide	r ₁ =3.6, r ₂ =4.9, h=8.6	r ₁ =3.7, r ₂ =4.9, h=8.6	271.5	299.4	483.0	499.4
Half Spheres						
10.16 cm diameter (4 in)	r=5.1	r=5.1	160.3	166.2	269.8	275.8
6.35 cm diameter (2.5 in)	r=3.2	r=3.2	64.7	64.0	69.2	66.9
5.08 cm diameter (2 in)	r=2.5	r=2.5	40.5	40.3	34.2	33.2
1.91 cm diameter (0.75 in)	r=1.0	r=1.0	6.3	5.4	2.1	1.6
Cylinder	r=3.3, h=8.6	r=3.4, h=8.6	211.6	211.8	292.0	294.1
Rectangular Box	l=8.9, h=8.9, w=6.7	l=8.9, h=8.9, w=6.8	316.0	320.5	526.7	527.7

^ah=height, r=radius, l=length, w=width.

Table 2

Intra- and interrater reliability and concurrent validity data for caliper, surface area and volume measurements of geometric solids on a flat board.

Reliability (intrarater)	ICC	P value
Caliper	0.999	<0.001
Surface Area	0.994	<0.001
Volume	0.999	<0.001
Reliability (interrater)		
Caliper	0.999	<0.001
Surface Area	0.999	<0.001
Volume	0.999	<0.001
Validity (manual vs. calc)		
Caliper	0.999	<0.001
Surface Area	0.993	<0.001
Volume	0.999	<0.001

Table 3

Results of two-way repeated measures ANOVA between geometrically calculated and SP calculated measurements of geometric solids. No significant differences were found.

Linear	P value
SP vs. manual	0.28
Surface area	
SP vs. manual	0.06
Volume	
SP vs. manual	0.17
SP vs. fluid displacement	0.80
Manual vs. fluid displacement	0.16

Table 4

Reliability and concurrent validity data for point to point (caliper), surface distance and surface area measurements of mannequin in 180 and 360 degree capture modes.

Reliability (interrater)	ICC	P value
Caliper	0.999	<0.001
Surface Distance	0.997	<0.001
Surface Area	0.999	<0.001
Validity in 180 (SP vs. manual)		
Caliper	0.999	<0.001
Surface Distance	0.998	<0.001
Validity in 360 (SP vs. manual)		
Caliper	0.980	<0.001
Surface Distance	0.995	<0.001

Table 5

Reliability and precision data for surface area and volumetric measurements of geometric solids attached to the mannequin in 180 and 360 degree capture modes.

Reliability (interrater)	ICC	P value
Surface Area	0.999	<0.001
Volume	0.999	<0.001
Precision (SP vs. manual)		
Surface Area	0.995	<0.001
Volume	0.999	<0.001