ELSEVIER

Applied Ergonomics 31 (2000) 445–451

APPLIED ERGONOMICS

www.elsevier.com/locate/apergo

Performance of a 2D image-based anthropometric measurement and clothing sizing system

Pierre Meunier[a,*], Shi Yin[b]

[a]*Systems Modelling Group, Defence and Civil Institute of Environmental Medicine, 1133 Sheppard Ave West, P.O. Box 2000, Toronto, Ont., Canada M3M 3B9*

[b]*Image and Vision Systems, 1202-30 Godstone, Toronto, Ont., Canada M2J 3C6*

Received 9 September 1999; accepted 31 March 2000

Abstract

Two-dimensional, image-based anthropometric measurement systems offer an interesting alternative to traditional and three-dimensional methods in applications such as clothing sizing. These automated systems are attractive because of their low cost and the speed with which they can measure size and determine the best-fitting garment. Although these systems have appeal in this type of application, not much is known about the accuracy and precision of the measurements they take. In this paper, the performance of one such system was assessed. The accuracy of the system was analyzed using a database of 349 subjects (male and female) who were also measured with traditional anthropometric tools and techniques, and the precision was estimated through repeated measurements of both a plastic mannequin and a human subject. The results of the system were compared with those of trained anthropometrists, and put in perspective relative to clothing sizing requirements and short-term body changes. It was concluded that image-based systems are capable of providing anthropometric measurements that are quite comparable to traditional measurement methods (performed by skilled measurers), both in terms of accuracy and repeatability. Crown Copyright © 2000 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Anthropometry; 2D-measurement system; Automated clothing sizing

1. Introduction

In spite of highly standardised protocols designed to maximize the degree of repeatability and accuracy of measurements, anthropometric data are not always as reliable as they appear. Many factors come into play during the measurement of human subjects, which can result in the appearance of numerous sources of error. Some of the important sources include posture, identification of landmarks, instrument position and orientation, and pressure exerted by the measuring instrument (Davenport et al., 1935). The difficulty in controlling all potential sources of error is such that it has been said that true values are seldom measured in anthropometry (Jamison and Zegura, 1974). Accuracy and precision of anthropometric measurements are at the mercy of the measurers who take them. Even if measured by highly trained observers, comparison of two populations may be meaningless (Bennett and Osborne, 1986). In a comparative study by Kemper and Pieters (1974), 50 boys (12 and 13 years of age) were measured independently by experienced observers in two institutes. Both teams of observers were trained to the same measurement techniques and used the same measuring instruments. In spite of this, systematic differences were found in nine of the 12 measurements taken. Pearson correlations between 0.872 (biacromial diameter) and 0.996 (stature) were found between the measurements taken by the two groups. Although the variable with the lowest correlation (biacromial diameter) did not present systematic errors, it suffered from repeatability problems (precision error). The results of these and many more studies show how difficult it is to measure humans, even under controlled conditions and after extensive training of the observers.

Computerized image-based systems offer an interesting alternative to traditional methods of measurement, especially in applications where individuals are measured to determine their size of clothing and equipment. They offer rapid body measurements that are quickly

* Corresponding author. Tel.: + 1-416-635-2093; fax + 1-416-635-2013.

E-mail address: pmeunier@dciem.dnd.ca (P. Meunier).

PII: S0003-6870(00)00023-5

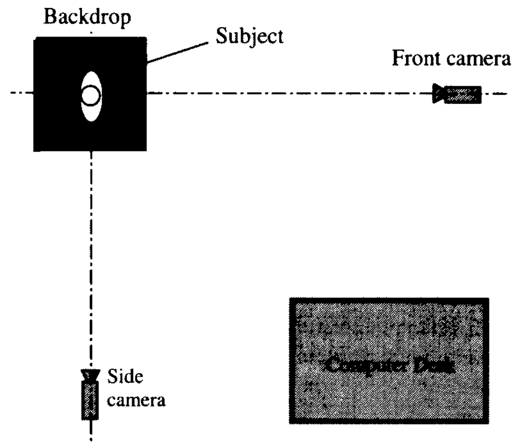

Fig. 1. Plan-view of image capture set-up.

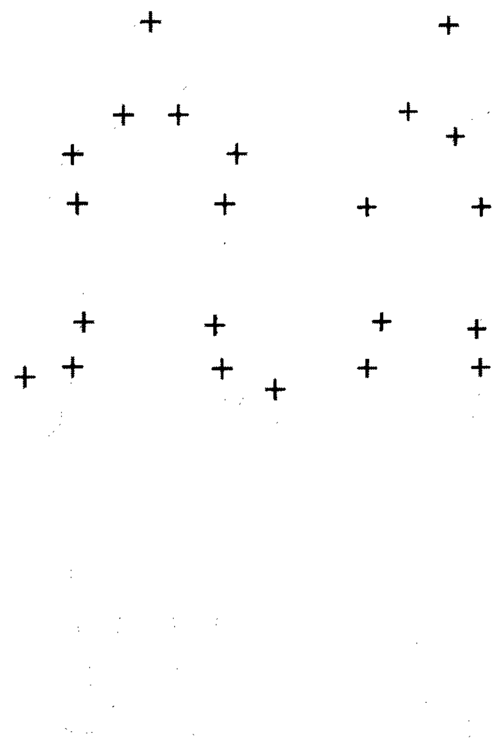

Fig. 2. Front and side silhouettes with landmarks.

translated into clothing size categories based on predetermined fit criteria. Such systems are capable of overcoming some of the problems of traditional anthropometry, but introduce error sources of their own. In image-based systems, the sources of error take the form of perspective distortion, camera resolution, camera calibration, landmarking error, and modelling error (since circumferences are not measured directly). The objective of this paper is to compare the accuracy and precision of measurements made from two-dimensional images of humans with those of highly trained anthropometrists, and put the results in perspective in the context of clothing and equipment sizing.

1.1. System description

The system under review is a PC-based system comprised of two Kodak DC120 colour digital cameras (1280×960 pixels) and a blue backdrop embedded with calibration markers (Fig. 1). The calibration of the cameras is performed using an algorithm developed by Tsai (1986). The system takes simultaneous front and side pictures of individuals standing with their arms straight and slightly abducted along their side. By taking both images simultaneously, the exact posture in space is captured, and it is possible to recover the object dimensions in 3D. Fig. 2 shows the final result, after image processing and automatic landmarking. The cross-hairs identify the location of the landmarks.

Potential sources of error can be found at each of the steps in the image analysis process illustrated in Fig. 3, namely in

(a) pre-processing of the front and side images,
(b) calibration of the cameras,
(c) segmentation of the body from the background,
(d) detection of the landmarks, and
(e) calculation of the anthropometric dimensions.

1.2. Theoretical assessment of error

The error of a measurement is defined as the difference between the measured value and the true value of the item being measured. Errors can be catalogued as either random (precision error) or systematic (bias error). *Precision* is defined as the difference in values obtained when measuring the same object repeatedly. It has an average value of zero. *Accuracy* is the difference between the measured and true values. Bias error, which occurs in the same way on each measurement, affects the accuracy of a measurement while random error affects precision.

The concept of error is useful, but it implies knowledge of the true value of what is being measured. Since any measurement contains error, the true value is never known. Although absolute error cannot be calculated, it can be estimated. Precision error, for instance, can be estimated by taking a large number of readings on an individual and using a statistical model to determine the expected spread of values at a given probability level. Bias error, on the other hand, requires comparison of measurements with a more accurate method/instrument. This is difficult to do in anthropometry, given that the best available method, i.e. traditional measurements taken by highly trained and skilled measurers, is one that contains non-negligible error itself, as discussed in the introduction.

Of the sources of error listed above (a–e), segmentation and modelling are probably the most significant. A rough estimate of segmentation error can be made from a theor-

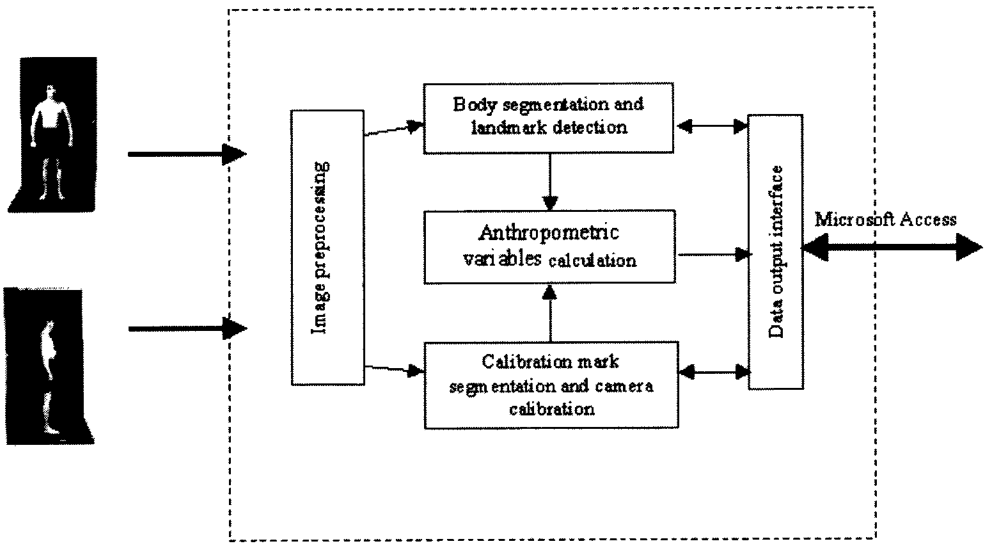

Fig. 3. Image analysis process.

etical perspective, using camera resolution as the starting point. Since the cameras used in this system have 1280×960 pixels covering an area that is approximately 2.5×1.8 m at the subject, the corresponding spatial resolution is slightly less than 2.0 mm/pixel. This resolution tends to remain constant regardless of camera position, since the camera set-up procedure requires a full view of the backdrop, which is achieved either by moving the camera and/or by adjusting the zoom. Assuming a segmentation error of plus or minus one pixel, direct measurements requiring two points (i.e. for breadths, depths, and heights) will fluctuate within ± 2 mm of the true value (1 pixel $\times$ 2 mm/pixel). The maximum error, which is obtained when both points err in opposite direction, would put the result within ± 4 mm of the true value (2 pixels $\times$ 2 mm/pixel).

Circumferences cannot be measured directly using only front and side pictures and must therefore be calculated using some form of mathematical modelling. The choice of model depends on the cross-sectional shape being measured, which varies among individuals. In two-dimensional systems, circumference measurement error depends on the accuracy of the model as well as on the breadth and depth measurements used in the calculation. Assuming a cylindrical object (i.e. no modelling error) and a one-pixel error on the diameter, the results would be likely to fluctuate within ± 6 mm ($\pi \times 2$ mm) of the true value.

2. Methodology

2.1. Accuracy assessment

The accuracy of the image-based system was assessed by comparing image-based measurements with manual measurements taken by anthropometrists during the 1997 survey of the Canadian Land Forces (Chamberland et al., 1998). Six dimensions were selected because of their relevance to clothing sizing, which is the main purpose of the system. These were: stature (as a general indicator of length measurements), neck circumference, chest circumference, waist circumference, hip circumference, and sleeve length (spine-wrist). The definitions for stature, and for neck, chest, and hip circumferences were virtually identical in both methods of measurement: stature was taken at the highest point, the neck at the Adam's apple perpendicular to the long axis of the neck, chest horizontally at the fullest part of the breast, and hip horizontally at the maximum protrusion of the buttock. The definitions for waist circumference were different: the traditional measurement of waist circumference was taken horizontally at the level of the navel, whereas the image-based system used what is sometimes referred to as "preferred" waist, i.e. where individuals wear the belt of their trousers. The side view in Fig. 2 shows that contrary to the traditional definition, the waist landmarks are not horizontal. The traditional spine-wrist sleeve length definition is somewhat incompatible with front and side views of a subject, since it requires a different posture than that used in the system and would require a top view of the subject. However, it is possible to obtain an equivalent measurement from measurements that are accessible from those views, namely sleeve outseam (from the acromion to the wrist) and biacromial breadth.

The test sample consisted of a subset of 349 subjects (95 females and 254 males) from the survey that had been measured both with traditional methods and with the image-based system. Each subject was measured once by each method. The image-based system was the last of seven anthropometric measurement stations in the survey (Chamberland et al., 1998), which means that both types of measurements were performed within a period of about 90 min, avoiding the effects of daily body variations. t-tests were performed to compare the means of all dimensions, and F-tests to compare the standard deviations. Waist circumference was excluded from this comparison due to the difference in measurement definition between the two methods.

2.2. Precision assessment

The precision of the image-based system was determined by performing ten repeated measurements on a full size plastic mannequin as well as on a human subject. All image capture and analysis sequences were performed in succession (every minute or so) such that camera calibration and lighting conditions were relatively constant. The mannequin was used in order to exclude variations due to breathing movement and postural differences from picture to picture. The human subject was instructed to stand with the arms slightly abducted along the side the body during picture taking, and to move away from the platform between measurements. Thus, contrary to the mannequin, the human precision estimates contain variability due to postural differences, breathing movement, and repositioning from one set of images to the other.

3. Results

3.1. Accuracy

The means and standard deviations for the subjects measured manually and digitally are listed in Table 1. t-tests did not indicate any significant difference between the manual and digital methods for either males or females. Table 1 also lists the Pearson correlation coefficients between manual and 2D image measurements.

3.2. Precision

Table 2 summarises the results of repeated measurements of a plastic mannequin.

Table 2
Mannequin repeatability results (mm)

Variable	Mean ($n = 10$)	Range	Std. dev.	1.96 Std. dev.
Stature	1822	2	1	1
Neck circumference	360	5	2	3
Hip circumference	946	12	4	7
Waist circumference	856	7	2	4
Chest circumference	960	10	3	6
Sleeve length	829	24	8	15

Table 3
Human repeatability results (mm)

Variable	Mean ($n = 10$)	Range	Std. dev.	1.96 Std. dev.
Stature	1817	5	2	3
Neck circumference	369	6	2	4
Hip circumference	978	11	4	8
Waist circumference	873	15	5	10
Chest circumference	964	16	6	11
Sleeve length	887	36	10	20

Table 1
Accuracy results (mm)

Measurement		Females			Males		
		Mean	Std. dev.	Correlation	Mean	Std. dev.	Correlation
Stature:		$n = 95$			$n = 254$		
	Manual	1633	60	0.98	1747	64	0.98
	2D system	1633	59		17487	63	
Neck circumference:		$n = 62$[a]			$n = 254$		
	Manual	329	18	0.88	395	23	0.94
	2D system	329	16		395	22	
Chest circumference:		$n = 88$[a]			$n = 254$		
	Manual	956	87	0.95	1024	83	0.94
	2D system	957	84		1024	78	
Hip circumference:		$n = 95$			$n = 238$[b]		
	Manual	1027	91	0.98	1005	72	0.94
	2D system	1026	89		1004	68	
Sleeve length:		$n = 95$			$n = 254$		
	Manual	799	34	0.79	876	35	0.76
	2D system	800	27		875	26	

[a]Some subjects were rejected due to interference of hair with landmarking.
[b]Some subjects were rejected due to clothing interference (boxer shorts).

The results of repeated measurements of a human subject are shown in Table 3.

4. Discussion

4.1. Accuracy

The results of t-tests comparing manual and image-based measurements showed no significant difference between the two, indicating that the two methods provide similar results. This is not surprising since the indirect measurement models of the image-based system were developed and optimized using the images and data collected in the survey. F-tests comparing the standard deviations of the two types of measurements also showed no significant difference, indicating that they were equally consistent in taking those measurements.

4.2. Precision

The theoretical assessment of the random measurement error made earlier suggested that an error of the order of ± 2 and ± 6 mm could be expected from a one-pixel segmentation error on linear and circumferential measurements, respectively. As shown in Table 2, the results of repeatability tests performed on the plastic mannequin showed the actual error to be within 1 mm of the mean for stature, 95% of the time. The fact that this is less than what could be expected from a one-pixel segmentation error indicates that the spread of measurements is probably due more to camera calibration or modelling fluctuations than segmentation. Where the mannequin's shape attributes were true to life (i.e. except for hinged joints, non-standard posture and unnatural shapes), reliable landmark positions were obtained using the automatic landmarking algorithms. Hinges at the shoulder, elbow and wrist hindered the repeatability of sleeve length measurements. Fluctuations in this measurement in particular were unavoidable because the automatic landmark detection software was developed to recognise real human shape. Circumferences were found to be fluctuate within 3–6 mm of the mean, 95% of the time (Table 2), with neck circumference exhibiting the highest degree of repeatability. This is thought to be due, in part, to the special attention paid to this measurement during development of segmentation and landmarking algorithms.

The results in Table 3 show that, for the most part, repeated measurements of a human subject exhibited the same basic trend as for the mannequin in that the linear measurements were more precise than circumferences, and neck circumference was more repeatable than other circumferences. In most cases, the human results exhibited more variability in measurement than in the case of the mannequin, which was anticipated. The largest difference between mannequin and human subject measurements were for waist and chest circumferences. This can be partly explained by torso movement during breathing (due to expansion and contraction of the rib cage and abdomen) and differences in posture from picture to picture (arm position, relaxed or tight posture).

4.3. Computer versus human measurement repeatability

The results of the repeatability study on a human subject were compared with those of recent large-scale surveys where accuracy and precision were monitored throughout. The first survey was conducted on Canadian Land Forces personnel in 1997 (Chamberland et al., 1998). The second survey was conducted on US Army personnel in 1988 (Gordon et al., 1989). Repeated measurements were part of the routine during both surveys, although the methodology was slightly different. In the Canadian Land Forces survey, subjects were re-measured by the same observers within minutes (10–90 min of the first measurement (see Forest et al., 1999 for details). This can be viewed as the best-case scenario in terms of repeatability, since it is assumed that the same observer will measure in the same way given the same landmarks. In the US Army survey, subjects were re-measured within minutes by a second observer. This case can be viewed as the best-case scenario for repeatability by different observers, since both observers were highly trained on the dimensions specific to their measuring station.

The technical error of measurement (TEM), which is essentially a form of standard deviation, was used as the basis for comparison. The TEM, or r, was calculated using the following equation:

$$r = \sqrt{\frac{\sum_{i=1}^{n}\left(\sum_{j=1}^{k} x_j^2 - (1/k)\left(\sum_{k=1}^{j} x_j\right)^2\right)}{n(k-1)}}. \quad (1)$$

Fig. 4 shows the TEMs for computer measurements (on a mannequin and on a human) and compares them to those obtained by trained human observers (single (Forest et al., 1999) and dual observer results (taken from Gordon and Bradtmiller, 1992)). Although the computer measurements contain an additional source of error due to automatic landmarking, the results indicate that the repeatability was similar to the single observer results for stature and neck circumference. The single observer results had the lowest TEMs for all other measurements, however, followed by computer measurements on a mannequin and on a human, followed by the measurements made by two observers.

The differences observed between mannequin and human repeatability results show the effect posture and breathing can have on measurements. Better precision could be obtained by controlling these factors, if

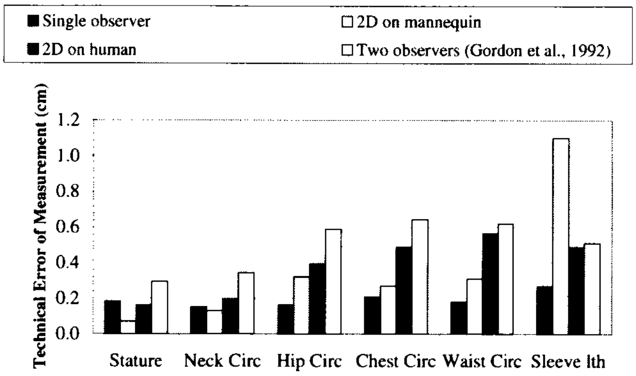

Fig. 4. Comparison of technical error of measurement (TEM) obtained by ICESS on a human subject and expert manual measurements.

Table 4
Reliability of a 2D image-based measurement system

	Reliability (%)
Stature	99.9
Neck circumference	99.3
Hip circumference	99.7
Chest circumference	99.6
Waist circumference	99.7

required. It should be noted that had the survey results included landmarking error (the survey subjects had the same landmarks during re-measurement), the results could have been more in favour of the computer measurements.

4.4. Reliability

Mueller and Martorell (1988) state that two pieces of information are sufficient to characterize the reliability of an anthropometric variable: the TEM and the reliability coefficient. The reliability coefficient (R) is an interesting metric in that it compares the variability due to measurement error (r^2) against the biological variability of that dimension (sample variance s^2). It is computed using the following equation:

$$R = 1 - (r^2/s^2), \qquad (2)$$

where r is the technical error of measurement, s is the sample standard deviation, n is the number of subjects and k is the number of measurements per subject.

If the measurement error is small compared to the standard deviation of the sample then the reliability of that measurement will be high. Reliabilities above 90–95% have been recommended for the selection of variables in a survey (Gordon and Bradtmiller, 1992). The reliability coefficients obtained by image-based measurement system were well above that, for the dimensions shown in Table 4.

Table 5
Manufacturing tolerances (on circumference) for Canadian Forces dress trousers and shirts

		Tolerance (mm)
Trouser	Waist circumference	± 13
	Inseam	± 13
Shirt	Neck circumference	± 3
	Chest circumference	± 13
	Sleeve length	± 13

4.5. Measurement accuracy requirements

The first part of the discussion dealt with the capabilities of the image-based measurement system when compared with skilled human measurement. But the answer to the question "how much measurement accuracy is required?" can only be answered in the context of the application. For clothing sizing, a large part of the answer comes from the manufacturing tolerances. In a sense, the manufacturing tolerances represent the limits of a trade-off between fit of the clientele and cost of the garment. They could be interpreted as an amount of fluctuation in garment size that has minimal impact on fit. By the same token, it could be said that given a garment size, the same amount of fluctuation in body measurement would also have minimal impact on the fit of a garment. An example of manufacturing tolerances for a military dress uniform is shown in Table 5. Using the above logic, Table 5 could be interpreted as an indication that a measurement error of ± 3 mm on neck circumference and ± 13 mm on the other dimensions would have minimal impact on fit, which suggest that a measurement accuracy within those ranges should be considered acceptable.

From a different perspective , it is also important to balance the accuracy against short-term body variations. These variations, which occur naturally, must be accommodated by the clothing regardless of their magnitude in order for the clothing to be acceptable. Several body dimensions can change substantially over short periods. Stature, for instance, has been known to change by 30–50 mm in a day depending on the amount of standing, walking and carrying done (NASA, 1978). Davenport et al. (1935) conducted experiments where repeated measurements of one subject were made at various times of day over a number of days by the same observer. The results (Table 6) show that measurements varied significantly, most notably for waist circumference, where 95% of the measurements were within ± 21 mm of the mean. It could be argued that the clothing worn on a day-to-day basis commonly accommodates the amount of variation reported by Davenport et al., and that those fluctuations should give an indication of the degree of accuracy required. Thus, it could be concluded that the highest

Table 6
Results of repeated measurements of a subject at various times of day over several days by one observer (Davenport et al., 1935)

	1.96 S.D. (mm)
Waist circumference	21
Chest circumference	15
Neck circumference	5

measurement accuracy would be required for neck circumference, i.e. 95% of the measurements should be within the range observed by Davenport, i.e. ± 5 mm.

5. Conclusions

The results of this analysis showed that image-based systems can provide anthropometric measurements that are quite comparable to traditional measurement methods performed by skilled anthropometrists, both in terms of accuracy and repeatability. The quality of the results depends, in large part, on the dependability of the automatic landmarking algorithms and the correct modelling of indirect measurements. Once that is achieved, however, this type of system can provide uniform measurement of a population regardless of where, when or by whom, it is operated.

While a high degree of accuracy and precision in anthropometric measurements is always desirable, it is not always necessary. The requirement for accuracy should be established on the basis of the application. Short-term body changes, clothing design, fit, and manufacturing tolerances were used as guides to estimate this requirement for the clothing sizing application. It was argued from these data that a body measurement system should be capable of measuring neck circumference within ± 5 mm in order to be effective, and all other dimensions within ± 15 mm. From the accuracy and precision analyses, it was found that the system under evaluation was capable of this performance.

References

Bennett, K., Osborne, R., 1986. Interobserver measurement reliability in anthropometry. Human Biology 58 (5), 751–759.

Chamberland, A., Carrier, R., Forest, F., Hachez, G., 1998. Anthropometric survey of the Land Forces (LF97) (Contractor report 98-CR-15). Defence and Civil Institute of Environmental Medicine, Toronto, Ontario.

Davenport, C., Steggerda, M., Drager, W., 1935. Critical examination of physical anthropometry on the living. Proceedings of the American Academy of Arts and Science Boston 69 (7), 265–285.

Forest, F., Chamberland, A., Billette, J., Meunier, P., 1999. Anthropometric survey of the Land Forces: measurement error analysis. DCIEM No. CR 1999-041, DCIEM, Toronto.

Gordon, C.C., Bradtmiller, B., 1992. Interobserver error in a large scale anthropometric survey. American Journal of Human Biology 4, 253–263.

Gordon, C.C., Churchill, T., Clauser, C.E., Bradtmiller, B., McConville, J.T., Tebbetts , I., Walker , R.A., 1989. 1988 anthropometric survey of US army personnel: methods and summary statistics. (NATICK/TR-89/044). US Army Natick Research, Development, and Engineering Center, Natick, MA.

Jamison, P., Zegura, S., 1974. A univariate and multivariate examination of measurement error in anthropometry. American Journal of Physical Anthropology 40, 197–204.

Kemper, H., Pieters, J., 1974. Comparative study of anthropometric measurements of the same subjects in two different institutes. American Journal of Physical Anthropology 40, 341–344.

Mueller, W., Martorell, R., 1988. Reliability and accuracy of measurement. In: Lohman, Rocha, Mortorell (Eds), Anthropometric Standardization Reference Manual. Human Kinetics, Champaign, IL, USA, pp. 83–86.

NASA, 1978. Anthropometric Source Book, Vol. I: Anthropometry for Designers (NASA reference publication 1024). NASA, Washington, DC.

Tsai, R., 1986. An efficient and accurate camera calibration technique for 3D machine vision. IEEE Conference on Computer Vision and Pattern Recognition, pp. 364–374.

#522068

CA024493