DEVELOPMENT OF THE BOUNDARY ZONE METHOD FOR GENERATION OF REPRESENTATIVE HUMAN MODELS

Kihyo Jung[1], Ochae Kwon[2], and Heecheon You[1]

[1]Division of Mechanical and Industrial Engineering, Pohang University of Science and Technology, Pohang, South Korea
[2]Design Team, Mobile Communication Division, Samsung Electronics, Seoul, South Korea

The present study developed a generation method of representative human models (RHMs) at a boundary zone which statistically accommodates a designated percentage of the target population. The *boundary zone* method proposed in the study consists of (1) identification of a boundary zone and (2) cluster analysis on cases within the identified boundary zone. The boundary zone of a designated accommodation percentage was formed by the normalized squared distance of each anthropometric case from the centroid of the target population. Cluster analysis was used to group homogenous cases within the boundary zone to reduce the number of the cases. A comprehensive evaluation under various combinations of anthropometric dimensions revealed that the average of multivariate accommodation percentages of the boundary zone method (91%) closer to the designated percentage (90%) than those of the existing generation methods (*square* method = 49%, *circular* method = 76%, and *rectangular* method = 96%).

INTRODUCTION

A small group of human models representing the target population is used for the ergonomic design and evaluation of products and workstations in a digital environment such as JACK® and RAMSIS®. Representative human models (RHMs) are a group of humans representing the body size characteristics of a designated percentage (e.g., 90%) of the target population (HFES 300, 2004). Use of RHMs provides designers with an efficient way to apply the body size characteristics of the target population to ergonomic design and evaluation (Jung et al., 2008). For example, You et al. (1997) evaluated the layout design of bus operator's workstation in JACK® using three RHMs.

Three generation methods (square, rectangular, and circular methods) have been developed which generate RHMs at the boundary of a designated accommodation percentage formed in the space of factors (or components). The square method proposed by Bittner et al. (1987) and Bittner (2000) generates RHMs at a square boundary formed in the space of factors extracted by factor analysis on the anthropometric dimensions under consideration. The rectangular method developed by Kim and Whang (1997) identifies RHMs at a rectangular boundary which is determined to statistically enclose a designated percentage of the population in the space of factors by reflecting their relative importance. Lastly, the circular method proposed by Meindl et al. (1993) generates RHMs at a circular boundary defined in the space of components extracted by principal component analysis.

The limitations of the existing generation methods using the data reduction techniques such as factor analysis have been discussed. The data reduction techniques can reduce the original set of anthropometric dimensions to a smaller set while most of the body size variability (e.g., 80%) is accounted. This information distillation process significantly simplifies the generation process of RHMs; however, some portion of the body size variability (e.g., 20%) is ignored in the generation process (Meunier, 1998). Furthermore, the conversion process from the values of RHMs defined in the space of factors to the sizes of anthropometric dimensions inflates estimation error if there are low correlations between the factors and anthropometric dimensions.

The present study developed a multi-variate generation method of RHMs at a boundary zone which statistically accommodates a designated percentage of the target population. The boundary zone method proposed in the present study formed a boundary zone for a designated accommodation percentage using the normalized squared distance from the centroid of the target population instead of the data reduction techniques. To compare the performances of the boundary zone and existing generation methods from each other, a comprehensive evaluation under various combinations of anthropometric dimensions was conducted in the study.

BOUNDARY ZONE METHOD

The present study proposed the boundary zone method consisting of (1) identification of a boundary zone and (2) cluster analysis for cases within the boundary zone. In the first step, the normalized squared distance of each anthropometric case from the centroid of the target population was calculated to identify the boundary zone accommodating a designated percentage of the target population. In the second step, cluster analysis was conducted for cases within the boundary zone to reduce the number of the cases by grouping cases with homogeneous characteristics.

Step 1: Identification of a Boundary Zone

To determine a boundary which accommodates a designated percentage of the target population, the values of anthropometric dimensions were converted into normalized squared distances by Equation 1. The normalized squared distance of normally distributed multivariate data follows Chi-square distribution with n (the number of variables) degrees of freedom (Johnson and Wichern, 1988). Hence, the boundary of a designated accommodation percentage can be formed by the corresponding Chi-square value. For example, for two anthropometric dimensions (n = 2), the normalized squared distances of 90% of the target population are those smaller than $\chi_2^2(1\text{-}0.9) = 4.61$.

$$D = (AD - \mu)^T \Sigma^{-1} (AD - \mu) \le \chi_n^2(1-p) \qquad (1)$$

where: D = normalized squared distance
AD = n-by-1 body size matrix
μ = n-by-1 average body size matrix
Σ = n-by-n covariance matrix
$\chi_n^2(1-p)$ = Chi-square value with n degree of freedom and p probability

The boundary zone which accommodates a designated percentage is formed by two boundaries as shown in Figure 1. The two boundaries are determined by a designated accommodation percentage plus and minus a tolerance (e.g., 90% ± 1%). Figure 1 illustrates the Chi-square values of two boundaries ($\chi_2^2(1\text{-}0.89) = 4.41$ and $\chi_2^2(1\text{-}0.91) = 4.81$) when the number of dimensions = 2, target percentage = 90%, and tolerance = 1%.

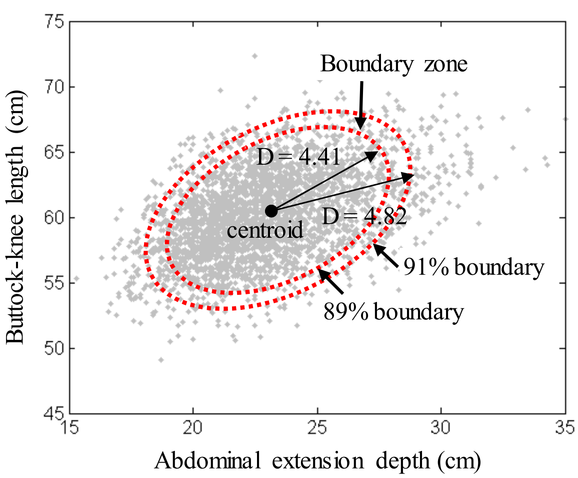
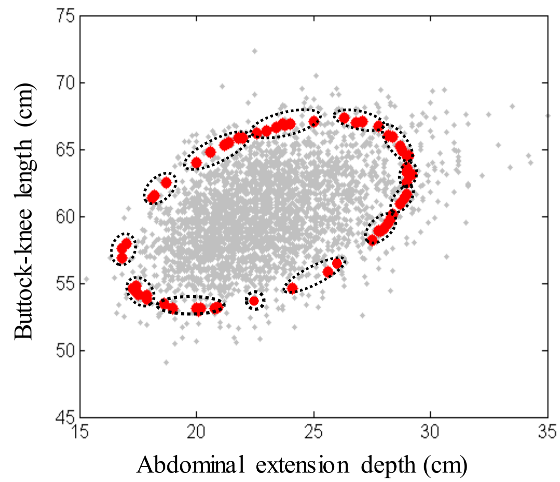


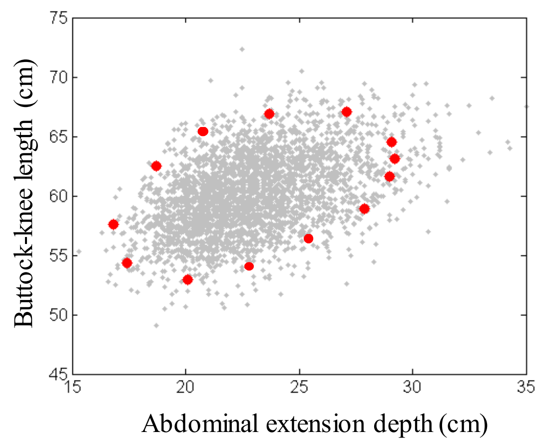
Figure 1. Illustration of a boundary zone formed by two boundaries

Step 2: Cluster Analysis in the Boundary Zone

To generate a small group of RHMs, K-means cluster analysis was applied to the cases within the boundary zone. Some of the cases within the boundary zone have similar body sizes; therefore, the homogeneous cases are grouped into clusters to reduce the number of RHMs as illustrated in Figure 2. An appropriate number of clusters can be determined by an in-depth analysis on multivariate accommodation percentage



(a) Homogeneous cases clustered



(b) RHMs generated by cluster analysis

Figure 2. Cluster analysis for the cases within a 90% boundary zone

(MAP) as the number of clusters increases. For example, in Figure 3 the optimal number of clusters can be selected as 34 by considering the target MAP value (90%) and MAP stabilization.

One case was selected from each cluster to construct an efficient group of RHMs. A RHM representing each cluster can be either the case nearest to the centroid or the centroid itself. To guarantee that the body size of the RHM is within the size range of the target population, the present study selected the nearest case to the centroid as the RHM representing the corresponding cluster.

EVALUATION METHODS

Anthropometric Data

The 1988 US Army anthropometric survey data (Gordon et al., 1988) was used in the present study to evaluate the generation methods of RHMs. The US Army anthropometric data provides a large, comprehensive database on 3,987 participants (female = 2,213, male = 1,774).

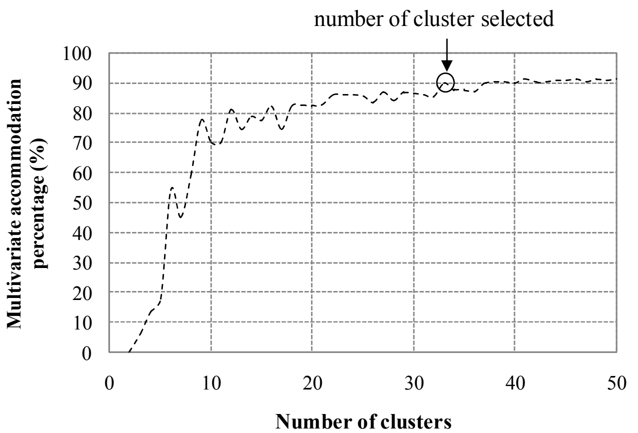


Figure 3. Determination of an optimal number of clusters by K-means cluster analysis (anthropometric dimensions = 10 and cases within the 90% boundary zone = 50)

To avoid an evaluation bias, the US Army anthropometric data was partitioned into a learning set and a testing set by the holdout validation method. When developing a prediction model, it is necessary to verify that the fitted model can be generalized to future data (Hawkins et al., 2003). Therefore, the present study randomly divided the US Army anthropometric data into a learning set (*n* = 2,982) and a testing set (*n* = 1,000) by referring the holdout validation method (Wikipedia, 2008; Blum et al., 1999).

Evaluation Conditions

To comprehensively evaluate the boundary zone and existing generation methods, various evaluation conditions (total = 21) were prepared in terms of the number of anthropometric dimensions and their combinations. The number of anthropometric dimensions considered was four levels (*n* = 5, 10, 15, and 20). For each number condition, five combinations of different anthropometric dimensions were considered. All evaluation conditions were prepared by random selection of anthropometric dimensions from the US Army data. In addition, a set of anthropometric dimensions (*n* = 10) which are applied to computer workstation design (ANSI/HFES, 2007) was considered.

Performance Measures

The RHM generation methods were evaluated in terms of three aspects: (1) multivariate accommodation percentage (MAP), (2) outlier proportion, and (3) the number of RHMs. First, MAP evaluates how much proportion of the target population is accommodated by a group of RHMs. Second, outlier proportion is the percentage of RHMs that are larger and/or smaller than the size ranges of the target population. Lastly, the number of RHMs evaluates the applicability of RHMs to ergonomic design and evaluation in a digital environment. The smaller the number of RHMs, the easier the application to ergonomic design and evaluation due to existing digital human modeling systems such as JACK® require significant amount of time and effort in creating humanoids and generating postures of the humanoids (Blome et al., 2006).

Evaluation Program

Matlab programs for each generation method were developed for efficient evaluation. The programs, first, prepared the evaluation conditions defined in the present study. Next, the programs generated groups of the RHMs accommodating 90% of the learning set as well as quantified MAP for the testing set. The algorithms for the existing three generation methods were developed by referring the generation processes specified in the previous studies (Bittner, 2000; Kim and Whang, 1997; Meindl et al., 1993).

EVALUATION RESULTS

Multivariate Accommodation

The MAP of the boundary zone method (91%) was quite close to the target percentage (90%) as shown in Figure 4. On the other hand, those of the square (49%) and circular (76%) methods were significantly lower than the target percentage (t (20) = -23.0, $p < 0.001$; t (20) = -8.6, $p < 0.001$). Conversely, the MAP of the rectangular method (96%) was significantly greater than the target percentage (t (20) = 9.8, $p < 0.001$).

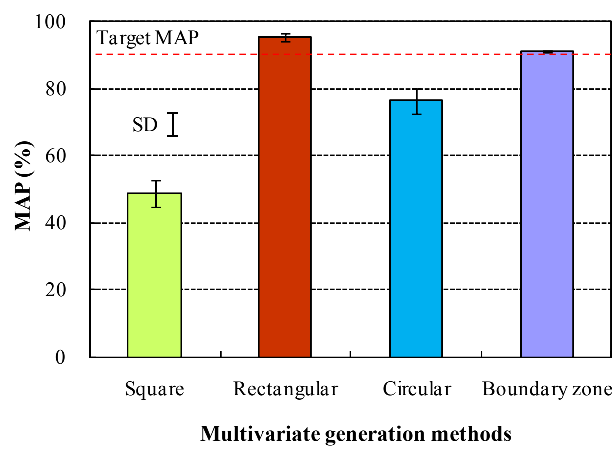
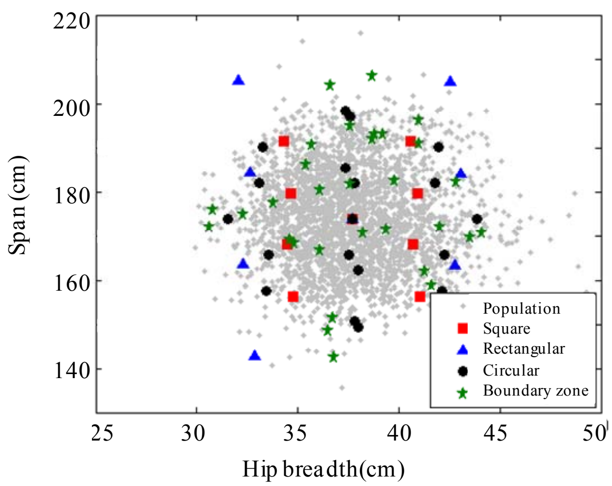


Figure 4. Multivariate accommodation percentages of the generation methods of RHMs

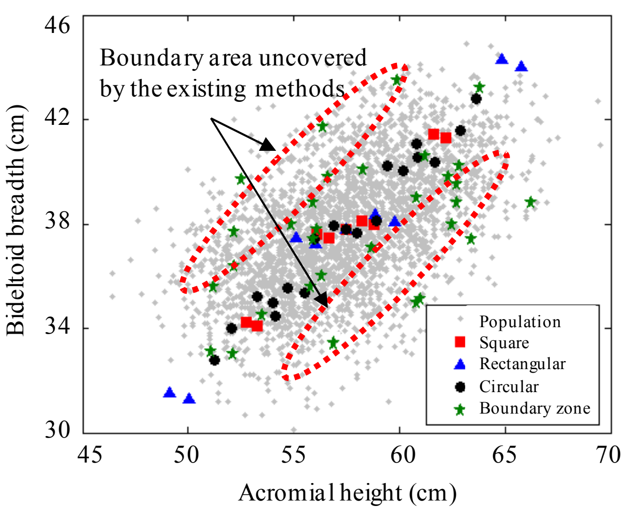
The variability of MAP of the boundary zone method was significantly lower than those of the existing generation methods. Figure 4 shows that the SD of the boundary zone method was 0.6%; On the other hand, the SDs of the square (8.2%), circular (7.3%), and rectangular (2.5%) methods were significantly greater than that of the boundary zone method (F (20, 20) = 169, $p < 0.001$ for square method; F (20, 20) = 15, $p < 0.001$ for circular method; F (20, 20) = 133, $p < 0.001$ for rectangular method).

A systematic analysis on the generated RHMs identified that the RHMs of the existing generation methods improperly represented the body size diversity of the target population.

Although the RHMs of the existing methods represented the body size diversity of the target population for the pairs of the anthropometric dimensions having different factor loading pattern as shown in Figure 5.a; the RHMs cannot properly represent the body size diversity for the pairs of the anthropometric dimensions having the similar factor loading pattern as shown in Figure 5.b.



(a) Pair of dimensions having different factor loading pattern



(b) Pair of dimensions having similar factor loading pattern

Figure 5. Bivariate distribution of RHMs

Outlier Proportion

Outlier analysis showed that the rectangular method generated RHMs that were smaller or larger than the size ranges of the target population. For 95% of the total evaluation conditions, the body sizes of at least one dimension were smaller or larger than the minimum or maximum of the body sizes of the target population. In addition, the outlier proportion of anthropometric dimensions in which the body sizes were apart from the range of the target population was strongly correlated ($r = 0.71$, $p < 0.001$) with the number of the factors extracted by factor analysis.

Number of RHMs

Figure 6 shows that the average number of RHMs of the boundary zone method was significantly larger than those of the existing generation methods (t (22) = -5, $p < 0.001$ for square and rectangular methods; t (22) = -2.6, $p = 0.02$ for circular method). In addition, the variability of the boundary zone method was significantly greater than the existing generation methods (F (20,20) = 15.6, $p < 0.001$ for square and rectangular methods; F (20,20) = 4.1, $p = 0.003$ for the circular method).

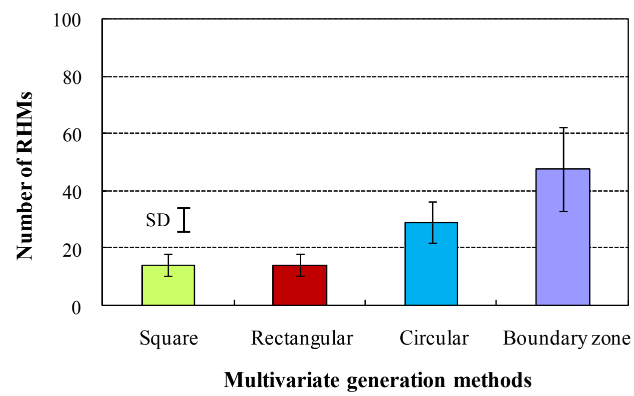


Figure 6. Average number of RHMs

The number of RHMs for all of the generation methods was found positively correlated with the number of anthropometric dimensions under consideration. The correlation coefficients of the square and rectangular methods ($r = 0.75$, $p < 0.001$) and circular method ($r = 0.77$, $p < 0.001$) were statistically significant and similar to each other. On the other hand, the coefficient of boundary zone method ($r = 0.94$, $p < 0.001$) was higher than those of the existing methods.

DISCUSSION

The present study developed a generation method of RHMs at a boundary zone which accommodates a designated percentage of the target population. To identify the boundaries surrounding a designated accommodation percentage, the boundary zone method used the normalized squared distance instead of the data reduction techniques, which have been known as the source of multivariate accommodation problem (Meunier, 1998). Thus, RHMs generated by the boundary zone method more appropriately represent the body size characteristics of a designated percentage of the target population.

The evaluation results indicate that the existing generation methods should be carefully used when the combinations of anthropometric dimensions are applied in ergonomic design and evaluation. The present study found that the RHMs generated by the existing methods cannot appropriately represent the size diversity of the target population for the pairs of anthropometric dimensions having similar factor loading patterns. It is noticeed that the body sizes of the target population cannot be appropriately applied to ergonomic design and evaluation if the combinations of anthropometric dimensions are simultaneously considered.

The boundary zone method requires a tolerance percentage (e.g., ±1 %) to form a boundary zone. If the tolerance percentage is too small, cases within the boundary zone are too small to represent the target population. Conversely, if the tolerance percentage is too large, unnecessary cases which are far apart from the boundary are included for the candidates of RHMs. Therefore, an appropriate level of tolerance should be decided by considering various technical aspects such as the size of anthropometric database and the number of anthropometric dimensions under consideration. The larger the anthropometric database, the smaller the tolerance required; the opposite becomes true for the size of database and the number of anthropometric dimensions.

To statistically identify the boundary, the boundary zone method assumes that anthropometric dimensions follow a multivariate normal distribution. HFES 300 (2003) stated that the distributions of body sizes follow normal distributions and Roebuck et al. (1975) indicated that the body sizes related to the length of bones statistically follow normal distribution. Conversely, Vasu and Mital (2000) reported that significant normality violations could exist with some anthropometric variables. On the other hand, in the present study, the difference between the designated percentage and the proportion of the cases within a boundary showed about 1% anthropometric dimensions violate the normality assumption. However, an in-depth validation research is required to check the validity of the assumption.

The boundary zone method generates a more number of RHMs than the existing generation methods. The more the number of RHMs generated, the higher the difficulty to apply the RHMs to ergonomic design and evaluation in a digital environment such as Jack® and RAMSIS®. The evaluation results showed that the boundary zone method requires a more number of RHMs to statistically accommodate a designated percentage of the target population than the existing methods. In addition, the number of RHMs required is rapidly increased as the number of anthropometric dimensions is increased. It might limit the applicability of the boundary zone method in a digital environment. However, this application limitation of the boundary zone method can be overcome if functional relationships between anthropometric dimensions and design dimensions are established in the design process, which has been under study as continuation of the present study.

REFERENCES

ANSI/HFES 100 (2007). *Human Factors Engineering of Computer*. Santa Monica, CA: Human Factors and Ergonomics Society.

Bittner, A.C., Glenn, F.A., Harris, R.M., Iavecchia, H.P. Wherry, R.J. (1987). CADRE: A family of manikins for workstation design. In S.S. Asfour (Ed.) *Trends in Ergonomics/Human Factors IV* (pp. 733-740). North Holland: Elsevier Science Publishers B.V.

Bittner, A. C. (2000). A-CADRE: Advanced family of manikins for workstation design. *Proceedings of the IEA 2000/HFES 2000 Congress*. San Diego. CA, 774-777.

Blum, A., Kalai, K., and Langford, J. (1999). Beating the hold-out: bounds for K-fold and progressive cross-validation. *Proceedings of the 25th Annual Conference on Computational Learning Theory*. 203-208.

Blome, M., Hanson, L., Lamkull, D., Nielsen, E., and Stam, E. (2006). Identifying time-consuming human modeling tool activities. *Proceedings of Digital Human Modeling for Design and Engineering Conference*, France

Chaffin, D. B. (2001). *Digital Human Modeling for Vehicle and Workplace Design*. Warrendale, Pa: SAE.

Gordon, C. C., Bradtmiller, B., Churchill, T., Clauser, C., McConville, J., Tebbetts, I., and Walker, R. (1988). 1988 Anthropometric survey of U.S. Army Personnel [Data file]. Available from National Technical Information Service Web site, http://www.ntis.gov.

Hawkins, D. M., Basak, S. C., and Mills, D. (2003). Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences*, 43, 579-586.

HFES 300 (2004). *Guidelines for Using Anthropometric Data in Product Design*. Santa Monica, CA: Human Factors and Ergonomics Society.

Johnson, R. A. and Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis* (2nd ed.). New Jersey: Prentice Hall.

Jung, K., You, H., and Kwon, O. (2008). A multivariate evaluation method for representative human model generation methods: Application to grid method. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1665-1669). Santa Monica, CA: The Human Factors and Ergonomics Society.

Kim, J., and Whang, M. (1997). Development of a set of Korean manikins. *Applied Ergonomics*, 28, 407-410.

Meindl, R. S., Hudson, J. A., and Zehner, G. F. (1993). *A Multivariate Anthropometric Method for Crew Station Design* (AL-TR-1993-0054): Armstrong laboratory, Wright-Patterson. (1993).

Meunier, P. (1998). Effects of a data reduction technique on anthropometric accommodation. *Proceedings of the 42nd Human Factors and Ergonomics Society*. 727-731.

Porter, J. M., Case, K., Marshall, R., Gyi, D., and Oliver, R. S. (2004). Beyond Jack and Jill: designing for individuals using HADRIAN. *International Journal of Industrial Ergonomics*, 33, 249-264.

Roebuck, J. A., Kroemer, K. H. E., and Thomson, W. G. (1975). *Engineering Anthropometry Methods*. New York: Wiley-Interscience.

Wikipedia (2008). Definition of holdout validation. Retrieved from www.wikipedia.com.

You, H., Bucciaglia, J., Lowe, B. D., Gilmore, B. J., and Freivalds, A. (1997). An ergonomic design process for a US transit bus operator workstation. *Heavy Vehicle Systems, A Series of the International Journal of Vehicle Design*, 4, 91-107.