Research on python-based regression model of female body measurement

Chang Ma, Zhengdong Liu*, Boxiang Xiao
Fashion Design and Engineering, Beijing Institute of Fashion Technology, Beijing, China

ABSTRACT

In accordance with the problems of body measurement technology and difficulty in obtaining body samples for common consumers in large-scale customization of clothing, simulation body samples and the machine learning method with positive images were proposed to be adopted to predict the body data. Firstly, simulated shooting of positive images from different angles was conducted on the basis of 116 simulation body samples. In addition, characteristic parameters were obtained through extracting the body silhouette and identifying key points and multiple linear regression and neural networks were used to conduct the machine learning of chest circumference, waistline and hipline, with the prediction model established. The experiment results showed that the error mean of two models in the sample set was less than 3cm and two models were equipped with the certain insensitivity to shooting angles, which may be promoted to clothing size customization of online large-scale customization of clothing.

Keywords: Anthropometry, multiple linear regression, MLP neural network

1. INTRODUCTION

Clothing fitness is closely relevant to the body data. Conventional body measurement method cannot meet the needs of digital clothing customization with the body measurement as the basis of clothing design and production, while the three-dimensional body scanning method has a low use rate in clothing production on the basis of big volume and expensive price[1]. Nevertheless, the two-dimensional non-contact body measurement technology is featured by high measurement efficiency and low price, with image acquisition, silhouette extraction, feature point recognition and circumference fitting as the basic thoughts. In addition, it is only needed to take some pictures for obtaining the body size with easy operation[2], fully indicating that the feasibility of such a measurement method in the clothing field. Currently, the existing two-dimensional image measurement method has achieved better effect in chest circumference, waistline and hipline and other sizes[3, 4]. However, collecting the real body size data from various angles is required and an easier measurement method of the chest circumference, waistline and hipline of female body was proposed for this. The chest circumference, waistline and hipline of the measured object may be measured through the picture based on the only need of a positive image and the height and weight of the measured object.

2. PARAMETRIC MEASUREMENTS

2.1. Experimental subjects

It can be found that most experimenters conduct the experiment through collecting a certain number of real body data on the basis of researching the previous two-dimensional non-contact body size prediction experiments. However, the problems of low sample coverage, high measurement cost and shortage of samples exist in terms of using the real body as the experiment object[5, 6], so that Make Human was considered to be used for manufacturing simulation body samples in the experiment. Make Human is the 3D character modeling software on the basis of numerous anthropological morphological characteristic data and face and limb models of male and female at different ages may be rapidly formed and accurately simulated.

The body model material was arranged as an Asian female in the experiment and the simulation body samples with body types of Y, A, B and C were randomly generated in accordance. Besides, simulation bodies of various body types were established through the careful modification of various parts of the simulation body, including slipped shoulder, square shoulder, hunchback, straight waist, abdominal retraction, abdominal bulge and so on. Finally, the data of various parts

*jsjlzd@bift.edu.cn

International Conference on Computer Application and Information Security (ICCAIS 2021),
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12260, 122601P · doi: 10.1117/12.2637633

of the simulation body were exported. In the research, simulation bodies of 116 Asian females with the height between 150.3cm and 180.4cm, the weight between 40.72kg and 73.91kg and the age between 16 and 73 were the experiment objects. The basic size information of the experiment objects is shown in Table 1.

Table 1. Basic size information of experiment objects.

Parameter (cm)	**Height**	**Weight**	**Hipline**	**Chest circumference**	**Waistline**	**Back length**	**Shoulder breadth**
Minimum	150.36	40.72	79.05	73.68	62.71	34.42	10.24
Maximum	180.42	73.91	116.18	108.09	100.78	42.71	19.28
Mean	165.51	56.28	96.79	90.57	77.16	38.22	13.27
Standard deviation	8.49	7.96	7.32	5.59	8.75	1.68	1.04

2.2. Measurement part determination

In the paper, the most widely available 5 body measurement sizes were adopted under default in terms of the body measurement data, namely the height, weight, chest circumference, waistline and hipline. The simulation body was arranged to stand naturally, look straight ahead, keep feet about shoulder-width apart and open hands diagonally downward and naturally at 30° to measure the height, weight, chest circumference, waistline and hipline of the simulation body.

2.3. Two-dimensional image processing

The positive image was obtained for the simulation body sample and the Make Human software was called to open the simulation body model. The simulation body was shot positively through the virtual camera simulating the real scenes and the camera was arranged to be 1.5m above the ground with 2.5m-3.5m from the shot object. The whole simulation body was always kept in the viewfinder and the simulation body with bare feet stood in the center point of viewfinder. In addition, the posture of the simulation body should be consistent with the measurement posture and 5 positive images with the resolution ratio of 800×600 should be shot from different angles for the simulation body for the purpose of practicing the insensitivity to the shooting angles. Moreover, hands cannot cover the body in the shooting process in order to avoid the difficulty in obtaining data in the subsequent experiment process. Totally 580 positive images of the simulation body were obtained, which are shown in Figure 1a.

Binary image conversion was conducted for the positive images in order to more intuitively identify the body silhouette. Each pixel corresponds to a gray value in the positive image and the threshold function (threshold) was applied to the positive image to compare the gray value of the positive image with the threshold[7]. In addition, the simulation body was arranged to be black with the white background, thus achieving the binarization processing as shown in Figure 1b. Then the target silhouette was extracted through the edge detection. With local transition appearing in the gray values of two adjacent regions, the image edge appeared within two adjacent regions. Moreover, the boundary between the two regions may be obtained through the detection of the nearby edge points and the edge in the direction may be modeled through the location parameters on the directional profile. Besides, the first derivative and the second derivative were adopted to detect the edge in terms of the gray changes in the edge location[8].

Baidu AI body key point recognition technology was used to generate the key point figure, which is shown in Figure 1c. 21 main key points of the simulation body in the positive image may be obtained by entering the positive image, containing the top of the head, five sense organs, the neck, limbs and other parts and the coordinate information of 5 key points may be extracted, such as the neck, the left shoulder, the right shoulder, the left hip and the right hip.

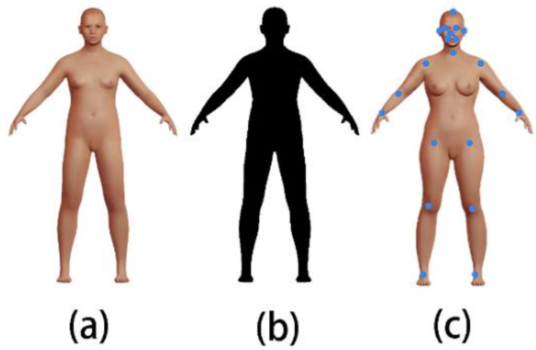

Figure 1. Schematic diagram of image processing.

2.4. Two-dimensional size extraction

The relevant two-dimensional image feature points (A: the foreneck point, B: the left lateral chest point, D: the right lateral chest point, E: the left lateral waist point, G: the right lateral waist point, H: the left lateral hip point and J: the right lateral hip point) should be determined in accordance with the obtained 5 body measurement sizes. The computer graphics knowledge was used to obtain the coordinate of various feature points, thus obtaining the pixel distance of various feature sizes, AC: neck-chest length, CI: chest -hip length, BD: lateral chest space, EG: lateral waist space and HJ: lateral hip space. In addition, the size extraction parts are shown in Figure 2.

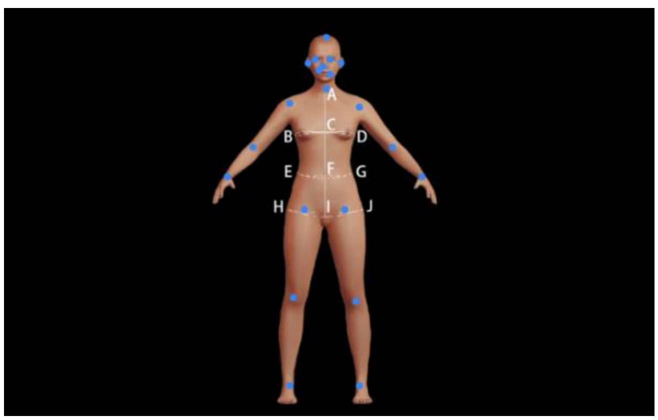

Figure 2. Schematic diagram of two-dimensional size extraction.

The coordinate information of A can be obtained directly through using the key point figure in accordance with Figure 2 and H and J can be found through the intersections of key points of the left and right hips in the key point figure respectively extending outwards to the silhouette edge. H and J were used to search along the silhouette edge from bottom to top until key points of left and right shoulders. The locations with the lowest horizontal distance from A were E and G. In addition, the locations with the maximum horizontal distance from A were B and D by searching along the silhouette edge from bottom to top and from E and G to key points of left and right shoulders. C, F and I were respectively the vertical intersections from A to BD, EG and HJ. Moreover, the feature sizes of 6 parts may be obtained by using the formula for distance between two points in accordance with the above principles. In addition, the real sizes of the body chest circumference, waistline and hipline in the image may be obtained through finding the conversion relationship between the image extraction size and the real size.

2.5. Data pre-processing

The errors must exist in the process of obtaining the data and the pre-processing of the data should be conducted for the purpose of improving the accuracy of the data. SPSS software was used to process the missing value and find the abnormal value for correction of the original data. 580 effective samples were determined and the data were numbered. Then 485 groups of data were selected randomly to be used for the model training and the left 95 groups of data were used for the model verification.

3. MODEL TRAINING

The height, weight, lateral chest space, lateral waist space and lateral hip space, neck-chest length and chest-hip length were used as the input vector and the chest circumference, waistline and hipline were used as the output vector in the experiment. The data used for the model training were randomly divided into the training set and the test set with the data in the training set for the training model and the data in the test set for the accuracy of the model after the assessment training. The training and test samples in the Paper were divided into four parts in accordance with the ratio of 80% and 20%: the partial input training samples, the partial input test samples, the partial output training samples and the partial output test samples. The linear regression model and MLP neural network model will be established respectively in the following content.

3.1. Establishment of the linear regression model

A linear regression model is a type of regression analysis that models the relationship between one or more independent and dependent variables using the least square function of a linear regression equation[9]. The linear regression model is featured by the advantages of simple calculation and the results easy to understand and the establishment steps of the linear regression model are as follows:

Linear Regression() in the machine learning function library was used to establish the linear regression model and the regression analysis was conducted for the training sample data. fit() was used to analyze the linear regression model parameters and predict () as the model established through the linear regression model parameters calculated by fit() was used to predict the input variables and obtain their values. Then the predict function was used to predict the chest circumference, waistline and hipline and the comprehensive analysis was conducted for their predicted values to determine the final predicted values. Then the model. intercept function was used to calculate the intercept and the model. coef function was used to calculate the coefficients of each parameter. The multiple linear regression equations of obtained chest circumference y_0, waistline y_1 and hipline y_2 are shown in Table 2.

Table 2. Multiple linear regression equations.

	Regression formula
Chest circumference	$y_0 = -0.02x_0 + 0.51x_1 - 0.14x_2 - 0.03x_3 - 0.04x_4 - 0.28x_5 - 0.56x_6 + 62.53$
Waistline	$y_0 = -0.02x_0 + 0.51x_1 - 0.14x_2 - 0.03x_3 - 0.04x_4 - 0.28x_5 - 0.56x_6 + 62.53$
Hipline	$y_2 = -0.37x_0 + 0.09x_1 - 0.36x_2 + 0.38x_3 - 0.17x_4 - 0.09x_5 + 1.06x_6 + 96.24$

2.2. Establishment of MLP neural network model

As a type of artificial neural network model, multilayer perceptron model (MLP) can describe a group of complex mapping relationship[10] from input variables to output variables through the nonlinear basis function. The establishment steps of MLP neural network model are as follows:

MLP Regressor() in the machine learning function library was called to conduct MLP neural network modeling for the chest circumference, waistline and hipline. Any one neuron at each layer of MLP will generate a connection coefficient with any one neuron at next layer and the output value of the output layer will be calculated in accordance with the connection coefficient after the input of the input layer, namely the predicted values of the chest circumference, waistline and hipline. In addition, they will be compared with the real values. The weight of various connection coefficients was adjusted with the times of repetition being the maximum number of the allowed iterations to obtain the neural network model meeting the requirements and then verify the model finally. In the paper, the arranged maximum iteration times were 10000 and the number of neurons of the input layer was 7, namely the number of the input vectors. Besides, the hidden layers were 2 with the number of neurons at the first hidden layer of 100 and the number of neurons at the second hidden layer of 50 and the number of the neurons of the output layer was 3, namely the number of the output variables. The basic structure of MLP neural network model is shown in Figure 3.

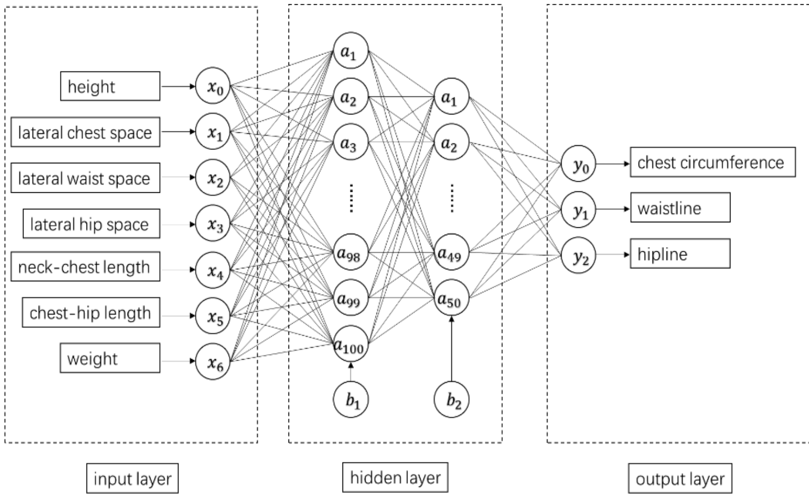

Figure 3. Basic structure figure of MLP neural network model.

2.3. Model evaluation

The index of R^2 value was adopted to evaluate the prediction accuracy and the degree of correlation between input vector and output vector of the above two models in the Paper. As a type of measurement, R2 value describes the difference between the predicted value and the actual value. In addition, when it is closer to 1, the prediction accuracy and the correlation are higher. The formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i^T - y_i')^2}{\sum_{i=1}^{N}(y_i^T - \bar{y}^T)^2}$$

y_i^T refers to the actual value and y_i' refers to the predicted value with $\bar{y}^T$ indicating the mean in the formula. The specific numerical values were obtained through calculation and the results of the verified model are shown in tables 3 and 4.

Table 3. Evaluation results of linear regression model.

Parameter	**R^2**	**Average error (cm)**	**Standard variance (cm)**	**Maximum error (cm)**	**Minimum error (cm)**
Chest circumference	0.88	1.52	1.16	5.42	0.03
Waistline	0.91	2.11	1.76	8.9	0.13
Back length	0.91	1.74	1.5	7.89	0.00

Table 4. Evaluation results of MLP neural network model.

Parameter	**R^2**	**Average error (cm)**	**Standard variance (cm)**	**Maximum error (cm)**	**Minimum error (cm)**
Chest circumference	0.81	1.93	1.61	7.00	0.04
Waistline	0.87	2.35	2.41	14.47	0.01
Back length	0.83	2.33	2.05	9.31	0.06

In accordance with Tables 3 and 4, the correlation coefficients of the chest circumference, waistline and hipline in the linear regression model and MLP neural network model all exceed 0.8 and the average error, standard variance, maximum error and minimum error of the linear regression model are all less than those of the MLP neural network model. As a consequence, the prediction accuracy of the linear regression model is better than that of the MLP neural network model and the prediction accuracy difference of the two models is minor.

4. MODEL VERIFICATION

It is required to verify the two models on whether the established linear regression model and MLP neural network model can be put into practical application. The left 95 groups of data after model training were selected to conduct the model verification and the data were respectively put into the model to view the predicted results. If the error between the predicted value and the actual value is minor, the models can be applied to the reality. The obtained error statistics of the chest circumference, waistline and hipline of the two models through calculation is shown in Tables 5 and 6.

Table 5. Error statistics of linear regression model.

Parameter (cm)	Minimum	Maximum	Mean	Standard deviation	Variance
Error of chest circumference	0.00	5.70	1.95	1.40	2.00
Error of waistline	0.00	6.10	2.19	1.74	3.03
Error of hipline	0.00	5.40	1.91	1.47	2.17

Table 6. Error statistics of MLP neural network model.

Parameter (cm)	Minimum	Maximum	Mean	Standard deviation	Variance
Error of chest circumference	0.10	6.60	2.11	1.52	2.31
Error of waistline	0.00	9.30	2.5	2.20	4.85
Error of hipline	0.10	8.70	2.8	2.21	4.89

In accordance with Tables 5 and 6, each parameter in Error Statistics of Linear Regression Model is less than error statistics of MLP neural network model, indicating the better fitting effect of the linear regression model. In addition, the error means of the chest circumference, waistline and hipline of the two models are all less than 3cm, indicating that both two models can be applied to the practice.

5. CONCLUSION

In this paper, the chest circumference, waistline and hipline of the body were predicted through the machine learning, establishment of linear regression model and MLP neural network model and measurement of characteristic parameters of height, weight, neck-chest length, waist-hip length, lateral chest space, lateral waist space and lateral hip space in terms of female body type recognition. Based on the evaluation and verification of the two models, the following conclusions may be obtained:

(1) The assessment effect of the linear regression model is better than that of the MLP neural network model with R^2 values of the chest circumference, waistline and hipline in the linear regression model respectively being 0.88, 0.91 and 0.91 and R^2 values of the chest circumference, waistline and hipline in the MLP neural network model respectively being 0.81, 0.87 and 0.83.

(2) In terms of the two models, the average relative error and standard error of the predicted value and the actual value are relatively small and more accurate prediction can be made for the body type recognition, providing certain reference values. More training samples will cause better assessment effect in terms of the neural network model. In addition, if more sample data are obtained for training, the accuracy of the MLP model may be enhanced.

ACKNOWLEDGMENTS

This paper is sponsored by Innovation Team Development Project of BIFT, Undergraduate teaching reform and innovation project of Beijing Higher Education and 2021 Graduate Research Innovation Project of BIFT (X2021-010).

REFERENCES

[1] Wang, T. and Gu, B. F., "Construction of 3D model of human neck and shoulder based on image," Journal of Textile Research, 42(01), 125-132 (2021). (in Chinese)

[2] Liu, X. Y. and Xie, H., "Research on non-contact anthropometry technology based on two-dimensional image," Knitting Industries, 5, 58-62 (2021). (in Chinese)

[3] Zhu, F. F. and Luo, R. L., "Human body measurement system based on incomplete data," Journal of Zhejiang Sci-Tech University, 28(05), 710-713 (2011). (in Chinese)

[4] Xia, M., Wang, W. L. and Zhang, W. B., "Shape fitting and girth size prediction of human feature section," Journal of Textile Science and Technology, 35(08), 76-80 (2014).

[5] Li, X. J. and Wang, Y. X., "Calculation of hip circumference in two-dimensional non-contact anthropometric system," Journal of Textile Research, 42(02), 98-100 (2004). (in Chinese)

[6] Ji, K. C., [Improvement of Circumference Fitting Based on Two-Dimensional Contactless Measurement System for Young Male Body], Soochow University, Suzhou, Master's Thesis, 53 (2014). (in Chinese)

[7] Wang, T., "Graphic recognition based on OpenCV in Android mobile terminal," Practical Electronics, 23, 64-65 (2020). (in Chinese)

[8] Ma, L., [Research and Implementation of Image-Based Contactless Anthropometry Algorithm], University of Electronic Science and Technology of China, Chengdu, Master's Thesis, 33 (2017). (in Chinese)

[9] Qiao, L. L., "Application of multiple linear regression prediction method in garment manufacturing," Journal of Yangtze University (Natural Science Edition), 7(01), 145-147 (2010). (in Chinese)

[10] Liu, S. and Wang, H. Y., "Modulation recognition of MLP Neural Network based on hybrid genetic algorithm optimization," Journal of Wuhan University (Natural Science), (01), 104-108 (2008). (in Chinese)