

## Research Article

# Application of Garment Customization System Based on AR Somatosensory Interactive Recognition Imaging Technology

Wenjuan Tao 

*Guangdong Polytechnic, Foshan, Guangdong 528000, China*

Correspondence should be addressed to Wenjuan Tao; 2013041119@stu.zjhu.edu.cn

Received 15 February 2022; Revised 18 March 2022; Accepted 30 March 2022; Published 15 April 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Wenjuan Tao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to realize the popularization of customized clothing and augmented reality technology (AR), a design scheme of clothing customization system based on AR somatosensory interactive recognition imaging technology is proposed and studied. Using software such as 3ds Max and Clo3d, the modeling of the human body model and the clothing model is realized, and the clothing deformation algorithm based on the input posture of the human body is applied to deform the clothing model, which presents the physical characteristics of the virtual clothing more realistically. At the same time, the image acquisition equipment is used to obtain the real image, and the clothing image transfer algorithm is used to superimpose the virtual clothes and the real image. The position of the joint points of the human body is obtained through the feature extraction technology, and the scaling ratio of the virtual clothing is calculated, so that the virtual clothes and the real image are closely combined. The 3D human modeling technology can obtain all the data from the human body in 10 to 30 seconds and create the most suitable clothes according to the parameters, which can be applied to different fields such as tailoring, personalized clothing customization, and 3D fitting, combined with the G-ERP system. The integration of online body measurement, clothing design, plate making and production is realized.

## 1. Introduction

AR is a new human-computer interaction technology rising in recent years. Through the comprehensive application of various technologies such as object target recognition, dynamic tracking, and virtual reality, it can truly realize the “seamless” integration of virtual information and real environment, so as to enhance users’ perception of the real world as shown in Figure 1. With the rapid development of the information age, the proportion of image information in the field of textile and garment is increasing, and it has become an important tool to obtain and transmit information [1]. People’s production and life style have also changed in the process of textile and garment production, and physical images are transformed into required information by acquiring and processing images, so as to provide information input for automatic production. In the process of display and sales of textiles and clothing as commodities, image retrieval, recognition, labeling, and other technologies with textiles and clothing as the content

provide convenience for users to screen commodity information, as well as personalized recommendation. The technology used is the image recognition technology in the field of image engineering and artificial intelligence. Its purpose is to let the computer process a large amount of physical information instead of human beings, process and analyze the image, and finally identify the target to be studied [2]. The research progress of image recognition technology is closely related to the theories and technologies of computer graphics, computer vision, machine learning, and fuzzy logic. Its application scope involves the fields of medicine, biology, industrial automation, public safety, and so on. Ar somatosensory interactive imaging technology is powerful. It provides automatic data generation imaging through three-dimensional human body digital scanning technology, so that the scanned files can be controlled in size and precision according to your needs. This perfect 3D human body scanning and imaging software guides users to edit, save, and reuse scanning results through its simple user interface. At the same time, it can

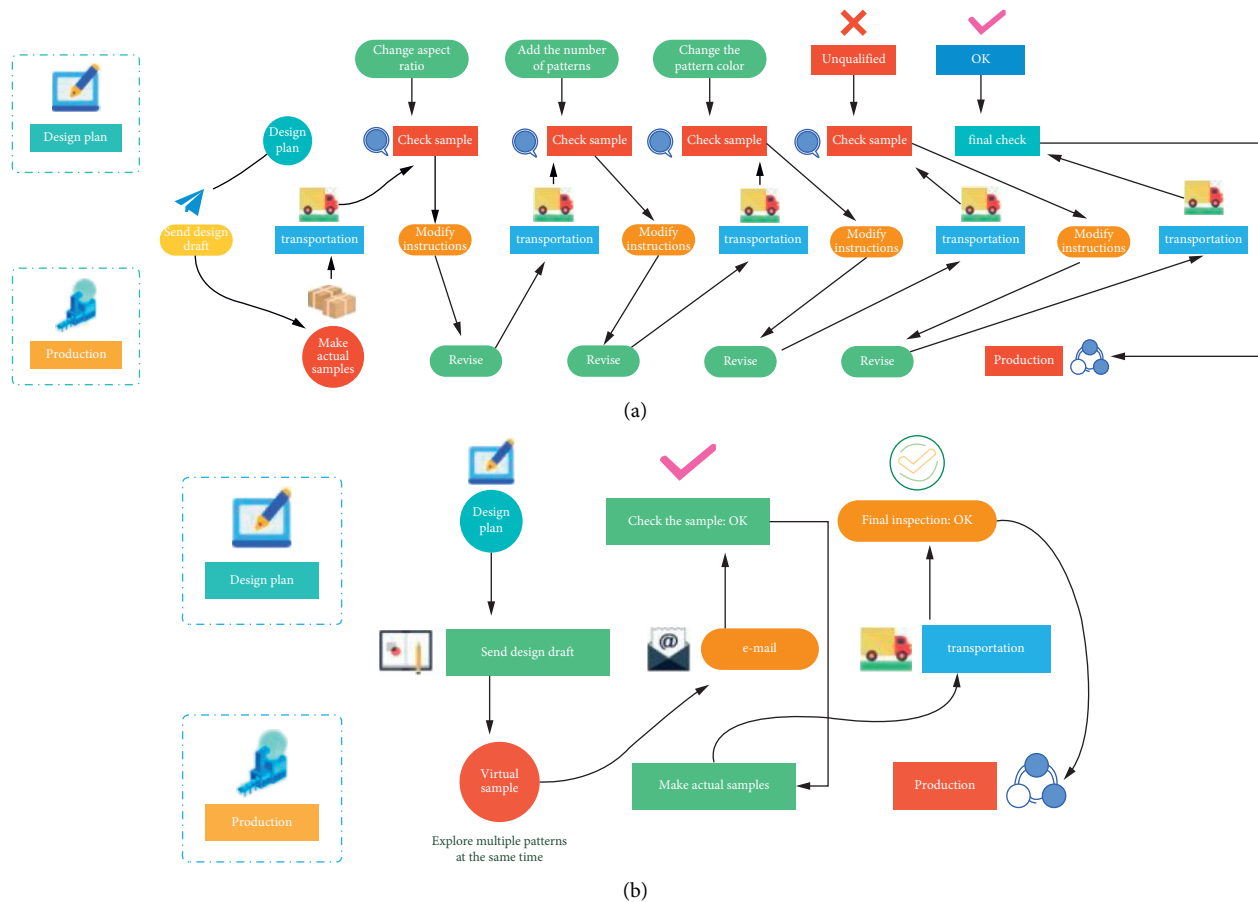


FIGURE 1: AR flow chart.

also be built into multiple garment design CAD software to make it easier for data processing [3].

## 2. Literature Review

Li and others said that in recent years, the research of virtual fitting system based on augmented reality technology and somatosensory interaction technology has attracted extensive attention [4]. Jiang et al. believe that the advantage of virtual fitting over traditional fitting lies in its convenient and fast interactive fitting mode, which brings users a new shopping experience and forms a free and fashionable shopping method [5]. Chiang et al. said that the textile and garment industry in the 21st century has entered a new stage, and more and more people like private customization. Masscustomized clothes will bring many problems, which are not suitable for all people, such as some people with very wide shoulders and some people with very large beer belly. It is difficult for such people to choose their own clothes in the mall [6]. King et al. believe that we urgently need a low-cost textile and garment customization system suitable for different people manufactured by scientific and technological means [7]. Wang et al. said that by using these advanced production equipment and high and new technologies, garment production enterprises can accelerate garment order undertaking and production, enhance R&D capacity

of sample garment, enrich product design, accelerate new product listing, shorten production and processing cycle, improve product quality, reduce production cost, improve production efficiency, and reduce their dependence on artificial technology. Thus, the market competitiveness of enterprises is greatly enhanced and the profit maximization is created [8]. Nemcokova et al. put forward three different development directions for the presentation of clothes: one is to express all clothes models and mannequins in the form of animation based on 3D special effect animation. When users try on, they can choose a model close to their own figure to replace their own fitting, and users can see the fitting effect on the fitting screen at 360 degrees. The other is to make 2D clothing pictures of clothes first and then use somatosensory technology. When users try on clothes, they choose to place the pictures on the human body, which belongs to a mapping method. The third is the photographing and fitting system. Users first select the clothes they want to try on from the existing clothes, then take pictures of themselves and clothes, respectively, and then synthesize them in the background to complete the fitting experience [9]. Zhu et al. said that the first method brings users a comprehensive fitting effect experience, but this method is not for users to try on themselves. The model used can only ensure that the user and the model are similar in size, and the user has a poor sense of fitting immersion [10]. Jones et al.

believe that the second method is to display clothes on the plane. Users can only obtain the three-dimensional effect of clothes on the body by imagination, so the display effect is not intuitive [11]. Moon et al. believe that the third one is also based on plane display, and the operation is complex, which affects the fitting effect of users [12]. In addition, Zeng et al. said that these three technologies only show the appearance of clothes but cannot be well represented for factors such as clothing fabric, texture, and drape effect and cannot automatically change the scaling coefficient of clothes according to the user's body shape characteristics, so the experience effect is relatively poor [13]. The virtual fitting system is to use the principle of computer imaging, use the created three-dimensional model of the user's human body to perform virtual fitting on the clothing selected by the user, complete the superimposition of the clothing picture onto the human body model, and display the wearing effect of the user's virtual fitting.

### 3. Method

#### 3.1. Introduction to Relevant Technical Principles of Virtual Fitting System

*3.1.1. Speech Recognition Technology.* The language recognition technology is used to design the system. Speech recognition technology means that the machine can use relevant technologies to recognize speech and convert it into information that the machine can "understand." Its main research goal is to make the computer understand the voice information sent by human beings and make correct response. The speech recognition technology is one of the key technologies in the fields of speech communication system, banking service, computer control, industrial control, data query, ticket booking system, artificial intelligence, security systems and so on. The main problem to be solved by the automatic speech recognition (ASR) technology is to enable the machine to obtain the important information in human voice and recognize it correctly. In order to enable the computer equipment to recognize human language, the ASR technology is essential for the speech recognition system of intelligent computer. The ASR technology is like bringing binaural to the computer system [14]. Although speech recognition technology has a broad prospect, there are still many problems to be solved. As early as the 1950s, speech recognition researchers in various countries have invested great energy in this research. Especially in the last two decades, many enterprises and research departments at home and abroad have joined the research field of speech recognition function technology, made great efforts, and achieved good research results. Although there are still many problems in speech recognition technology until today, it does not prevent speech recognition technology from gradually entering people's daily life [15]. Nowadays, speech recognition technology has been widely used in acoustics, digital signal processing, linguistics, statistical pattern recognition, and many other disciplines. Modern speech recognition systems have been successfully applied in many scenarios, and the technologies used in different applications

are slightly different. A schematic diagram of a relatively general speech recognition system as shown in Figure 2.

The behavior process of the speech recognition system is as follows: the original speech (the object to be recognized) is transformed into an electrical signal (speech signal) through the microphone and added to the input of the recognition system. First, it needs to undergo preprocessing, including antialiasing filtering, preemphasis, and endpoint detection. Then, it carries out feature extraction. The requirements for feature parameters are as follows: (1) the extracted feature parameters can be used to distinguish the speech features of track markers; (2) there is good independence between the parameters of each order; (3) it can calculate the characteristic parameters efficiently. The last is pattern recognition. General speech recognition system has two key points: one is the establishment of speech model base, which can usually be constructed through repeated training of historical data or directly by experts through accumulated experience. The second is the pattern recognition part, which accurately extracts important speech features and matches the model base for pattern recognition [16]. Based on the human-computer interaction interface of virtual fitting based on Kinect-based voice recognition system, the fitting person in front of the virtual fitting mirror only needs to say the model, brand, style, or other menu operations of the desired clothes (the following one, with a scarf, etc.), so that he can quickly select clothes, set samples, and try on clothes. The main task of the virtual fitting speech recognition system is to convert the voice input of the fitting into text output, so as to quickly, accurately, and conveniently operate the menu of the virtual fitting and provide an excellent user experience. The structure is mainly divided into two parts: training part and recognition part. Usually, the training part is completed in the off-line stage of the system. The signal processing method and speech mining method are used to combine the recognized speech to obtain the "language model" and "acoustic model" required by the speech recognition system. The online phase of the system usually completes the speech recognition process, and the well-trained speech model can automatically recognize the speech input by the user in real time. The recognition process can generally be divided into front module and rear module: the front module mainly carries out endpoint detection (used to remove long-time silence and noise), feature extraction, noise reduction, etc.; the function of the rear module is to extract the feature vector of the user's speech with the trained acoustic model and language model and then carry out statistical pattern recognition to obtain the text information contained therein [17]. In addition, the adaptive feedback module in the rear module can carry out adaptive learning according to the user's historical speech data, so as to train and correct the acoustic model and speech model and improve the accuracy of speech recognition. The block diagram of Kinect-based speech-to-character recognition system is shown in Figure 3.

The speech recognition system based on Kinect provides the rapidity and matching degree that the traditional physical fitting mode cannot achieve for virtual fitting, which is one of the important links to improve the user experience of virtual fitting [18]. In addition, the user voice commands

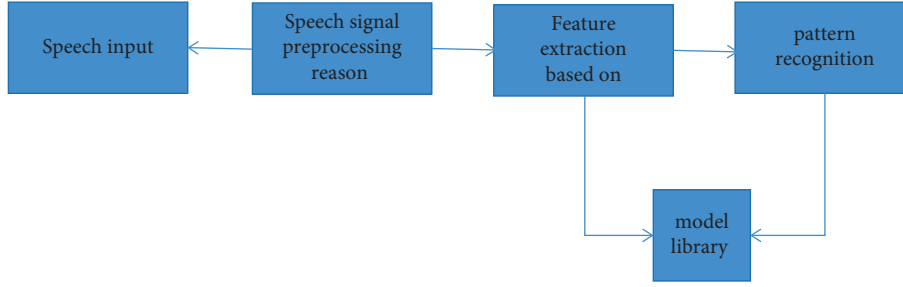


FIGURE 2: Schematic diagram of speech recognition system.

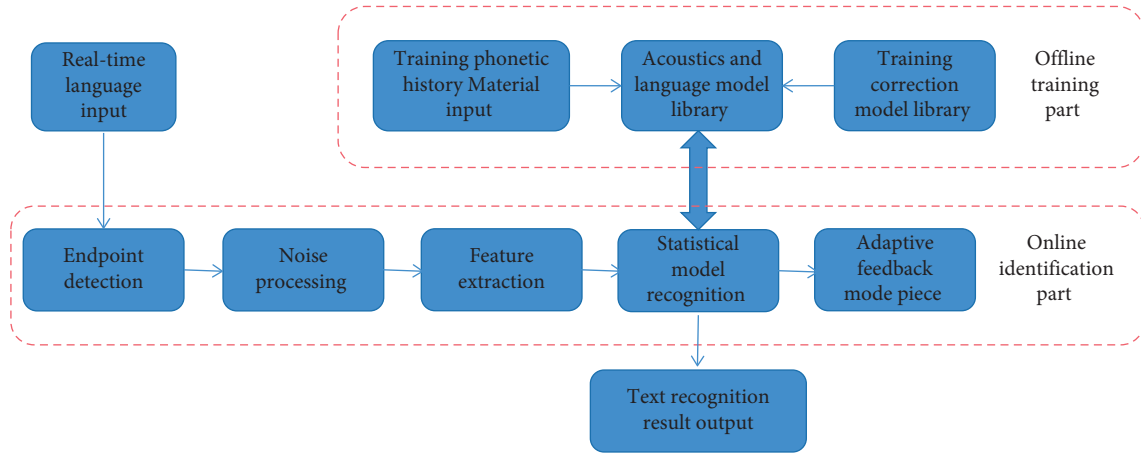


FIGURE 3: Block diagram of Kinect-based speech-to-character recognition system.

and fitting feelings in the acoustic and voice model library will accumulate all the time, which can not only achieve the unlimited experience accumulation and the ability to quickly learn the latest information that the waiters in the physical fitting store cannot have, but also provide a large amount of user feedback information, becoming a valuable information database for the perfect reproduction of clothes [4].

**3.1.2. Image Recognition Technology.** The image scale space is to introduce a parameter regarded as scale into the visual information (image information) processing model, obtain the visual processing information at different scales by continuously changing the scale parameters, and then synthesize the information to deeply dig the essential features of the image. The scale space of an image can be defined as the convolution operation of the original image and Gaussian kernel, and the scale size of the image can be expressed by Gaussian standard deviation. In computer vision, the scale space is represented as an image pyramid. In SIFT algorithm, the input image function is convolved with the kernel of Gaussian function repeatedly and the secondary sampling is carried out repeatedly. However, because each layer of image depends on the previous layer of image and the image size needs to be reset, the computation is large. The difference between SURF algorithm and SIFT algorithm in the use of pyramid principle lies in that SURF algorithm applies for increasing the size of image core, allowing multiple layers of scale space to be processed at the same time, and there is no

need for secondary sampling of images, thus improving the algorithm performance. The SURF algorithm uses scale interpolation and  $3*3*3$  nonextreme suppression method to extract feature points with constant scale.

Let  $I(x, y)$  be the input image,  $*$  be the convolution operator, and  $\delta$  be the scale variable, then  $G(x, y, \delta)$  is defined as follows:

$$G(x, y, \delta) = \frac{1}{2\pi\delta^2} e^{-(x^2+y^2)/2\delta^2}. \quad (1)$$

The Hessian equation of scale  $\delta$  at point  $(x, y)$  can be defined as follows:

$$\begin{aligned} H(x, y, \delta) &= L_{xx}(x, y, \delta)L_{yy}(x, y, \delta), \\ H(x, y, \delta) &= L_{xy}(x, y, \delta)L_{yy}(x, y, \delta). \end{aligned} \quad (2)$$

Among them,

$$\begin{aligned} L_{xx}(x, y, \delta) &= \frac{\partial^2 G(x, y, \delta)}{\partial x^2}, \\ L_{yy}(x, y, \delta) &= \frac{\partial^2 G(x, y, \delta)}{\partial y^2}, \\ L_{xy}(x, y, \delta) &= \frac{\partial^2 G(x, y, \delta)}{\partial x \partial y}. \end{aligned} \quad (3)$$

The SURF algorithm uses a box filter and integral image to approximate LoG of different scales. The coefficients of

the box filter are constrained by the  $F$  norm of the Hessian matrix and expressed as the determinant of

$$\frac{\|L_{xy}(1.2)\|_F \cdot \|D_{xx}(9)\|_F}{\|L_{xx}(1.2)\|_F \cdot \|D_{xy}(9)\|_F} = 0.912 \approx 0.9. \quad (4)$$

It can be concluded that

$$\det(H) = D_{xx}D_{yy} - (0.9D_{xy})^2. \quad (5)$$

According to equation (5), in order to ensure that the  $F$  norm of Hessian matrix is constant on the filter template of any size, the response of the filter is normalized according to the size of the template.

**3.2. Research on Related Technologies of Clothing Deformation Simulation Algorithm.** Due to the limitation of dimension, the calculation of distance measurement based on body surface difference is huge, and a large amount of memory may be required. Therefore, sensitive distance measurement is an effective method for clothing deformation sample selection. It is most appropriate when the distance measurement is joint angle or body surface. They only indirectly measure the difference between the input pose and the sample pose. Therefore, the adopted distance measurement method can directly obtain the sensitivity information and obtain the approximate difference between clothing grids [19]. In addition, the difference between the two clothing meshes can be expressed as  $\|y(p^{\text{in}}) - \bar{y}^j\|_2$ , where  $y(p^{\text{in}})$  is the actual clothing mesh of the queried posture and  $\Delta\bar{y}^j$  is the prediction calculation using SOR. However, how to get  $y(p^{\text{in}})$  and whether the queried pose should not be included, the key is to use sensitivity analysis to estimate  $y(p^{\text{in}})$  from the example clothing shape. Therefore, the distance measurement is called the sensitivity-based distance measurement. When the unit number of joint angle is much less than the number of clothing mesh, the calculation is more effective. The clothing mesh is further decomposed into several areas, so that the nearby clothing samples can be represented by each area, which can effectively reduce the number of examples in the database, and each area can make clothing deformation relatively independently. The bones are manually divided into area  $N_G$ . Then, the clothing weight  $\hat{w}_{g,y}$  is calculated for each clothing vertex  $y$ . When  $J=1$  of the body mesh closest to this area, the clothing vertex weight of each area can be calculated as the sum of skin weights. In this method, the garment vertex  $y$  connecting the bones belongs to the same region  $g$ , and the weight of this region will be 1, while the weight of the remaining region will be 0. For those unrelated garment vertices that are not in the same region, the sum of their region weights is guaranteed to be 1. When the weight of region  $g$  exceeds the garment mesh, it is also stable, so this method can maintain the smoothness around the joint [18]. In each area, the difference between the clothing shape of the input pose and the sample pose is calculated by the approximate weight sum of the clothing mesh difference.  $D_g^j(p^{\text{in}})$  as the weight sum of the joint angle difference is as follows:

$$\begin{aligned} D_g^j(p^{\text{in}}) &\approx \sum_{y \in Y} \hat{w}_{g,y} \|y^{\text{in}} - \bar{y}^j\|^2 \\ &\approx \sum_{y \in Y} \hat{w}_{g,y} \sum_{m=1}^{3N_L} \|(s_m^j - \bar{s}_m^j) \Delta\Theta_m\|^2 \\ &\approx \sum_{m=1}^{3N_L} U_{g,m}^j \|\Delta\Theta_m\|^2. \end{aligned} \quad (6)$$

The weight  $U_{g,m}^j = \sum_{y \in Y} \hat{w}_{g,y} \|s_m^j - \bar{s}_m^j\|^2$  here represents the clothing difference caused by the difference of multiple joint angles, which can be predicted by sensitivity analysis. In the offline stage, the weight of each data point is calculated to obtain  $s_m^j$  and  $\bar{s}_m^j$ , because this weight is the dimension of  $NG * (3nl)$  size and is very compact, so the system can quickly calculate the distance as long as a small storage space is used. The weight scheme is based on posture. The transformation of clothing mesh from one posture to another can be expressed as  $\|s_m^j - \bar{s}_m^j\|^2$ , where  $m$  is selected from the right shoulder joint to represent the weight change of relevant joint points, and its number represents the weight change from left to right [20]. Given a sensitivity based on distance  $D_g^j(p^{\text{in}})$ , the corresponding distance weight can be calculated from the clothing shape derived from each example pose, and the influence of distance weight between similar example poses can be controlled by increasing  $K$ . Experiments show that the fusion of clothing deformation and the smoothing of similar weight can better represent the deformation details of clothing. The blending weight  $w^j(p^{\text{in}})$  of each garment vertex can be expressed by its region weight and distance weight as follows:

$$w^j(p^{\text{in}}) = \sum_{y=1}^{N_G} \left( \frac{\hat{w}_g w_g^j(p^{\text{in}})}{\sum_{j=1}^{N_j} w_g^j(p^{\text{in}})} \right). \quad (7)$$

When the system obtains an input pose, the corresponding garment deformation will be generated by using sor and the deformation of the mixed sample garment based on sensitivity distance measurement. In many cases, the sensitivity-based weight scheme between the input pose and the sample pose is described as a coherent difference that changes smoothly with time. However, for sudden changes in input posture, the weight scheme may undergo sudden changes and may produce continuous mutations in the synthesis stage of clothing [21]. This problem can be prevented by introducing damping. The damping value can be calculated from the value of the current time period and the value of the previous time period. The clothing changes continuously in 0.05 seconds. The damping scheme can reduce the attenuation of clothing deformation response when the human body moves rapidly.

It is assumed that the bone studied has 22 bones and each bone has a certain range of rotation angle. As long as the surface of the manikin can change continuously with the input parameter  $P$ , the hypothetical scheme will not be affected and limited by the parameter change of the receptor table. In all examples, the double quaternion hybrid skinning method is used, and each volume net is composed of 12K triangles and 6K vertices. Use Pinnocchio to describe a given

TABLE 1: Clothing database parameters.

Clothing	T-shirt	Long-sleeved shirt	Shorts	Trousers
Number of garment vertices	12K	13K	11K	13K
Number of triangles in clothing mesh	24K	23K	14K	21K
System operating frequency (FPS)	63	59	68	61
Joint number	140	110	90	160
Database size (MB)	48	42	35	47

character and its associated bone skin weights. First, a database is created for four clothing models: T-shirts, long-sleeved shirts, shorts, and trousers of female body models. Table 1 lists some parameters of database construction of female body model, and PCA method is used to reduce memory occupation [22]. In addition, in order to reduce the reconstruction time, the weights in the database are divided into 6 to 8 clusters, and then local PCA is performed for each cluster. This setting can reduce the memory size required for database reconstruction.

For each posture of the input system, the garment deformation method based on the input human posture can be used to obtain realistic garment deformation effect, and the same material parameters can be applied to all kinds of garments. The scheme expresses the nonlinear and global relationship between body and clothing deformation. It can be seen that the clothes on the abdomen are raised and form new wrinkles, showing the full impact of the thickness and fineness of the cloth on the body geometry. When the model raises the arm, a new cloth wrinkle will appear according to the nonlinear deformation behavior. It tries to get a balance between the authenticity and speed of physical simulation and can achieve high-quality garment deformation details [23]. However, the artificial products of clothing animation, such as unnatural wrinkles and penetration, can still be observed by users sometimes. The Kinect is put on the top of the display and then the nearest and farthest moving distance of the model is limited, according to the placement height and opening angle of the acquisition device Kinect. The relationship between the distance range of human joint points observed by the Kinect and its placement height is listed in Table 2.

The experimental verification shows that the higher the Kinect is placed, the farther the nearest observation distance is, and the smaller the corresponding furthest observation distance is, thus the tracking ability of the lower body joint points will be reduced. Therefore, the Kinect is placed at 0.9 m, which can obtain the appropriate nearest and farthest observation distance, and the distance between people and Kinect is 2.0–3.0 m, which can ensure good image quality [24]. In the image acquisition process of Kinect, in addition to saving the image of each frame, it is also necessary to save the corresponding bone position information. In Stefan’s research method, the human contour feature points are used as the matching keywords for contour matching, that is, the model is allowed to wear the specified clothes into the specially arranged image information collection room, pose at will, collect 10 frames of images and save them locally, and then carry out contour extraction, matting, and other postprocessing on the images. The image acquisition method

TABLE 2: Relationship between placement height and human body distance.

Kinect placement height	Nearest observation distance	Maximum observation distance
0.9	1.814	4.179
1.0	1.957	4.157
1.1	2.084	4.124

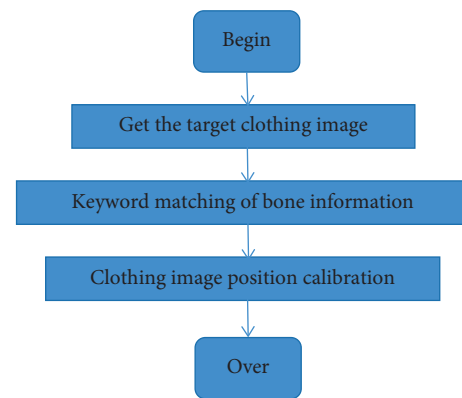


FIGURE 4: Matching transfer flow chart of garment image.

uses computer image processing to obtain image information, replaces the original method of using 8 cameras to obtain image information, and uses the image recognition algorithm to calculate the bone information joint point coordinates, which is used as the matching keyword of the clothing transfer algorithm. Then, the contour extraction algorithm is used to extract the image contour, which avoids some inevitable defects and huge workload caused by human manual matting [25]. In the final clothing image transfer calibration, the method of scaling the clothing model is used to calibrate the fit between clothing and human body, which reduces the amount of calculation and the use of storage space. The workflow of garment image transfer matching is shown in Figure 4.

#### 4. Results and Analysis

In order to obtain the start time of effective audio stream and prevent noise from occupying the processing system, it is necessary to detect the start and end endpoints of the audio stream input to the system. The detection method is to set a threshold. In order to highlight the resolution between short-time energy and zero-crossing energy, it is necessary to set at least the low threshold and zero-crossing threshold

of short-time energy. The low threshold can locate the real starting and ending point of the collected voice stream. It should be noted that it is basically the real endpoint of the voice stream that crosses the high threshold. However, what crosses the low threshold may not necessarily be the start and end endpoints of speech but may be just a burst of noise. Combined with the timing measurement index, after determining the fuzzy starting point of speech with high threshold, the reverse deduction method carefully determines the accurate starting point of speech flow with low threshold. The judgment of the end point has to be similar to the two threshold method. In a few cases, the noise energy will be large enough to break through the high threshold. In this case, the double-gate limited speech starting point will fail. Because the duration of high-energy noise is very short, we can judge whether it is noise or speech by counting the maintenance time of collected speech samples. In this way, the modified short-time energy double-threshold audio stream endpoint determination method is completed, as shown in Figure 5.

First, the double-threshold short-time energy method is used to locate the starting position of speech. The steps are as follows: (1) first set double thresholds, high and low thresholds  $E_H$  and  $E_L$ , respectively. The short-time energy segment  $N_{1U} \sim N_2$  higher than the high threshold  $E_H$  in the speech sample must be the main speech segment. (2) Then extend and  $N_{1U}$  to search the periphery of the voice segment. When the continuous short-term energy can still be greater than the set low threshold  $E_L$ , it is determined that the extension part still belongs to the speech segment. (3) The only difference between the two is that the zero-crossing rate of the silent segment is much lower than that of the light segment. Therefore, it is necessary to set an additional threshold of zero-crossing rate and extend the search from  $N_1$  and  $N_{2U}$  to both ends again. When the short-time zero-crossing rate is lower than 3 times (or more times) of the zero-crossing rate threshold, the current time is the starting point of the voice signal. Of course, in terms of practical effect, the gap between searching for the periphery or searching for light tone segments should be limited to the length of one frame of voice (generally 25 ms). As shown in Figure 6, the voice signal is detected by the endpoint detection method [26].

Audio stream requires complex speech processing to improve speech effect, usually including AGC, echo cancellation, and noise reduction. The class of Kinect voice stream can flexibly configure the working mode of its microphone array to produce the required processing effect. At the same time, it can detect the arrival sequence of audio stream in a specific direction and control the tracking direction of microphone array.

- (1) *Echo Cancellation*. Echo refers to that the user hears the voice of his previous speech during the call, and the echo voice will lag in the line for a period of time. The important idea of echo cancellation algorithm is to compare the speaker's previous voice frequency amplitude information and offset his previous voice information from the received audio information, so as to eliminate the echo.

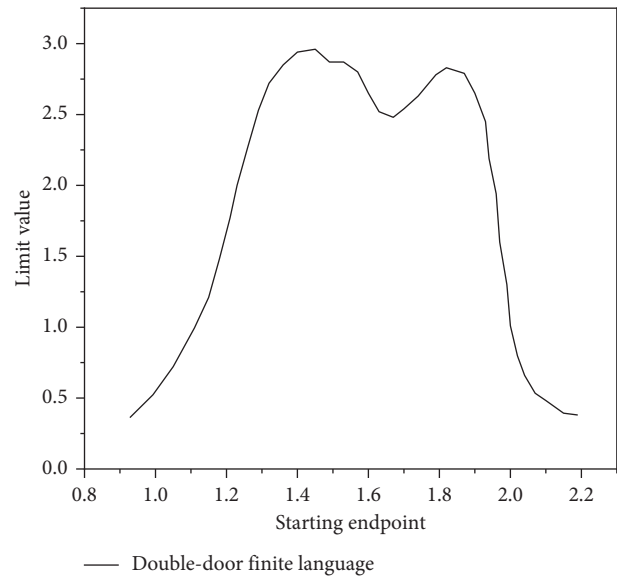


FIGURE 5: Determination of start and end points of audio stream by double-threshold method of short-time energy.

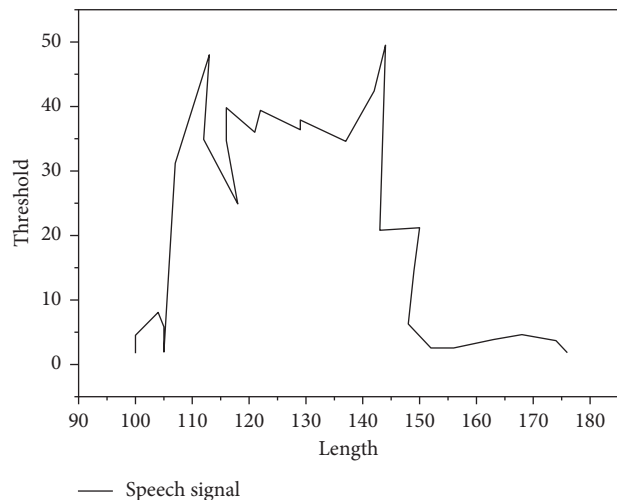


FIGURE 6: Voice signal detection by endpoint method.

- (2) *AGC (Automatic Gain Control)*. It is used to balance or maintain the consistency of the user's voice strength for a period of time. It prevents the volume from surging or suddenly weakening when the user is close to and away from the microphone. AGC smooths the volume for a period of time, so that the volume of the call is as natural as talking face to face. The Kinect adopts the gain boundary technology. In addition, even if the signal segment with too high sound intensity is successfully obtained, it will be discarded because it is too saturated and easy to be distorted.
- (3) *Noise Reduction Processing*. Noise reduction methods mainly include noise filling and noise suppression. Noise filling is to add a small amount of noise to some signals after the residual echo signal is

removed by pointing to the center clipping, which can obtain better QES than silent signal (user experience). Noise suppression is to eliminate the nonverbal information received in the microphone. After deleting the background noise, the actual voice is clearer, less blurred, and less burred.

The Kinect has a class for noise characteristic control. It is EchoCancellationMode. In order to match AEC, the EchoCancellationSpeakerIndex property needs to be assigned an integer value. This integer value is used to specify the specific user who needs to control noise. The search and power on drive functions of the microphone array are run spontaneously by the SDK. Echo cancellation is a core technology of Kinect. The testing process is very complex. First, we need to add a CheckBox on the interface and then create IsAECOn to configure the properties. Finally, the IsCheck attribute of the CheckBox is bound to IsAECOn. In terms of speech recognition, the research results of this study can identify the source direction of speech, and the user interface can generate corresponding actions with the change of speech, but the current speech recognition technology of the fitting system can only perform specific fitting-specific speech. In case of recognition, it is still impossible to recognize the voice of the user's random speech.

## 5. Conclusion

With the rapid development of virtual reality technology, the research and development of virtual fitting system have attracted a large number of researchers, but the key technology of virtual fitting is still immature, and some problems have not been well solved. Relying on the virtual fitting development and research project of an institute, the theoretical research and corresponding application level programming of three key technologies including bone skin, clothing deformation, and speech recognition operation are carried out. Taking 3D virtual fitting as the starting point, the key technologies of virtual fitting are studied, and the applications of human bone skin technology, clothing deformation algorithm, clothing image transfer algorithm, and speech recognition technology in virtual fitting system are analyzed. In clothing deformation, a clothing deformation algorithm based on sensitivity is proposed, which first uses the obtained posture bone skin, then uses the clothing database to detect the similarity of each part of the input human posture, then deforms the corresponding regions respectively, and finally mixes the deformation of each region. For the garment image transfer algorithm, bone joint points are proposed as the eigenvalues of garment transfer, which improves the accuracy of garment transfer and reduces the amount of computation. In the operation mode of the virtual fitting system, the development and implementation of key technologies of speech recognition based on Microsoft Kinect acquisition device are proposed.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] X. Ren, "Research on a software architecture of speech recognition and detection based on interactive reconstruction model," *International Journal of Speech Technology*, vol. 24, no. 4, pp. 1–9, 2021.
- [2] Y.-P. Zhang and X.-P. Xu, "Design of a wearable human-computer interaction system based on bioelectrical signal recognition technology," *Journal of Interdisciplinary Mathematics*, vol. 21, no. 5, pp. 1049–1054, 2018.
- [3] X. Liu and Z. Yang, "Research on legibility of English text based on improved decision tree and intelligent interactive system," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 4, pp. 5737–5747, 2020.
- [4] P. Li and J.-H. Chen, "A model of an e-customized co-design system on garment design," *International Journal of Clothing Science & Technology*, vol. 30, no. 5, pp. 628–640, 2018.
- [5] L. Jiang, J. Ye, L. Sun, and J. Li, "Transferring and fitting fixed-sized garments onto bodies of various dimensions and postures," *Computer-Aided Design*, vol. 106, no. 1, pp. 30–42, 2019.
- [6] H.-T. Chiang and C.-T. Chang, "Introduction to and application analysis of taiwan's nhi-medcloud system," *Journal of Service Science Research*, vol. 11, no. 1, pp. 93–115, 2019.
- [7] King and R. Pan, "Review of printed fabric pattern segmentation analysis and application," *Autex Research Journal*, vol. 20, no. 4, pp. 530–538, 2020.
- [8] J. R. Wang and G. X. Zhang, "Research on application and development of segmented srm," *Guti Huojian Jishu/Journal of Solid Rocket Technology*, vol. 39, no. 4, pp. 451–455, 2016.
- [9] R. Nemcokova and B. Musilova, "Implementing mass customization into clothing production," *Vlakna a Textil*, vol. 20, no. 4, pp. 12–19, 2013.
- [10] X.-j. Zhu, H. Lu, and M. Rättsch, "An interactive clothing design and personalized virtual display system," *Multimedia Tools and Applications*, vol. 77, no. 20, pp. 27163–27179, 2018.
- [11] M. R. Jones and V. L. Giddings, "Tall women's satisfaction with the fit and style of tall women's clothing," *Journal of Fashion Marketing and Management: International Journal*, vol. 14, no. 1, pp. 58–71, 2010.
- [12] H. Moon and H. H. Lee, "Consumer's preference fit and ability to express preferences in the use of online mass customization," *The Journal of Research in Indian Medicine*, vol. 8, no. 2, pp. 27–32, 2014.
- [13] X. Zeng, Y. Chen, W. Yu, M. Happiette, P. Bruniaux, and R. Ng, "A new method of ease allowance generation for personalization of garment design," *International Journal of Clothing Science & Technology*, vol. 20, no. 3, pp. 161–173, 2008.
- [14] M. L. Mpampa, P. N. Azariadis, and N. S. Sapidis, "A new methodology for the development of sizing systems for the mass customization of garments," *International Journal of Clothing Science & Technology*, vol. 22, no. 1, pp. 49–68, 2010.
- [15] M. M. Senanayake and T. J. Little, "Mass customization: points and extent of apparel customization," *Journal of Fashion Marketing and Management: International Journal*, vol. 14, no. 2, pp. 282–299, 2010.
- [16] H. Moon and H. H. Lee, "Product recommendation service in online mass customization: consumers' cognitive and affective



- responses,” *Journal of the Korean Society of Clothing & Textiles*, vol. 36, no. 11, pp. 60–65, 2012.
- [17] S. Inui, T. Yamada, Y. Horiba, and M. Hashimoto, “A preliminary study of a cloth guiding mechanism for automatic sewing system,” *International Journal of Clothing Science & Technology*, vol. 24, no. 1, pp. 6–14, 2012.
- [18] M. Senanayake, A. Raheja, and Y. Zhang, “Automated human body measurement extraction,” *International Journal of Clothing Science & Technology*, vol. 30, no. 2, pp. 175–194, 2018.
- [19] Y. Zhong, D. Li, G. Wu, and P. Hu, “Automatic body measurement based on slicing loops,” *International Journal of Clothing Science & Technology*, vol. 30, no. 3, pp. 380–397, 2018.
- [20] X. Bugao, S. Junqiang, and Guolian, “Development of individualized pattern prototype based on classification of body features,” *International Journal of Clothing Science and Technology*, vol. 27, no. 6, pp. 895–907, 2015.
- [21] S. Gill, “Sizing in clothing: developing effective sizing systems for ready-to-wear clothing,” *Journal of Fashion Marketing and Management: International Journal*, vol. 12, no. 4, pp. 579–581, 2008.
- [22] L. Toettrup, S. F. Atashzar, D. Farina, E. N. Kamavuako, and W. Jensen, “Nerve injury decreases hyperacute resting-state connectivity between the anterior cingulate and primary somatosensory cortex in anesthetized rats,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 99, pp. 1–8, 2020.
- [23] Koomiran, “A study on the shape of hem-line of semi-flare skirts by p/w blended fabric’s-based on the comparison between real clothing and 3d virtual clothing,” *Journal of Korean Traditional Costume*, vol. 13, no. 2, pp. 145–155, 2010.
- [24] S. Guo, T.-M. Choi, B. Shen, and S. Jung, “Inventory management in mass customization operations: a review,” *IEEE Transactions on Engineering Management*, vol. 66, no. 3, pp. 412–428, 2019.
- [25] D. Villani, E. Gatti, S. Triberti, E. Confalonieri, and G. Riva, “Exploration of virtual body-representation in adolescence: the role of age and sex in avatar customization,” *SpringerPlus*, vol. 5, no. 1, pp. 740–813, 2016.
- [26] K. Yang and A. P. Young, “The effects of customized site features on internet apparel shopping,” *Journal of Fashion Marketing and Management: International Journal*, vol. 13, no. 1, pp. 128–139, 2009.