Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1503325, 11 pages
https://doi.org/10.1155/2021/1503325

Research Article

Dynamic Gesture Contour Feature Extraction Method Using Residual Network Transfer Learning

Xianmin Ma and Xiaofeng Li

Department of Information Engineering, Heilongjiang International University, Harbin 150025, China

Correspondence should be addressed to Xiaofeng Li; lixiaofeng@hiu.net.cn

Received 2 August 2021; Revised 10 September 2021; Accepted 24 September 2021; Published 13 October 2021

Academic Editor: Deepak Gupta

The current dynamic gesture contour feature extraction method has the problems that the recognition rate of dynamic gesture contour feature and the recognition accuracy of dynamic gesture type are low, the recognition time is long, and comprehensive is poor. Therefore, we propose a dynamic gesture contour feature extraction method using residual network transfer learning. Sensors are used to integrate dynamic gesture information. The distance between the dynamic gesture and the acquisition device is detected by transfer learning, the dynamic gesture image is segmented, and the characteristic contour image is initialized. The residual network method is used to accurately identify the contour and texture features of dynamic gestures. Fusion processing weights are used to trace the contour features of dynamic gestures frame by frame, and the contour area of dynamic gestures is processed by gray and binarization to realize the extraction of contour features of dynamic gestures. The results show that the dynamic gesture contour feature recognition rate of the proposed method is 91%, the recognition time is 11.6 s, and the dynamic gesture type recognition accuracy rate is 92%. Therefore, this method can effectively improve the recognition rate and type recognition accuracy of dynamic gesture contour features and shorten the time for dynamic gesture contour feature recognition, and the F value is 0.92, with good comprehensive performance.

1. Introduction

Gesture is an intuitive and convenient way of communication, and natural and comfortable human-computer interaction can be realized through gesture recognition [1]. At present, dynamic gestures are more intuitive than static gestures and are widely used in flexible human-computer interaction applications. It can not only manipulate virtual objects in a virtual reality environment but also can be widely used in smart home appliances and corresponding automatic control fields [2–3]. However, due to the complexity of the human hand structure, dynamic gestures are presented as diverse. The recognition of dynamic gestures has always been one of the difficulties in research. Traditional gesture recognition methods are usually implemented by background filtering and feature extraction. They are easy to be affected by external factors such as light, reduce the performance of the algorithm, and are difficult to obtain satisfactory results.

In recent years, with the rise of deep learning algorithms, a deep residual network has been widely used as a new deep learning algorithm. This learning algorithm adopts end-to-end learning strategy, which can efficiently and independently realize image feature analysis, explore the internal features of images, and help improve the effect of image analysis. However, the deep learning model is complicated for sample training and learning process. Therefore, transfer learning is introduced to migrate the learned sample data or model parameters to the new model to avoid the lack of zero learning in the deep learning network [4]. There are many related studies on the transfer learning of residual network. The literature [5] proposes the method of combining wide residual and long-term and short-term memory network. Convolutional neural network is used to extract features synchronously in space and time dimensions, respectively, and the long-term and short-term memory network is used to analyze features synchronously and input them into the residual network, so as to finally realize gesture recognition.

The literature [6] uses deep neural network and residual learning technology to study nonlinear wavefront sensing. The deep residual learning method expands the usable range of Lyot-based low order wavefront sensors (LLOWFS) by more than one order of magnitude and can improve the closed-loop control of the system with large initial wavefront error. This paper shows the advantages of residual network. The literature [7] proposes a deep residual convolutional neural network (CNN) model to self-learn the hidden median filter traces in JPEG lossy compressed images. In order to alleviate the over fitting problem of the deeper CNN model, this paper adopts a data enhancement scheme in training to increase the diversity of training data, so as to obtain a more stable median filter detector.

To solve the above problems, this paper proposes a dynamic gesture contour feature extraction method based on residual network transfer learning.

(1) The sensor is used to integrate the dynamic gesture information. The distance between the dynamic gesture and the acquisition device is detected by transfer learning, and the dynamic gesture image is segmented in the background

(2) The feature contour image is initialized, the dynamic gesture feature contour image was replaced, and the dynamic gesture feature recognition model was trained. The residual network method is used to accurately identify the contour and texture features of dynamic gestures

(3) The processing weights are fused to obtain a significant contour response, and the contour features of dynamic gestures are traced frame by frame. According to the geometric features of the dynamic gesture region, the dynamic gesture part is projected longitudinally, and the dynamic gesture contour area is processed by gray and binarization to realize the dynamic gesture contour feature extraction

2. Related Work

At present, a large number of scholars in this field have conducted research on it and achieved certain research results. The literature [8] proposes a multiloss recurrent residual network, a multitask-based method that performs gesture recognition and time detection at the same time. It adopts a double loss function, which assigns the category of each frame of video to a gesture class, and determines the frame interval related to each gesture, so as to extract gesture contour and gesture motion features to complete dynamic gesture recognition. The dynamic gesture recognition speed of this method is fast, but the contour and texture of dynamic gesture recognition are not considered, resulting in low accuracy of dynamic gesture recognition. The literature [9] proposed a dynamic gesture recognition method based on a feature fusion network and variant convolution Long Short-Term Memory (ConvLSTM) and designed the gesture recognition architecture. The architecture extracts spatiotemporal feature information from local, global, and deep and combines feature fusion to reduce the loss of feature information. The local spatiotemporal feature information is extracted from video sequences by using the three-dimensional residual network of channel feature fusion. The variation ConvLSTM is used to learn the global temporal and spatial information of dynamic gestures, and the attention mechanism is introduced to change the gate structure of ConvLSTM to realize dynamic gesture recognition. This method can effectively improve the accuracy of dynamic gesture recognition, which has the problem of long time of dynamic gesture recognition. The literature [10] proposed dynamic gesture recognition based on selective spatiotemporal feature learning, combined with the dynamic selection mechanism of selective spatiotemporal feature learning of ResC3D network and ConvLSTM, improved the architecture of dynamic gesture fusion model, adaptively adjusted dynamic gesture data, extracted short-term and long-term spatiotemporal features of dynamic gesture learning, and recognized dynamic gestures. The dynamic gesture recognition rate of this method is high, but the accuracy of dynamic gesture type recognition is low. The literature [11] proposes a dynamic gesture recognition method based on short-time sampling neural network. The short-time sampling neural network is used to integrate verified modules to learn short-term and long-term features from video input. Each video input is divided into a fixed number of frame groups, one frame is selected and represented as RGB image and optical flow snapshot, and the convolution neural network is input to extract features and output long-term short-term memory network to recognize dynamic gestures. This method has good robustness. However, the recognition efficiency is low. The literature [12] proposed a dynamic gesture recognition algorithm based on simultaneous detection and classification of wide residual network and long-term and short-term memory network. Firstly, the spatiotemporal features are extracted from the fine-tuning three-dimensional convolutional neural network. Secondly, the bidirectional convolutional long-term and short-term memory network is used to further consider the time aspect of image sequence. Finally, these advanced features are sent to a wide range of residual networks for final gesture recognition. The success rate of gesture recognition is high, but the accuracy of gesture recognition is low.

Relevant achievements have also been made in China. The literature [13] proposes a multifeature dynamic gesture recognition method, which uses the somatosensory controller leap motion to track the dynamic gesture to obtain data, extract the displacement vector angle and inflection point judgment count, train the dynamic gesture using the hidden Markov model, and recognize the multifeature dynamic gesture according to the matching rate of the gesture to be tested and the model. This method can effectively improve the recognition rate of similar gestures, but the recognition time is long. Literature [14] proposed a feature extraction method based on the geometric distribution of gestures, normalized the segmented gesture image, calculated the width length ratio of the minimum circumscribed rectangle of the gesture main direction and the gesture contour, preliminarily identified it by using the similarity function, counted

the distribution of gesture contour points by using the contour segmentation method, and recognized the gesture contour according to the modified Hausdorff distance similarity measurement method. The recognition time of this method is shorter, but the recognition accuracy is lower.

In order to solve the shortcomings of research, a dynamic gesture contour feature extraction method based on residual network transfer learning is proposed, and the performance of the proposed method is verified by experiments. The results show that the proposed method has high dynamic gesture contour feature recognition rate and gesture type recognition accuracy rate, and the recognition time is short, only 11.6 s, and the average F value of feature extraction is as high as 0.92.

3. Dynamic Gesture Contour Feature Extraction Method Based on Residual Network Transfer Learning

3.1. Dynamic Gesture Segmentation Image. Complicated background and external lighting can easily affect the acquisition of dynamic gesture images, so it is particularly important to choose a suitable gesture collection device. The sensor is used to segment the depth information of the palm gesture from the complex background. The distance between the dynamic gesture and the collection device is detected by transfer learning.

Assume that the source domain data sample of transfer learning is expressed as D_s, the label of the source domain data sample is expressed as L_s, the target domain data sample is expressed as D_t, and the label of the target domain data sample is expressed as L_t. Nonlinearly change the characteristics of the source and target domains, and align the source and target domain characteristics to perform second-order statistics. If the dynamic gesture contour feature covariance matrix is expressed as C_t, the loss between the contour features of the dynamic gesture on each feature layer of the source domain and the target domain is

$$l_s = \frac{1}{4C_t}\|D_s - D_t\|^2\|L_s - L_t\|^2. \tag{1}$$

For the classification model, the feature adaptation method is adopted, if the source domain data sample classifier is U_f, the target domain data sample classifier is U_l, and then the classification loss $P(D_s)$ of the source domain data sample D_s of transfer learning can be expressed as

$$P(D_s) = \min \frac{1}{C_t} P\left(U_f, U_l\right), \tag{2}$$

where $P(U_f, U_l)$ is cross entropy function. The classification loss $P(D_s)$ of the target domain data sample D_t of transfer learning can be expressed as

$$P(D_t) = \min \frac{1}{C_t} K(U_l), \tag{3}$$

where $K(U_l)$ is entropy function. Thus, the loss function of constructing the migration learning architecture is expressed as

$$S(h) = \min \frac{1}{C_t} P(D_s) + \frac{1}{\gamma} P(D_t) + \chi l_s, \tag{4}$$

where γ is the trade-off parameters of the target domain data sample and χ is loss trade-off parameters between dynamic gesture contour features. In gesture human-computer interaction, the palm is always in front of the camera, so by selecting an appropriate depth distance threshold, the palm gesture information can be separated from the background. However, the selection of the distance threshold is very difficult, and the threshold selection is inappropriate, which easily leads to the segmented gestures including arms. Or when the gesture is close to the body, the palm information cannot be segmented [15]. In order to overcome the influence caused by the selection of the threshold, the depth mining of image information is carried out by detecting the information of the gesture skeleton node. The sensor is used to collect the gesture bone node data, find the position of the palm node, and search for dynamic gestures in the palm node range. When all the pixels on the entire palm are close to the camera, setting a distance difference threshold can separate the gesture information from the background, and the dynamic gesture segmentation process. It is shown in Figure 1.

As the collector tracks the dynamic gesture skeletal node, the phenomenon of node drift is prone to occur. At this time, the distance between the palm node and the collector is not the actual distance between the palm node and the collector. When the distance difference threshold is segmented, the gesture segmentation will fail. Therefore, an approximate method for determining the position of the palm node is designed [16]. The position coordinates of all self-color pixels are averaged in a circle with the palm node as the center and the distance r between the palm node and the wrist node as the radius. This can be calculated by using the mean value to represent the position coordinates of the palm node. Therefore, the (x_p, y_p) equation of the palm node collection is

$$\left(x_p, y_p\right) = \left(\frac{1}{T}\sum_{i=1}^{T} x_i, \frac{1}{T}\sum_{i=1}^{T} y_i\right), \tag{5}$$

where T is the number of self-color pixels in the circle, x_i is the horizontal vector of the i self-color pixel, and y_i is the vertical vector of the i self-color pixel. After the position of the palm node is found, the gesture is segmented by judging the distance difference between the palm node and the pixels in the surrounding area to the collector. When the palm node is found, the gesture pixels need to be divided around the palm node [17]. In order to prevent the deviation of the gesture pixel points due to the drift of the palm node, the gesture pixel points are segmented in a large rectangular area centered on the palm node. The algorithm process is as follows: suppose the distance from the bone palm node

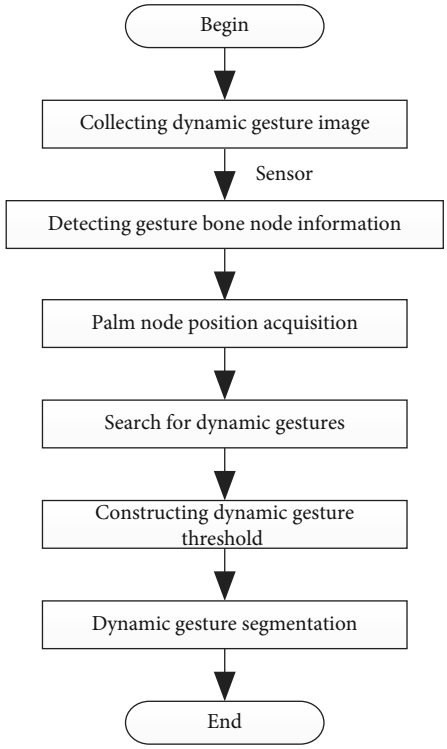


Figure 1: Dynamic gesture segmentation process.

extracted by the collector to the camera of the collector is d, the position of the palm node is (x_p, y_p, d_p), and the position of the ankle node is (x_r, y_r, d_r). Perform dynamic gesture pixel segmentation in a rectangular pixel centered on the palm node with a width of W and a height of H. It is shown in

$$W = \frac{1}{X\left(x_p, y_p\right)} Hd\sqrt{\left(x_p - x_r\right)^2 + \left(y_p - y_r\right)^2}. \quad (6)$$

The dynamic gesture segmentation is carried out through equation (6), and the basic pixel points of each dynamic gesture image are obtained, so that the subsequent recognition of the dynamic gesture contour feature is more unified, thereby increasing the rate of dynamic gesture extraction.

3.2. Recognize the Characteristic Contours of Dynamic Gestures. Recognize the segmented dynamic gesture image. First, initialize the value of the feature contour image, output the network model of the last layer of the training connection layer, and replace it with the dynamic gesture feature contour image. On this basis, the dynamic gesture feature recognition model is trained, the training set is recognized by the model obtained during each training, and the training accuracy is obtained. The dynamic gesture feature contour image is recognized to obtain the accuracy of the test set. When the current method performs dynamic gesture feature contour recognition, it cannot accurately recognize the contour and texture of the dynamic gesture, resulting in a low accuracy of dynamic gesture recognition. This is caused by the deep learning neural network layer of the recognition machine. Introduce the residual network model, before the level output, through the use of identity mapping, let the output layer cross the previous layer for data entry, and execute the identity mapping signal to avoid the network layer being too deep [18].

If the residual network has a total of l layers, the fully connected layer belongs to the last layer in the residual network, and the output of the residual mapping is expressed as

$$F\left(x^{l-1}\right) = g\left(\sigma_i \otimes g\left(\sigma_i \otimes x^{l-1}\right)\right). \quad (7)$$

Then, the output of the residual unit in the residual network model is expressed as

$$H\left(x^{l-1}\right) = g\left(F\left(x^{l-1}\right) + \kappa_i \otimes x^{l-1}\right). \quad (8)$$

where σ_i is convolution kernel parameters, x^{l-1} is $l-1$ layer input, κ_i is convolution kernel parameters for convolution operation, $\otimes$ is the convolution operation, and $g(x)$ is activated function. It is shown in

$$g(x) = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{if } x \le 0, \end{cases} \quad (9)$$

Due to the clever structure of each residual unit, when the gradient is transmitted backward, the error transmission mode is converted from the form of continuous multiplication to the form of addition to the upper level error sensitive items, thereby avoiding the gradient explosion and gradient caused by continuous multiplication. The disappearance of the problem ensures the effectiveness of error transmission.

The actual label and the model predicted label error are expressed as J. When parameters at layer $l-1$ are updated, the gradient of the current layer needs to be passed to this layer, and the back propagation algorithm is used to calculate the error sensitive term of this layer as

$$\vartheta^{l-1} = \frac{\partial J}{\partial x^{l-1}}, \quad (10)$$

If the gradient transfer error sensitive term of the $l+1$ layer is expressed as

$$\vartheta^{l+1} = \frac{\partial J}{\partial H\left(x^{l-1}\right)}, \quad (11)$$

then, in the residual network model, when the shortcut is not connected, $H(x^{l-1}) = F(x^{l-1})$, ϑ^{l-1} can be expressed as

$$\vartheta^{l-1} = \frac{\partial J}{\partial H\left(x^{l-1}\right)} \frac{\partial H\left(x^{l-1}\right)}{\partial x^{l-1}}. \quad (12)$$

When the shortcut is connected, $H(x^{l-1}) = F(x^{l-1}) + x^{l-1}$,

in back transmission, ϑ^{l-1} can be expressed as

$$\vartheta^{l-1} = \frac{\partial J}{\partial H\left(x^{l-1}\right)}\left(\frac{\partial H\left(x^{l-1}\right)}{\partial x^{l-1}} + 1\right). \tag{13}$$

During the transfer process, due to the residual unit structure of the residual network, the gradient of the previous layer can be directly transferred to the next layer and will not disappear due to the continuous multiplication of the transfer. In order to expand the scale of the training and recognition of dynamic gesture feature contours, the residual network's processing of the training set images is enhanced. The processing process includes flipping, rotating, filtering, cropping, and deformation. Because the difference of dynamic gesture changes is small, it needs to rely on a certain subtle feature to distinguish. Therefore, the focus of dynamic gesture feature contour recognition is to use the residual network to accurately identify the various features of dynamic gestures. However, if data enhancement methods such as cropping and deformation are used, important features may be lost, and the image after data enhancement may become unreliable. For different gestures, take the center node of the wrist as the relative coordinate, and set the wrist coordinate point as (x, y, z), and suppose the coordinates corresponding to the feature points of the gesture are $p = (x_i, y_i, z_i)$, $i = 1, 2, 3, 4 \cdots n$. Therefore, the distance (S_i) between the wrist coordinates and the dynamic gesture feature is

$$S_i = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}. \tag{14}$$

According to the above equation, within a certain time T, the motion sequence of the dynamic gesture skeleton node is expressed as $(D = S_i^1, S_i^2, S_i^3, \cdots, S_i^n)$, and the time series of the test gesture is $(Y = T_1, T_2, T_3, \cdots, T_n)$. Find the best point pair between the two sequences so that the sum of the distances between the corresponding points is the smallest, expressed as

$$Q(x, y) = \vartheta^{l-1} S_i \sum_{n}^{i=1} \varpi(x_i, y_i, z_i), \tag{15}$$

where ϖ is the corresponding node coefficient. The coordinates corresponding to the feature points of the gesture are $p = (x_i, y_i, z_i)$. Through the optimal sum of the distances of the points, the dynamic gesture images of the training set are recognized, and three methods are used for enhancement processing: rotation, flip, and filtering. Rotation is a random selection between -300 and 300. After rotation, the margins of the edge are filled with the adjacent background; the image is flipped horizontally. Image filtering adopts Gaussian filtering method, and the filtered image adds noise and is clearer than the original image [19]. After the original dynamic gesture images were enhanced in three ways, the training set image data set was expanded to 7044 images. After processing, there are many kinds of gesture recognition methods, such as neural network, SVM, and convolutional neural network. SVM has obvious effects on the two-classification processing, and the data outside the training set can achieve accurate prediction, low generalization error rate, and good real-time performance. Therefore, the residual network transfer learning is used to recognize the segmented gesture image and evaluate the normality of the gesture. The residual network migration learns the results of dynamic gesture recognition, tests the credibility of gestures and standard gestures, and provides a basis for dynamic gesture trace points.

3.3. Scan Feature Contour Frame by Frame. By identifying the feature contours of dynamic gestures, the basic data of dynamic gesture images and the skeleton data of gesture nodes are obtained. Consider the physiological characteristics of a single perception field in the visual pathway or the contextual adjustment between optic nerves [20]. For the same image, preserving contour information and removing background texture are usually contradictory. In this paper, the salient contour information of some spatial frequency tuning channels is transferred to the primary visual cortex in parallel, and the weight fusion processing is realized, and the salient contour response $\text{total} - r(x, y)$ is finally obtained. It is shown in

$$\text{Total} - r(x, y) = \sum_{j}^{n'}\left(\beta_j \cdot R_j(x, y)\right), \tag{16}$$

where X is the value of the contour feature point of the dynamic gesture on the x-coordinate axis, y is the value of the contour feature point of the dynamic gesture on the y-coordinate axis, β_j is the reference parameter of the node, R_j is the error parameter of the node, and n' is band fusion weight of the contour response on the frequency tuning channel of the n space, whose value is $[0, 1]$. Considering that on the low-frequency spatial frequency tuning channel, the contour map mainly represents the overall contour of the image [21]. Therefore, based on the fusion coding of the frequency-divided visual information stream under the primary visual cortex, and considering the influence of the frequency-divided characteristic parameters on the fusion weight, a value model for drawing points is proposed. It is shown in

$$\beta_j = [\text{total} - r(x, y)]\frac{e^{-\partial_j}}{\sum_{j=1}^{n'} e}, \tag{17}$$

where $e^{-\partial_j}$ is the outer value of feature trace node and e is inner value of feature trace node. Through equation (7), the contour feature of the dynamic gesture is traced, and the coordinate data of the skeleton node of the contour of the dynamic gesture is extracted to realize the feature trace of the dynamic gesture. Next, there is a need to go through the standard dynamic gesture detection of the dynamic gesture sample library, and the contour features of the dynamic

gesture by matching the optimal residual network transfer learning method were extracted.

3.4. Extract Contour Features of Dynamic Gestures. The color space model is used to check the area of the dynamic gestures. After obtaining the dynamic gesture region, according to the geometric features in the dynamic gesture sample library, the dynamic gesture region is traced a second time, and the gesture region square is extracted. According to the color characteristics between the images, grayscale changes and binarization remove the brighter interference items such as the background and then remove the darker interference items such as nails and dark backgrounds. Finally, the projection filtering operation is performed, and the vertical projection operation is performed on the dynamic gesture part. Set a certain threshold. If the vertical projection is less than the threshold, it will be set to black to remove interference items such as irrelevant lines [22, 23]. Perform skin color model processing on the obtained dynamic gesture contour area window to detect the approximate range of the dynamic gesture area. Determine the skin area; it is shown in

Suppose A is the area of the skin area of the dynamic gesture contour and B is to identify the total area of the image. Then, it is determined that the skin area is a dynamic gesture contour area. The area of the filtered area was sorted, and the largest value is the area square of the final detection. Assign 0 to the nondynamic gesture contour area in the original image, that is, paint it in black. The method is to first assign the original image to 0 and black it, and then the detected dynamic gesture contour area window to 1 white was assigned, and then the AND operation was performed with the unprocessed original image to obtain an image with only the dynamic gesture contour area square. Perform grayscale processing on the above image, and then binarize it; only the dynamic gesture contour area exists, and it is white. After multiplying by the unprocessed original image, a picture with only dynamic gesture outline area (including nails and darker background) is obtained [24–25]. Perform grayscale processing on the image obtained above (set a certain threshold). The white area obtained after binarization is the dynamic gesture contour area. According to the method process of the image projection filtering operation, the white connected areas of the dynamic gesture area are labeled, and each line of each area is traversed. Add each row of the area and project the image longitudinally to obtain the processed dynamic gesture contour feature image, which realizes the dynamic gesture contour feature extraction.

3.5. Method for Feature Recognition of Dynamic Gesture Contours. The residual network migration learning is used to realize dynamic gesture contour feature and dynamic gesture type recognition; the dynamic gesture contour feature recognition method is described as

(i) Input: obtain the loss function of the migration learning architecture through the source and target domains with or without labels in the migration learning $S(h)$. Set the input of a layer in the residual network to τ. After passing some convolutional layers and activation functions, a residual mapping $F(x^{l-1})$ is obtained.

(ii) Output: extraction result of dynamic gesture contour feature potential $T(\lambda)$.

(1) The sensor is used to segment the depth information of the palm gesture from the complex background. The distance between the dynamic gesture and the collection device is detected by transfer learning. Construct the loss function of the transfer learning architecture $S(h)$. Calculate the position coordinates of the palm nodes, and segment the pixels of the dynamic gesture image

(2) Initialize the contour image of the dynamic gesture feature, and use the residual network method to calculate the error-sensitive items ϑ^{l-1}. Obtain the distance between the wrist coordinate and the dynamic gesture feature, and accurately identify the contour and texture feature of the dynamic gesture through the optimal sum of the distances of the points

(3) Obtain significant contour response total $- r(x, y)$ through weight fusion processing. Considering the influence of the frequency-division characteristic parameters on the fusion weight, the β_j is constructed to draw the point value model, and the dynamic gesture contour features are drawn frame by frame

(4) The gray scale and binarization process the dynamic gesture contour area, detect the approximate range of the dynamic gesture area, and obtain the dynamic gesture contour feature extraction result as $T(\lambda)$

$$T(\lambda) = \frac{1}{S(h)} F\left(x^{l-1}\right) \beta_j r(x, y) \tag{18}$$

(5) End

In summary, to achieve the feature extraction of dynamic gesture contours, the specific process is shown in Figure 2.

4. Experimental Analysis and Results

4.1. Experimental Environment and Data Set. In order to verify the effectiveness of the dynamic gesture contour feature extraction method based on residual network migration learning, the experimental hardware configuration is i72.6GHz processor, 16 GB RAM graphics workstation, and Kinect 2.0 sensor. VS2008, OpenCV, and WindowsSDK2.0

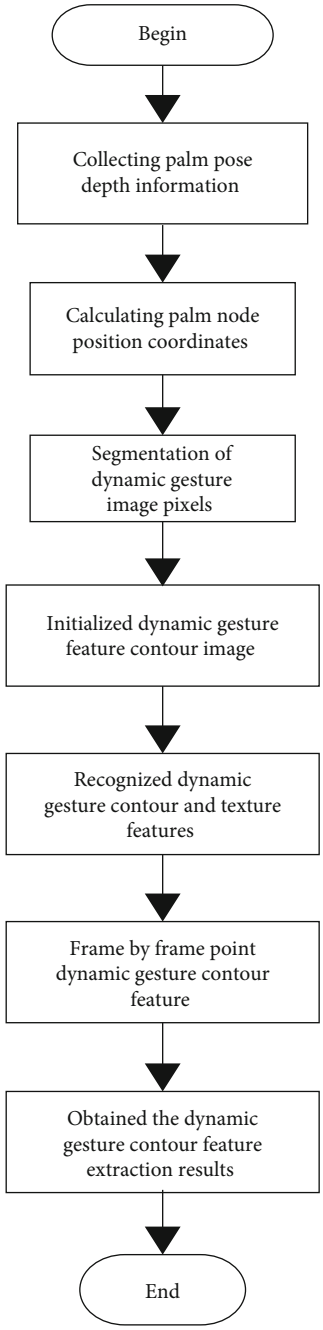


Figure 2: Dynamic gesture contour feature extraction process.

software development platforms are used to conduct simulation experiments.

The experiment uses CGD gesture recognition data set, 11k Hands data set, and HandNet data set as data sources:

(1) CGD gesture recognition data set: this data set mainly includes 24-letter images, collected by 5 people under different illumination and different height conditions.

(2) 11k Hands data set: this data set covers 11,076 hand images, the age of the collected objects is between 18 and 75 years old, and the pixels are 1600×1200 pixels. The gesture images in this data set are all taken from the back and palm sides of the hand.

(3) HandNet data set: gesture images in this data set are taken at different locations, and they are all taken from different directions using RGB-D cameras to form nonrigid deformed images.

After obtaining data from the above three data sets, distinguish similar or nonstandard gestures, and calculate the bone feature point data obtained by the proposed method, select 4 dynamic gestures with larger differences for analysis, and establish a sample library. At the same time, 18 3D feature points such as the palm, wrist, elbow, shoulder, and shoulder center were selected to analyze the gesture changes, and the influence of other regional feature points on gesture recognition was not considered. The established sample library contains human palm movements with different illumination and different heights, and a total of 400 sets of gesture image data are selected.

4.2. Evaluation Criteria

(1) Taking the first image of the hand as the object, the computer is used to draw the process of gesture image segmentation, contour recognition, contour tracing, and feature recognition, in order to show the intuitive effect of using the proposed method to extract the contour features of dynamic gestures

(2) Dynamic gesture contour feature recognition rate refers to the ratio of the number of relevant dynamic gesture contour feature points to the total number in the gesture recognition database. The calculation equation is

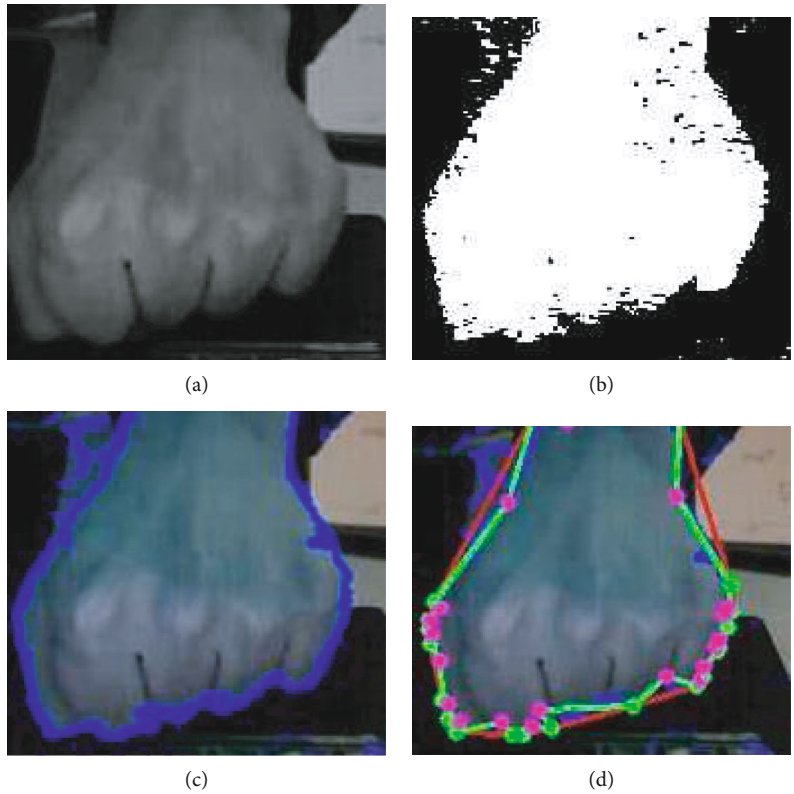
$$D_s = \frac{\alpha_z}{\sigma_z} \times 100\%, \tag{19}$$

where α_z is the number of identified contour feature points of related dynamic gestures and σ_z is the total number of contour features of dynamic gestures to be recognized

(3) In the recognition time of dynamic gesture contour feature, taking the recognition time of dynamic gesture contour feature as a criteria, combine the proposed method with Bastos et al.'s [8] method, Peng et al.'s [9] method, Tang et al.'s [10] method, Zhang et al.'s [11] method, and Liang and Liao's [12] method which are compared to verify the performance of the proposed method

(4) Dynamic gesture type recognition accuracy refers to the correctness of dynamic gesture type recognition, which reflects the accuracy of dynamic gesture contour feature recognition

$$D_z = \frac{\beta_z}{\sigma_z} \times 100\%, \tag{20}$$



(a) (b) (c) (d)

Figure 3: Dynamic gesture contour feature extraction.

where β_z is the number of correctly identified dynamic gesture contour features.

(5) The F value of dynamic gesture contour feature extraction is the weighted harmonic average of the precision rate and the recall rate, which can take into account the precision and recall rate of the method, and is a comprehensive evaluation criteria

4.3. Results and Discussion. Taking a fist holding gesture image as the original image, the proposed method is used to segment and recognize the gesture image and finally realize the contour feature extraction. The process diagram is shown in Figure 3.

It can be seen from Figure 3 that the proposed method can be input to the computer to draw the overall dynamic gesture contour feature extraction process and can effectively trace the feature contour to complete the feature extraction.

Select 800 dynamic gesture contour feature points and 4 dynamic gesture types. Among them, there are 150 contour feature points for the dynamic gesture type of hand grasping fist with index finger and middle finger extending, 230 contour feature points for the dynamic gesture type of hand grasping fist with thumb up, and the contour feature of hand grasping fist with thumb and little finger extending dynamic gesture. There are 220 points, and there are 200 contour feature points of the five-finger open palm dynamic gesture type. Select 90 sets of data as the training set, 4 sets of average values as the test samples, respectively, using the proposed method and Bastos et al.'s [8] method, Peng et al.'s [9] method, Tang et al.'s [10] method, Zhang et al.'s [11] method, and Liang and Liao's [12] method which are compared, and the recognition rate of dynamic gesture contour features of different methods is obtained. It is shown in Figure 4.

According to Figure 4. When there are 800 dynamic gesture contour feature points, the average dynamic gesture contour feature recognition rate of Bastos et al.'s [8] method is 60%, and the average dynamic gesture contour feature recognition rate of Peng et al.'s [9] method is 57%, the average dynamic gesture contour feature recognition rate of Tang et al.'s [10] method is 82%, the average dynamic gesture contour feature recognition rate of Zhang et al.'s [11] method is 76%, and the average dynamic gesture contour feature recognition rate of Liang and Liao's [12] method is 10%. The average recognition rate of dynamic gesture contour features is 91%. It can be seen that the dynamic gesture contour feature recognition rate of the proposed method is relatively high. Because the proposed method uses transfer learning to segment dynamic gesture images, feature contour images were initialized, and dynamic gesture feature recognition models were trained. Using the residual network method, each feature of the dynamic gesture contour is accurately recognized, and the influence caused by the excessively deep network layer is avoided, thereby improving the recognition rate of the dynamic gesture contour feature.

On this basis, the dynamic gesture contour feature recognition time of the proposed method is further verified, using Bastos et al.'s [8] method, Peng et al.'s [9] method,

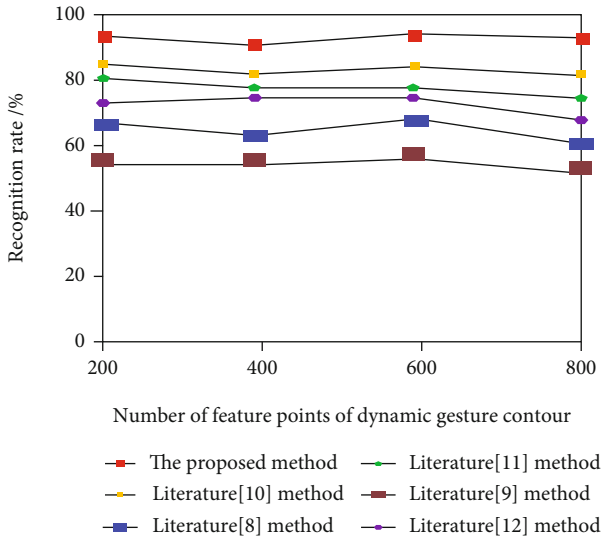


Figure 4: Comparison results of dynamic gesture contour feature recognition rate with different methods.

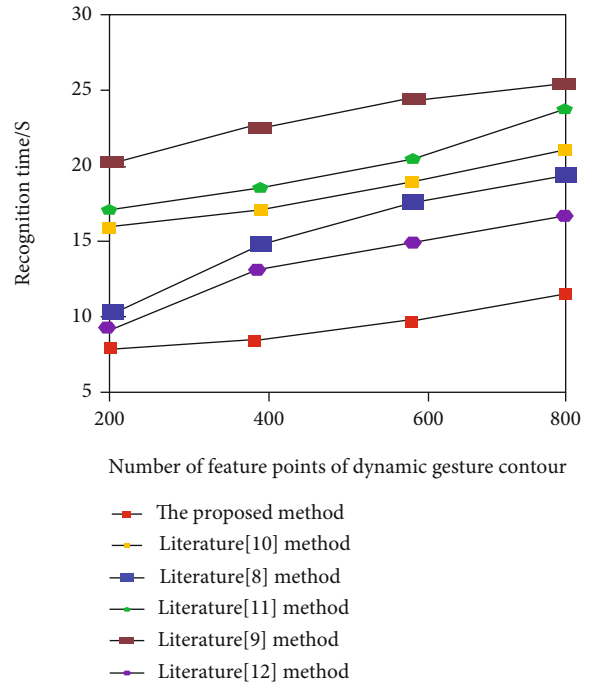


Figure 5: Recognition time comparison results of dynamic gesture contour feature with different methods.

Tang et al.'s [10] method, Zhang et al.'s [11] method, and Liang and Liao's [12] method for dynamic gesture contour feature recognition. The comparison results of different methods of dynamic gesture contour feature recognition time are obtained. It is shown in Figure 5.

According to Figure 5, as the number of contour feature points of dynamic gestures increases, the recognition time of dynamic gesture contour features increases. When the number of dynamic gesture category recognition feature points is 800, the dynamic gesture contour feature recognition time of Bastos et al.'s [8] method is 18.8 s, and the dynamic gesture contour feature recognition time of Peng et al.'s [9] method is 25.3 s, the dynamic gesture contour feature recognition time of Tang et al.'s [10] method is 19.5 s, the dynamic gesture contour feature recognition time of Zhang et al.'s [11] method is 23.8 s, and the dynamic gesture contour feature recognition time of Liang and Liao's [12] method is 16.9 s. The dynamic gesture contour feature recognition time of the proposed method is only 11.6 s. It can be seen that the dynamic gesture contour feature recognition time of the proposed method is shorter. Because the proposed method detects the distance between the dynamic gesture and the acquisition device through transfer learning and divides the dynamic gesture image to obtain the basic pixel points of each dynamic gesture image, the dynamic gesture extraction rate is increased, and the dynamic gesture contour feature recognition time is shortened.

In order to verify the accuracy of the dynamic gesture type recognition of the proposed method, Bastos et al.'s [8] method, Peng et al.'s [9] method, Tang et al.'s [10] method, Zhang et al.'s [11] method, and Liang and Liao's [12] method are compared with the proposed method, respectively. Thus, the comparison results of the accuracy of dynamic gesture type recognition of different methods can be obtained. It is shown in Figure 6.

According to Figure 6, for different dynamic gesture types, the average dynamic gesture type recognition accuracy of Bastos et al.'s [8] method is 42%, the average dynamic gesture type recognition accuracy of Peng et al.'s [9] method is 71%, and the average dynamic gesture type of Tang et al.'s [10] method. The recognition accuracy rate is 81%, the average dynamic gesture type recognition accuracy rate of Zhang et al.'s [11] method is 61%, and the average dynamic gesture type recognition accuracy rate of Liang and Liao's [12] method is 56%. The average dynamic gesture type recognition accuracy of the proposed method is 92%. It can be seen that the dynamic gesture type recognition accuracy of the proposed method is relatively high. Because the proposed method adopts the residual network method through transfer learning, the dynamic gesture contour is accurately recognized, and the processing weight is merged to obtain a significant contour response. According to the geometric characteristics of the dynamic gesture drawing area, the dynamic gesture part is projected longitudinally, and the dynamic gesture contour area is binarized, so as to effectively improve the accuracy of dynamic gesture type recognition.

In order to further verify the comprehensiveness and accuracy of the method proposed, the F value is selected as a criteria, and the proposed method is compared with the method of [8], method of [9], method of [10], method of [11] and method of [12], and the results are shown in Table 1.

The value range of F value is $[0, 1]$, and the larger the value of F in the value range, the better the output effect of the proposed method. According to Table 1, it can be seen that the average F value of the dynamic gesture contour feature extraction proposed method is 0.92, the average F value

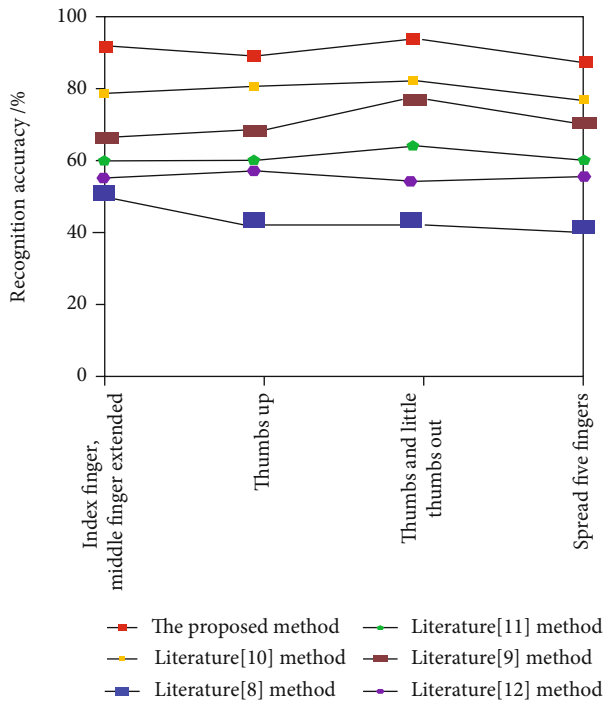


Figure 6: Comparison results of dynamic gesture type recognition accuracy with different methods.

Table 1: Comparison results of F values of different methods.

Method	F values
The proposed method	0.92
Literature [8]	0.84
Literature [9]	0.72
Literature [10]	0.84
Literature [11]	0.67
Literature [12]	0.68

of [8] is 0.84, the average F value of [9] is 0.72, and the average F value of [10]. The average value is 0.84, the average F value of [11] is 0.67, and the average F value of [12] is 0.68. It can be clearly seen that this method has absolute advantages, and the F value is large, which shows that this paper uses residual network transfer learning to extract dynamic gesture contour features, which can balance the accuracy and recall of the method and obtain better feature extraction effect.

5. Conclusions

In order to improve the recognition rate of dynamic gesture contour features and the accuracy of dynamic gesture type recognition and shorten the recognition time of dynamic gesture contour features, a dynamic gesture contour feature extraction method based on residual network migration learning is proposed. Sensors are used to integrate dynamic gesture information, transfer learning is used to detect the distance between dynamic gestures and the collection device, and dynamic gesture images are segmented in the background to obtain basic pixels of each dynamic gesture image, which shortens the time for dynamic gesture contour feature recognition. Initialize the feature contour image, train the dynamic gesture feature recognition model, and use the residual network method to accurately identify the dynamic gesture contour and texture feature, merge the processing weights, and get a significant contour response. The contour features of dynamic gestures are traced frame by frame, and the contour area of dynamic gestures is processed by gray scale and binarization, which improves the accuracy of dynamic gesture type recognition. However, the proposed method can fully suppress the contour texture and cannot achieve better results. In the future, the research on contour texture should be enhanced. Based on the frequency division characteristics of the perception field, visual processing mechanisms such as peripheral texture suppression guided by multifeature information, and frequency division visual flow fusion of the primary visual cortex should be proposed. This can not only ensure the continuity and completeness of the contour to the greatest extent but also effectively suppress texture noise.

Data Availability

The data used to support the findings of this study are included within the article. Readers can access the data supporting the conclusions of the study from CGD data set and 11k Hands data set and HandNet data set.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Ministry of Education Science and Technology Development Center Fund under grant number 2020A050116 and the Natural Science Foundation of Heilongjiang Province of China under grant number LH2021F040.

References

[1] D. Pandey, V. Namdeo, and P. Kshirsagar, "Motor imagery feature extraction cum optimization for detection of ALS disease," *Solid State Technology*, vol. 64, no. 1, pp. 739–758, 2021.

[2] F. Ye, Y. Guo, Z. Xia, Z. Zhang, and Y. Zhou, "Feature extraction and process monitoring of multi-channel data in a forging process via sensor fusion," *International Journal of Computer Integrated Manufacturing*, vol. 34, no. 1, pp. 95–109, 2021.

[3] L. Zhu, G. Wang, F. Huang, Y. Li, W. Chen, and H. Hong, "Landslide susceptibility prediction using sparse feature extraction and machine learning models based on GIS and remote sensing," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2021.

[4] B. Espejo-Garcia, N. Mylonas, L. Athanasakos, E. Vali, and S. Fountas, "Combining generative adversarial networks and agricultural transfer learning for weeds identification," *Biosystems Engineering*, vol. 204, pp. 79–89, 2021.

[5] Z. Liang and S. Liao, "Research on dynamic gesture recognition integrating wide residual and long-term and short-term memory networks," *Computer application research*, vol. 36, no. 12, pp. 3846–3852, 2019.

[6] G. Allan, I. Kang, E. S. Douglas, G. Barbastathis, and K. Cahoy, "Deep residual learning for low-order wavefront sensing in high-contrast imaging systems," *Optics Express*, vol. 28, no. 18, pp. 26267–26283, 2020.

[7] S. Luo, A. Peng, H. Zeng, X. Kang, and L. Liu, "Deep residual learning using data augmentation for median filtering forensics of digital images," *IEEE Access*, vol. 7, no. 99, pp. 80614–80621, 2019.

[8] L. O. Bastos, V. H. C. Melo, and W. Robson Schwartz, "Multi-loss recurrent residual networks for gesture detection and recognition," in *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 170–177, Rio de Janeiro, Brazil, 2019.

[9] Y. Peng, H. Tao, W. Li, H. Yuan, and T. Li, "Dynamic gesture recognition based on feature fusion network and variant ConvLSTM," *IET Image Processing*, vol. 14, no. 11, pp. 2480–2486, 2020.

[10] X. Tang, Z. Yan, J. Peng, B. Hao, H. Wang, and J. Li, "Selective spatiotemporal features learning for dynamic gesture recognition," *Expert Systems with Applications*, vol. 169, 2021.

[11] W. Zhang, J. Wang, and F. Lan, "Dynamic hand gesture recognition based on short-term sampling neural networks," *IEEE/-CAA Journal of Automatica Sinica*, vol. 8, no. 1, pp. 110–120, 2021.

[12] Z. Liang and S. Liao, "Dynamic gesture recognition based on wide residual networks and long short-term memory networks," *Application Research of Computers*, vol. 36, no. 12, pp. 3846–3852, 2019.

[13] G. Yu, H. Xiaohai, W. Xiaohong, W. Zhengyong, and Z. Yukun, "Dynamic gesture recognition method based on leap motion," *Computer system application*, vol. 28, no. 11, pp. 208–212, 2019.

[14] H. Xiao, Z. Jing, and L. Yuelong, "Gesture recognition based on geometric distribution of gestures," *Computer science*, vol. 46, no. S1, pp. 246–249, 2019.

[15] H. Huan, P. Li, N. Zou et al., "End-to-end super-resolution for remote-sensing images using an improved multi-scale residual network," *Remote Sensing*, vol. 13, no. 4, p. 666, 2021.

[16] Y. Qing and W. Liu, "Hyperspectral image classification based on multi-scale residual network with attention mechanism," *Remote Sensing*, vol. 13, no. 3, p. 335, 2021.

[17] Y. Li, S. Wang, H. He, D. Meng, and D. Yang, "Fast aerial image geolocalization using the projective-invariant contour feature," *Remote Sensing*, vol. 13, no. 3, p. 490, 2021.

[18] G. Song, "Accuracy analysis of Japanese machine translation based on machine learning and image feature retrieval," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2109–2120, 2021.

[19] M. Kowdiki and A. Khaparde, "Automatic hand gesture recognition using hybrid meta-heuristic-based feature selection and classification with dynamic time warping," *Computer Science Review*, vol. 39, 2021.

[20] L. Hao, Y. L. Lin, and Z. X. Li, "Train driver dynamic gesture recognition method based on machine vision," *Transducer and Microsystem Technologies*, vol. 40, no. 2, pp. 34–37, 2021.

[21] S. B. Reed, T. Reed, and S. M. Dascalu, "Spatiotemporal recursive hyperspheric classification with an application to dynamic gesture recognition," *Artificial Intelligence*, vol. 270, pp. 41–66, 2019.

[22] B. Huang, T. Xu, Z. Shen, S. Jiang, B. Zhao, and Z. Bian, "SiamATL: online update of Siamese tracking network via attentional transfer learning," *IEEE Transactions on Cybernetics*, vol. 1, pp. 1–14, 2021.

[23] C. Chao, "Contour feature extraction of moving image based on multi threshold optimization," *Journal of Shenyang University of technology*, vol. 41, no. 3, pp. 315–319, 2019.

[24] G. Krishnan, R. Joshi, T. O'Connor, F. Pla, and B. Javidi, "Human gesture recognition under degraded environments using 3D-integral imaging and deep learning," *Optics Express*, vol. 28, no. 13, pp. 19711–19725, 2020.

[25] Z. Xia, J. Xing, C. Wang, and X. Li, "Gesture recognition algorithm of human motion target based on deep neural network," *Mobile Information Systems*, vol. 2021, Article ID 2621691, 12 pages, 2021.