Research on using hand gesture to control manipulator based on Kinect camera

Chi Zhang, Qingsong Lin, Longlong Zhu

Henan University of Science and Technology, Luoyan471000, Henan Province, China

Abstract: To solve the problem of not accurate acquisition of external data in human computer interaction by ordinary cameras and the forward dynamics and reverse dynamics of the manipulator, we present a method to control the motion of the manipulator using the gesture recognition technology based on the Kinect camera. Use Image binarization, morphological filtering and feature extraction operations to process the captured gestures image to complete the recognition of hand gestures; use D-H parameter method to establish the mathematical model of the manipulator to analyze the motion of the manipulator; finally, send the gesture recognition results message to myRIO, myRIO processes it as the information of the joint angle of the manipulator and the control information of the end executor, so as to control the rotation of each joint of mechanical arm and the gripping and releasing of end effector.

Keywords: human-computer interaction; gesture recognition; Kinect; manipulator

1. INTRODUCTION

With the continuous development of science and technology, robots have more and more extensive applications to replace human activities and tasks that are dangerous or difficult to carry out. Generally, the operation and control of the robot is closely related to the computer. The control algorithm of the robot is developed and adjusted by the computer. The control algorithm is the core of the robot operation. Therefore, when human beings control the robot, the information exchange transmission between human and robot can smoothly transfer to the information interaction between human and computer. As an important part of human-computer interaction, gesture recognition has an great influence on the naturality and flexibility of human-computer interaction [1-2]. In practical applications, gestures are usually in a complex environment, such as bright light and dark, which will bring many uncertain factors and challenges to gesture recognition [3]. At present, using gesture recognition technology to control mechanical arm movement is mainly controlled by hardware sensors or machine vision. Compared with the machine vision, the use of hardware sensors makes a lot of inconvenience to users, so the machine visual gesture control method is more promising.

To solve above problems, this article is based on Kinect camera to design an efficient and accurate method of carving up user's gestures from the complex environment to control the mechanical arm to complete a series of activities

2. IMAGE PROCESSING AND GESTURE RECOGNITION

2.1 Image processing

The main purpose of the image preprocessing is to do some operations with collected RGB image, such as binarization processing, morphological filtering operation to reduce noises in the images, eventually extract the useful image features from gestures, to be prepared for the follow-up work. This paper mainly use the Kinect camera and the IMAQ Vison Builder module in LabVIEW for image acquisition and preprocessing.

The binarization of the image refers to converting gray or color images into a binary image by selecting the appropriate threshold. After using the IMAQ Color Threshold module in LabVIEW to set the threshold range of the image, the target image can basically reveal, but there are still some clutter in the image. This would require the median filtering method to eliminate or suppress the noise of image, the function of filtering can not only make certain characteristics more prominent, but also eliminate the noise of images, but the premise is not to destroy the image of the original information and the filtered image will have good clarity. When salt and pepper noise appeared, make the pixel of one point is different from the others, then filter out the brightest and darkest point after pixels rearrange, keep the middle point. Finally reach the purpose of removing noise

In addition also need to use the image morphology for image processing, image morphology is based on set theory method to simplify the image data, the basic operations: the erosion of the binary image, dilation, opening and closing. The purpose of expansion is to enlarge the image, and the purpose of corrosion is to reduce the image. Opening is the operation of first corrosion and then expansion of the binary image, and the closure is the operation of first expansion and then corrosion. The binary morphological opening and closing algorithm combining also can remove the noise, and the advantages of opening and closing doesn't change the target area, at the same time, smooth the edges of the target image.

After the above image preprocessing steps, we can extract the meaningful image, but in order to do qualitative or quantitative analysis of the target image, usually we also need to do some important parsing for features of the target. Common features include: NMI features, Contour moments [4], Fourier descriptors[5]. In this paper, the features of NMI are mainly analyzed, and the two-dimensional image f(x, y) can be viewed as M*N pixels, and the center of image centroid is (cx, xy), and the calculation formula is shown below.

$$cx = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} xf(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} f(x,y)} \tag{1}$$

$$cy = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} yf(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} f(x,y)} \tag{2}$$

The rotational inertia J of the image around the center of mass and the normalized rotational inertia NMI, the calculation formula are as follows.

$$J = \sum^{M}\sum^{N}((x - cx)^2 + (y - cy)^2)f(x, y) \tag{3}$$

$$NMI = \frac{\sqrt{J}}{m} = \frac{\sqrt{\sum_{x=1}^{M}\sum_{y=1}^{N}((x - cx)^2 + (y - cy)^2)f(x, y)}}{\sum_{x=1}^{M}\sum_{y=1}^{N} f(x, y)} \tag{4}$$

The m value is the sum of the gray values of all image pixels.The experiment also verified that NMI has the characteristics of rotation, scaling and translation invariance[6].

2.2 Gesture recognition

Gesture recognition is divided into static gesture recognition and dynamic gesture recognition. The gesture recognition method based on template matching method [7], dynamic time neat method (DTW) [8] hidden markov model (HMM) [9], etc., the template matching method usually for static hand gestures recognition. In the literature[10], static gesture recognition has been introduced in detail, which is not discussed here. In this paper, the dynamic gesture recognition process is analyzed in detail. The specific flow chart of gesture recognition is as follows:

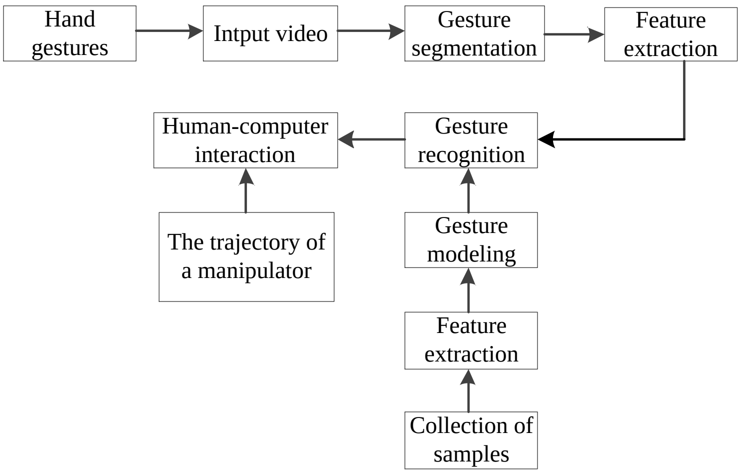

Fig.1. Gesture recognition flow chart

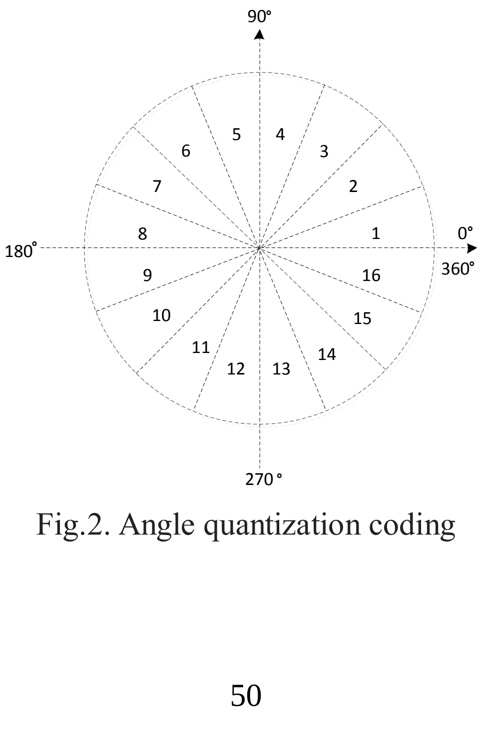

Fig.2. Angle quantization coding

Dynamic hand gestures are complex, and in many literatures, the direction angle is the most representative of gesture characteristics, and the direction angle feature of each point and the previous point is used as the dynamic gesture feature. In order to facilitate the calculation, the direction angle is divided into 16 grades to reduce the calculation. The specific description is shown in figure 2.

Because HMM can simultaneously model spatio-temporal characteristics and related characteristics, it has good recognition effect in dynamic gesture recognition system[11]. In this paper, a dynamic gesture recognition method based on HMM is adopted, and the characteristic of the change angle will as the input parameter of the HMM. The hidden Markov process is a double stochastic process, which includes two stochastic processes: implicit state transition and explicit observation output. Generally speaking, HMM model is composed of Markov chain and observation stochastic process. The Markoff chain describes the transition between the states with the implicit state transfer probability matrix A and the initial state probability matrix, thus producing the state sequence q; and the observation random process is described by the observation probability matrix and produces the observation value sequence. The model is shown in Figure 3.

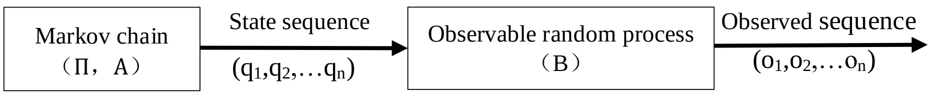

Fig.3.The composition of the HMM

A HMM can be seen as a five tuple: $\lambda = \{N, M, \Pi, A, B\}$ [12], N is the number of implicit states; M is the number of observation values (the direction angle is divided into 16 grades in the above, so the number of observations is 16); the original state transition matrix is $\Pi = (\pi_1, \pi_2, \dots, \pi_n)$

$$\pi_i = P(q_1 = S_i) \quad 1 \le i \le N \tag{5}$$

(S_i is the state probability of Q at t=1, state sequence $Q = (q_1, q_2, \dots, q_T)$,at the same time $\pi_i \ge 0$ and $\sum_{i=1}^{N} \pi_i = 1$)

A is the state transition probability matrix, the matrix element are as follow:

$$a_{ij} = P\left(q_{t+1} = S_j \middle| q_t = S_i\right) \quad 1 \le i \le N, 1 \le j \le N \tag{6}$$

$B = \{b_j(\mathrm{k})\}_{N \times M}$ is an observation probability matrix, the matrix element are as follows:

$$b_j(\mathrm{k}) = \mathrm{P}\left(\mathrm{o}_t = v_k \middle| q_t = S_j\right) \quad 1 \le j \le N, 1 \le k \le \mathrm{N} \tag{7}$$

(V is the observable value of the observed state; the observable state of the moment is $\mathrm{o}_t \in \{(v_1, v_2, \dots, v_n)\}$)

The HMM method is used to train gestures in the case of the known observation sequence O, and the model parameter of $\lambda = \{A, B, \Pi\}$ is adjusted to maximize the probability value of $P(O|\lambda)$.The basic steps are as follows:

(1) initializing hidden Markov model: initialize the state transfer probability matrix A and let the status i can only be returned to the current or transferred to the state of j=i+1.that is the Initialization of the the model parameter of $\lambda = \{A, B, \Pi\}$.

(2) The hidden state sequence of O can be found out by using Viterbi algorithm, after the initial model is determined, we calculate the $P(O|\lambda)$ under this model. The initial HMM model is recalculated by the Baum Welch algorithm, and the parameters of the HMM model are reestimated in this process, and we will get the new λ: $\bar{\lambda} = \{\bar{A}, \bar{B}, \bar{\Pi}\}$. Then the Viterbi algorithm is used to calculate the $P(O|\bar{\lambda})$ and $P(O|\bar{\lambda}) - P(O|\lambda)$, and comparing the magnitude of the difference value and the convergence threshold value. If the value is less than convergence threshold ε, the convergence $P(O|\bar{\lambda})$is obtained, otherwise the repeated calculation will be done until $P(O|\bar{\lambda})$ is convergent, so the trained HMM model parameters and the hidden Markov model which are closest to the observation value are obtained .

And HMM's optimization algorithm can effectively improve accuracy and real-time performance[13] .

3. DESIGN OF MOTION CONTROL SYSTEM FOR MECHANICAL ARM

In this paper, the results of gesture recognition are used to control the movement of the manipulator, that requires a motion analysis of the arm first. The kinematics of manipulator is divided into two aspects: forward and inverse kinematics[14]. The forward kinematics problem is to find the position and attitude of the end-effector of the manipulator relative to the reference frame in the premise of knowing the geometrical parameters and joint variables of the connecting rod. While the inverse kinematics problem is to find the joint variables of the manipulator under the premise of knowing the desired position of the end-effector.

In general, the D-H method [15] is used to establish the coordinate system of the manipulator model, for each link of the manipulator, a coordinate system is established to describe the relative positions of these coordinates with the homogeneous transformation. By recursively obtaining the homogeneous transformation matrix of the terminal actuator relative to the basis coordinate system, the motion equation of the mechanical arm will be obtained, and the description of the space state of the manipulator is represented by position and attitude [16].

The local coordinate system is established at each joint first, and then the relative position and posture of each joint are obtained through the coordinate transformation, finally, we use the base frame to represent the end-frame of the actuator. The transformation from j coordinate system of joint unit to j+1 is shown in Fig. 4.

First, we rotate about θ_j degrees around the z_j axis, so that the x_j axis is parallel to the x_{j+1} axis; then move $d_j$parallel to the z_j axis, make the x_j axis on a line with the x_{j+1}axis; and move L_{j+1} parallel to the x_{j+1} axis, make the z_j axis parallel to z_{j+1} axis; finally, we rotate about α_{j+1} degrees around the x_{j+1}, make the z_j axis on a line with the z_{j+1}axis; and the coordinate matrix ${}^{j+1}_{j}T$ (the pose matrix of frame {j+1} relative to {j})is derived by rotation matrix and transfer matrix, the formula is derived as follows:

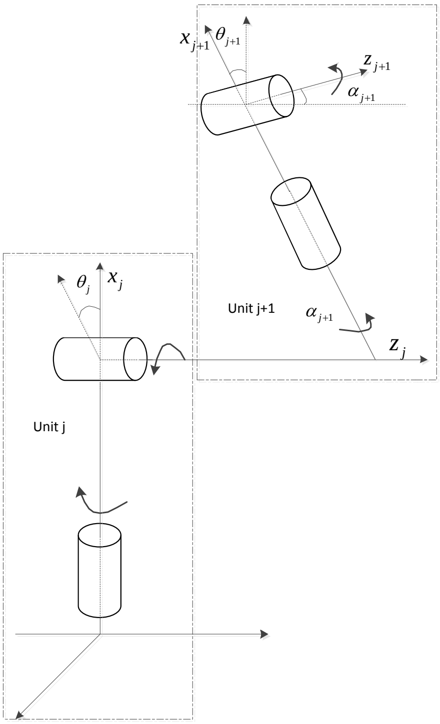

Fig.4. Relationship between adjacent joint units.

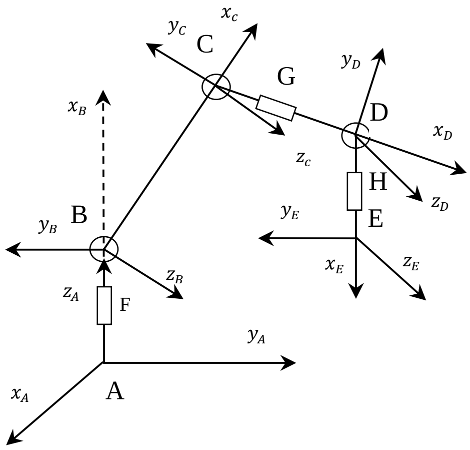

Fig. 5. Simplified mechanical arm structure.

$$
{}^{j+1}_{j}T = Rot(z,\theta_j)\cdot trans(0,0,d_j)\cdot trans(L_{j+1},0,0)\cdot Rot(x,\alpha_{j+1})
= \begin{pmatrix} cos\theta_j & -sin\theta_j cos\alpha_{j+1} & sin\theta_j sin\alpha_{j+1} & L_{j+1}cos\theta_j \\ sin\theta_j & cos\theta_j cos\alpha_{j+1} & -cos\theta_j sin\alpha_{j+1} & L_{j+1}sin\theta_j \\ 0 & sin\alpha_{j+1} & cos\alpha_{j+1} & d_j \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (8)
$$

According to the above coordinate equation (8), we can deduce the position coordinate transformation of unit j relative to unit i, the formula is as follows:

$$
{}^{j}_{i}T = {}^{i+1}_{i}T \cdot {}^{i+2}_{i}T \cdots {}^{j}_{j-1}T \quad (9)
$$

The base coordinate system is used to represent the end-frame of the actuator by the method of recursive again, then the homogeneous transformation matrix of the end-effector of the manipulator is obtained, and this is the solution to the forward motion equation. Due to the six degrees of freedom manipulator used in this paper, in order to simplify the analysis calculation, we simplified the characteristics of the structure unit of the manipulator as shown in the figure below:

Compared with the forward kinematics problem, the inverse problem is more difficult to solve, and the existence of the inverse kinematic solution depends on whether the desired position is within the range of the robot end-effector. The calculation of inverse kinematics is large and each joint solution may not be unique, so we need to find the optimal solution by fully consider. In this paper, the inverse kinematics problem of joint angle is determined based on the known terminal position ${}_A^ET$, the formula is derived as follows:

$$ {}_A^ET = {}_A^BT(\alpha_1) \cdot {}_B^CT(\theta_1) \cdot {}_C^DT(\theta_2, \alpha_3) \cdot {}_D^ET(\theta_3, \alpha_4) = \begin{pmatrix} n_x & o_x & a_x & E_x \\ n_y & o_y & a_y & E_y \\ n_z & o_z & a_z & E_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (10) $$

The geometric solution is to solve the equation by using geometric knowledge of the mechanical arm, and the solving steps are as follows:

(1) Calculation of D point coordinates

From the formula (10) we can get the E coordinate value $E = \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}$, and the azimuth matrix $R_E = \begin{pmatrix} n_x & o_x & a_x \\ n_y & o_y & a_y \\ n_z & o_z & a_z \end{pmatrix}$, and $\vec{n} = \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix}$ is the direction vector that relative to the polar coordinate system{A} and $\overrightarrow{DE}$, the expression is as follows:

$$ \overrightarrow{DE} = E - D = L_4 \cdot \vec{n} \quad (11) $$

This can also deduce the coordinates of the D point:

$$ \mathrm{D} = \mathrm{E} - L_4 \cdot \vec{n} \quad (12) $$

(2) Calculation of the joint angle of α_1

From Figure 5, we can know that A, B, C and D four points on one plane and α_1 is located in the coordinate system{A}, the rotation angle of yAz plane to this plane. From the formula (12) we can get that $D = (D_x, D_y, D_z)^T$, then the formula of α_1 is shown as follows:

$$ \alpha_1 = \arctan\left(-\frac{D_x}{D_y}\right) + k\pi, k \in z \quad (13) $$

(3) Calculation of the joint angle of θ_1

From Figure 5, we can know that the C and D points are all in the xBy plane. Based on the obtained D point coordinates, the coordinates of the coordinates of the {B} coordinate system are obtained by coordinate transformation. The formula is shown below:

$$ {}_BD = ({}_A^BT)^{-1} \cdot D \quad (14) $$

assuming that: ${}_BD = ({}_BD_x, {}_BD_y, 0)^T$, ${}_BC = (L_2 cos\theta_1, L_2 sin\theta_1, 0)^T$,

by known conditions: $L_3 = |CD| = |{}_BD - {}_BC|$

so we can get that:

$$(L_2cos\theta_1 - {}_BD_x)^2 + (L_2sin\theta_1 - {}_BD_y)^2 = {L_3}^2 \tag{15}$$

Then, $\theta_1$can be obtained from the following formula:

$$\theta_1 = \pm arccos\frac{{L_2}^2 + |BD|^2 - {L_3}^2}{2L_2 \cdot |BD|} + \varphi + 2k\pi\,, k \in Z \tag{16}$$

and $|BD| = \sqrt{{}_BD_x + {}_BD_y}$,$\cos\varphi = \frac{{}_BD_x}{|BD|}$,$\sin\varphi = \frac{{}_BD_y}{|BD|}$.

(4) Calculation of the joint angle of θ_2

By known conditions: ${}_B^CT = Rot(z,\theta_1) \cdot trans(L_2, 0,0) \cdot Rot(x,\alpha_2)$,by using matrix coordinate transformation, we can get the coordinates of D in {C} which are shown as below:

$${}_CD = ({}_B^CT)^{-1} \cdot {}_BD \tag{17}$$

assuming that${}_CD = ({}_CD_x\,, {}_CD_y\,, {}_CD_z)^T$,then:

$$\begin{cases} {}_CD_x = L_3cos\theta_2 \\ {}_CD_y = L_3sin\theta_2 \end{cases} \tag{18}$$

The formula for calculating θ_2 is as follows:

$$\theta_2 = sgn({}_CD_y) \cdot \arccos(\frac{{}_CD_x}{L_3}) + 2k\pi\,, k \in Z \tag{19}$$

(5) Calculation of the joint angle of α_3

By known conditions:$\overrightarrow{a_D} = \pm\overrightarrow{CD} \times \overrightarrow{DE}$, since z_C and z_B is in the same azimuth, we can get that:$\overrightarrow{a_c} = (cos\alpha_1\,, sin\alpha_1\,, 0)^T$

And we know that by the vector inner product:

$$a_D \cdot a_c = |\overrightarrow{a_D}| \cdot |\overrightarrow{a_c}|cos\alpha_3 \tag{20}$$

The formula for calculating α_3 is as follows:

$$\alpha_3 = sgn\left((\overrightarrow{a_D} \times \overrightarrow{a_c}) \cdot \overrightarrow{CD}\right)\arccos\left(\frac{\overrightarrow{a_D} \cdot \overrightarrow{a_c}}{|\overrightarrow{a_D}| \cdot |\overrightarrow{a_c}|}\right) + k\pi\,, k \in z \tag{21}$$

(6) Calculation of the joint angle of θ_3

Similar to the derivation of α_3, the solution of $\theta_3$and α_4 is shown in the following formula:

$$\theta_3 = sgn\left((\overrightarrow{CD} \times \overrightarrow{DE}) \cdot \overrightarrow{a_D}\right) \cdot \arccos\left(\frac{\overrightarrow{CD} \cdot \overrightarrow{DE}}{L_3 \cdot L_4}\right) + k\pi\,, k \in Z \tag{22}$$

$$\alpha_4 = sgn\left((\overrightarrow{Z_D} \cdot \overrightarrow{Z_E}) \cdot \overrightarrow{DE}\right) \cdot \arccos\left(\frac{\overrightarrow{Z_D} \cdot \overrightarrow{Z_E}}{|\overrightarrow{Z_D}| \cdot |\overrightarrow{Z_E}|}\right) + 2k\pi\,, k \in Z \tag{23}$$

So far, all the joint angles of the manipulator are obtained, and the inverse kinematics problem is solved completely. The control mode of the manipulator adopts the master-slave control mode, the main control machine is an ordinary computer, while assistant controller is the myRIO hardware platform. The whole system process is to use the Kinect camera to collect gestures first, then the computer carries out the image processing and completes the gesture recognition. According to the computer processing results, the information of the joint angle of the manipulator and the control information of the end executor will be obtained, thus the steering gear of the manipulator is actuated. The master-slave control system framework is shown in the figure below.

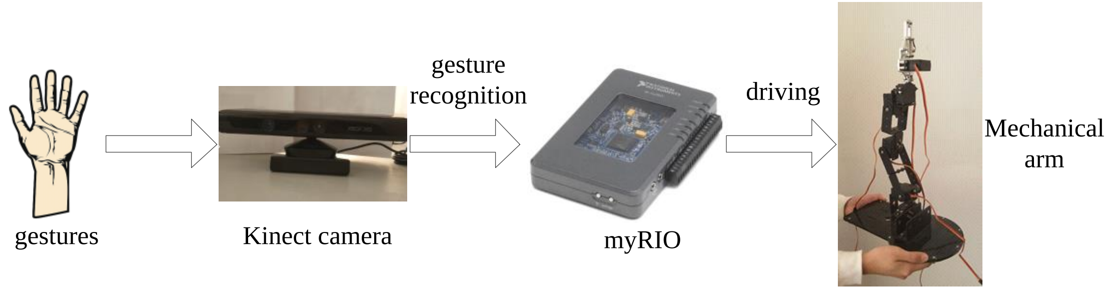

Fig.6. The master-slave control system framework

4. CONCLUSION

In this paper, gesture technology is applied to the research of manipulator control so that the user can directly manipulate the trajectory of the manipulator through gestures. And this non-contact man-machine interaction has great application value in this system.

(1) The original image is collected by using the Kinect camera, then segment the required image from the original image, finally the required image is extracted by filtering and morphological processing techniques.

(2) The gesture recognition is divided into two parts, and the static gestures are matched by the template. The dynamic gestures are identified by the HMM method based on the time and space trajectory, each of them is analyzed at the end.

(3) In the kinematics processing of the manipulator, the mathematical model of the manipulator is described first, and on this basis, the forward kinematics and inverse kinematics of the manipulator are analyzed and calculated.

(4) The control mode of the manipulator adopts the master-slave control mode, the main control machine is an ordinary computer, it is mainly responsible for image acquisition and gesture recognition, and send the results of processing to myRIO through WIFI,t hen the steering gear is actuated from the control machine myRIO.

REFERENCES

[1] Ghani J A,Deshpande S P. Task characteristics and the experience of optimal flow in human-computer interaction[J]. Journal of Psychology, 2016,128(4):381-391.

[2] Gupta H P,Chudgar H S,Mukherjee S,et al. A Continuous Hand Gestures Recognition Technique for Human-Machine Interaction Using Accelerometer and Gyroscope Sensors[J]. IEEE Sensors Journal,2016,16(16):6425-6432.

[3] Moreira T,Alcantara M,Pedrini H,et al. Fast and Accurate Gesture Recognition Based on Motion Shapes[C]// Iberoamerican Congress on Pattern Recognition. Springer International Publishing,2015: 247-254.

[4] Gu J S,Silva C W D. Development and implementation of a real-time open-architecture control system for industrial robot systems[J]. Engineering Applications of Artificial Intelligence,2004,17(5):469-483.

[5] Zhao G, Qin S, Wang D. Interactive segmentation of texture image based on active contour model with local inverse difference moment feature[J]. Multimedia Tools & Applications, 2018(4): 1-28.

[6] Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework.[J]. Journal of Computational & Graphical Statistics, 2006, 15(3): 651-674.

[7] Duan H, Xu C, Liu S, et al. Template matching using chaotic imperialist competitive algorithm[J]. Pattern Recognition Letters, 2010, 31(13): 1868-1875.

[8] Hartmann B, Link N. Gesture recognition with inertial sensors and optimized DTW prototypes[C]// IEEE International Conference on Systems Man and Cybernetics. 2010: 2102-2109.

[9] Raheja J L, Minhas M, Prashanth D, et al. Robust gesture recognition using Kinect: A comparison between DTW and HMM[J]. Optik - International Journal for Light and Electron Optics, 2015, 126(11–12): 1098-1104.

[10] Liu Y,Gan Z,Sun Y. Static Hand Gesture Recognition and its Application based on Support Vector Machines[C]. Ninth Acis International Conference on Software Engineering,Artificial Intelligence,Networking,and Parallel/distributed Computing. IEEE Computer Society,2008:517-521.

[11] Godoy V, Britto A S, Koerich A, et al. An HMM-Based Gesture Recognition Method Trained on Few Samples[C]// IEEE, International Conference on TOOLS with Artificial Intelligence. IEEE Computer Society, 2014: 640-646.

[12] Guo X L, Yang T T. Gesture recognition based on HMM-FNN model using a Kinect[J]. Journal on Multimodal User Interfaces, 2016, 11(1): 1-7.

[13] Jung H J, Park H, Kim D. Improvement of Gesture Recognition using 2-stage HMM[J]. Journal of Institute of Control Robotics & Systems, 2015, 21(11): 1034-1037.

[14] Sun L, Jiang M A, Xiao-Gang R. Trajectory Planning and Simulation of 6-DOF Manipulator[J]. Control Engineering of China, 2010, 17(3): 388-392.

[15] Manocha D,Canny J F. Efficient inverse kinematics for general 6R manipulators[J]. IEEE Trans on Robotics & Automation,1994,10(5):648-657.

[16] Cai C, Somani N, Knoll A. Orthogonal Image Features for Visual Servoing of a 6-DOF Manipulator With Uncalibrated Stereo Cameras[J]. IEEE Transactions on Robotics, 2016, 32(2): 452-461.