Research on Gesture Recognition of Smart Data Fusion Features in the IoT

Chong Tan[1], Ying Sun[1,2], Gongfa Li[1,2,4*], Guozhang Jiang[2,3], Disi Chen[5], Honghai Liu[5]

[1] Key Laboratory of Metallurgical Equipment and Control Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China;
[2] Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China;
[3] Research Center of Biologic Manipulator and Intelligent Measurement and Control, Wuhan University of Science and Technology, Wuhan 430081, China;
[4] Institute of Precision Manufacturing, Wuhan University of Science and Technology, Wuhan 430081, China;
[5] School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK;
*Corresponding Author Email: ligongfa@wust.edu.cn

Abstract: With the rapid development of Internet of Things (IoT) technology, the interaction between people and things has become increasingly frequent. Use simple gestures instead of complex operations to interact with the machine, the fusion of smart data feature information and so on has gradually become a research hotspot. Considering that the depth image of the Kinect sensor lacks color information and is susceptible to depth thresholds, this paper proposes a gesture segmentation method based on the fusion of color information and depth information, in order to ensure the complete information of the segmentation image, a gesture feature extraction method based on Hu invariant moment and HOG feature fusion is proposed, by determining the optimal weight parameters, the global and local features are effectively fused. Finally, the SVM classifier is used to classify and identify gestures. The experimental results show that the proposed fusion features method has a higher gesture recognition rate and better robustness than the traditional method.

Keywords: Gesture recognition; Fusion Features; Smart Data Aggregation; Hu moment; SVM

1 Introduction

With the development of Internet technology and communication technology, the Internet of Things technology has gradually been developed and applied. The Internet of Things technology mainly refers to the development of corresponding functions in the real world to transmit, process and execute smart data. This requires the Internet of Things to have the corresponding computing power and perceptual power, so that these real-world data can be converted into smart data, and ultimately achieve mutual interaction between people and things[1-2]. The research background of this paper is the somatosensory interaction technology based on the overall environment of the Internet of Things. At present, research on gesture recognition technology has become an important research direction in digital image processing, artificial intelligence, computer vision, pattern recognition and other related fields[3]. In recent years, the gesture recognition method based on Kinect sensor has been widely used in gesture recognition because it can separate gestures from complex backgrounds and is less affected by illumination, and can accurately track and locate gesture motions. However, the Kinect sensor needs to further improve in the resolution of the depth image and the lack of color information and the recognition of complex gesture movements. Therefore, this paper studies and analyzes the gesture recognition based on Kinect for the above problems.

The first part of this paper is about the segmentation of gesture images, and then proposes an image segmentation method based on the fusion of depth information and color information, the Kinect sensor is used to track and locate the gesture area of interest, and then the depth image is displayed in color. The second part is about the feature extraction of gesture images. To ensure the extraction of all the information of the

segmentation image gestures, the corresponding features are extracted from the global and local parts of the image. The third part is about the dimension reduction of the feature data. Because the extracted HOG feature dimension is high, the calculation is cumbersome when classifying, this paper uses PCA to reduce the dimension. The fourth part is about the experiment and result analysis of gesture recognition, which proves the superiority of this method from the recognition performance of gesture under different scales, different rotation angles and different illumination conditions.

2 Related Works

Gesture recognition methods based on data gloves have been widely studied for a long time. This method mainly uses gloves to measure various kinds of hand information, and then uses these measurements for gesture modeling and recognition. In recent years, the use of data gloves for gesture recognition has triggered a wave of gesture recognition research[4-5]. Some scholars propose a Kinect-based gesture recognition method that uses FEMD algorithm to achieve stable static gesture recognition, but the algorithm requires a large of data for training[6-7]. Some researchers have combined BPNN and PSO algorithms on the basis of traditional algorithms, which greatly reduces the training time and improves the accuracy of gesture recognition[8]. Although the data glove-based gesture recognition method has high recognition rate and high speed, but due to the different collection of individual and the layout of the electrode position, the data glove is very inconvenient in practical application, which greatly affects the effect of classification and recognition. At this stage, it is mainly applied to the training of myoelectric artificial hand.

Recently, because EMG signal is not easily affected by the external environment, and has better real-time performance, many researchers have begun to study EMG control. Yinfeng Fang et al.[9] independently developed a new type of electromyography electrode arrangement, using 16-channel surface EMG signal for gesture acquisition, and studied the relationship between EMG and human hand movement in the two-dimensional test chart. Some people use the patch electrode attached to the arm to measure the EMG signal when writing, and use the Dynamic Time Warping (DTW) algorithm matching algorithm to complete the writing recognition[10]. In the gesture recognition of EMG signals, many studies have shown that Hidden Markov Model (HMM) and artificial neural network methods can improve the recognition rate of gestures, but HMM methods and neural network classifiers need to train a large number of samples and algorithms is complex, so not suitable for fast gesture recognition applications.

Kinect sensor can collect RGB color image and depth image, and use Kinect platform to carry out human-computer interaction, then use computer graphics technology, machine learning algorithm and intelligent control theory to recognize and classify gestures, and make corresponding response. Therefore, this paper studies the recognition of gestures based on Kinect sensing equipment, which can be roughly divided into the following processes: Gesture segmentation, tracking and feature extraction, and classification recognition.

Since the Kinect sensor has been applied to the field of gesture recognition, the Kinect-based gesture recognition method has achieved good recognition results. However, the Kinect sensor still needs to be improved on the problems of the lack of resolution and color information of the depth image and the recognition of complex gesture movements. The resolution of the depth image is related to the depth value. If the closer to the sensor, the depth data cannot be obtained. If the sensor is too far away, the resolution is lower and the color information cannot be retained. The more complicated the gesture motion, the lower the recognition rate. Therefore, the combination of multiple methods to achieve segmentation and feature extraction and recognition of complex gestures in different environments needs further study.

3 Gesture Image Segmentation

Gesture segmentation is to separate the gesture area from the acquired image and other unrelated backgrounds. Commonly used static gesture segmentation methods include background subtraction, skin color model and depth information[11-12]. Among them, the background subtraction method is extremely vulnerable

to the limitation of background movement, and it is difficult to perform the task of gesture segmentation. The skin color model method is simple and has high real-time performance, and can adapt to changes in gesture shape. Since skin color features are not affected by rotation, scaling, etc., the segmentation is not subject to wear restrictions. And the skin color has a clustering property in a specific color space, however, it is highly susceptible to the influence of light and skin-like background. The depth information method has extremely high definition and resolution for images acquired in the visual range, and can accurately distinguish the background from the object to be measured, and the noise it produces is small and can be ignored, but it is easy to be affected by the depth value and lose the color information of the object[13-14]. Aiming at the advantages and disadvantages of the above gesture segmentation method, this paper adopts the segmentation method combining depth information and color information.

3.1 Segmentation of gestures of interest regions

The color image and the depth image are acquired by the Kinect sensor, and the depth value of each pixel of the depth image is stored in the two-dimensional array $depthing[i][j]$, and the following determination is made:

$$\begin{cases} If\,(depthing[i][j] \le X \;\&\&\; depthing[i][j] \ge Y) \\ depthing[i][j] = 1; \\ Else\;\; depthing[i][j] = 0; \end{cases} \tag{1}$$

Where X and Y represent the farthest and the nearest depth distance of the region of interest to the Kinect depth sensor, and the pixel value between the depth threshold $[X, Y]$ is 1, otherwise 0. The depthing $[i][j]$ two-dimensional array is mapped with the depth image[15], and then the gesture coordinate points in the depth coordinate space after the mapping process are converted into a color coordinate space, and finally displayed in color, as shown in Figure 1.

(a) Color image (b) Depth image (c) Segmentation image

Fig.1 Segmentation the region of interest

3.2 Selection of color space

The color space can be represented by a three-dimensional coordinate system. Each color is fixed in the position of the determined color space. In different color spaces, the way of representing the image is different[16]. The HSV color space and the YCbCr color space are two commonly used color spaces.

YCbCr color space is commonly used in digital video, where luminance information is stored separately in Y, and chrominance information is stored in Cb and Cr. Cb reflects the difference between the blue component and the luminance value, and Cr reflects the difference between the red component and the luminance value[17-20]. The Y value ranges from 16 to 235, and the Cb and Cr ranges from 16 to 240. The conversion formula of the RGB color space and the YCbCr color space is as shown in equation (2).

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2}$$

Compared with the HSV color space, the Y component in the YCbCr color space is also independent of the Cb and Cr components. The conversion of YCbCr and RGB color space is linear, and the calculation is relatively easy. In addition, the clustering effect of skin color in the YCbCr color space is more compact and easier to segment than the HSV color space[21-23]. So this paper chooses to segment the gesture image in the YCbCr color space.

3.3 Selection of skin color model

Under the single white background, a small number of holes will appear in the threshold model segmentation gesture, and more background will appear in the Gauss model segmentation. But the foreground gesture image segmentation is better than the threshold model. The foreground gesture segmentation effect of elliptic model and threshold model is similar and the background segmentation effect is the best. In different non white backgrounds, the Gauss model has the worst segmentation effect, while the elliptical model is the best. Based on the background of gesture image segmentation in this paper, the elliptic model is selected to segment the gesture image of depth extraction.

After selecting the YCbCr color space, this paper uses the statistical principle to analyze the skin color. By detecting the skin color area of all pixels in the test statistics image, then the image is segmented according to the distribution of the pixel points of each image in the skin color area measured in the test[24].

A large number of Statistics found on the skin color in the YCbCr color space, the distribution of the ellipse in the $Cb-Cr$ coordinates to fit the color is more appropriate[25], and obtains the calculation formula of elliptical model:

$$\begin{cases} \dfrac{(x-C_x)^2}{a^2} + \dfrac{(y-C_y)^2}{b^2} = 1 \\ \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} Cb - x_0 \\ Cr - y_0 \end{bmatrix} \end{cases} \tag{3}$$

After statistical analysis, the parameters in the $Cb-Cr$ plane of the model are: the center of the ellipse is $Cb_0 = 124.53,\ Cr_0 = 135.30$, the angle of inclination of the ellipse is $\theta = 2.27$ (radian), and the major and minor axes of the ellipse are $a = 29.99, b = 17.62$, $C_x = 1.83,\ C_y = 2.67$, respectively. The skin color judgment criteria are set to:

$$D(Cb,Cr) = \begin{cases} 1 & \dfrac{(x-C_x)^2}{a^2} + \dfrac{(y-C_y)^2}{b^2} \le 1 \\ 0 & other \end{cases} \tag{4}$$

As shown in Figure 2, elliptic curve is established on the skin color distribution by the elliptic model equation. It can be seen that most of the skin color points are included in the elliptical region, and the blank region is relatively small, and the elliptic model is relatively fast[26-28]. The elliptical model uses elliptical regions to contain skin color pixels[29], and the segmentation effect is best. Therefore, this paper selects the ellipse model to segment the depth-extracted gesture image, and the segmentation effect is shown in Figure 3.

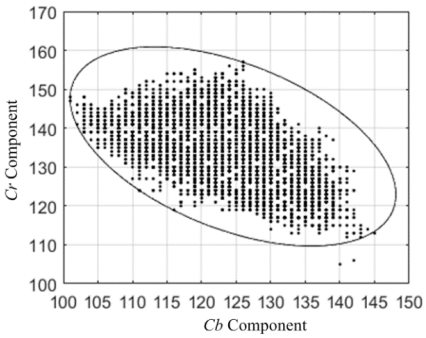


Fig.2 Elliptical model

(a) Depth segmentation image (b) Elliptical skin segmentation image

Fig.3 Elliptical model segmentation effect

3.4 Post-processing of gesture images

After the original image is segmented by the depth segmentation and the elliptical skin model, a binary image of the gesture image with a large number of backgrounds is obtained, but there are burrs on the gesture boundary or holes in the gesture area, which will interfere with subsequent feature extraction and classification operations[30-33]. Therefore, it is necessary to perform morphological processing and image enhancement.

In order to reduce the subsequent interference on the feature extraction of the gesture image, a gesture of elliptical model image segmentation is performed in the YCbCr color space expansion processing, and it can eliminate a small amount of holes in the gesture[34], the processing effect is shown in Figure 4 (a). Secondly, median filtering is performed, the contour of the gesture edge becomes very smooth[35], and the processing effect is shown in Figure 4(b). Then, the etching process is performed to restore the original size of the gesture, so that the gesture contour is more rounded[36], and the processing result is shown in Figure 4(c). Finally, the outline of the gesture image is extracted as shown in Figure 4 (d).

(a)Expanded image (b) Smooth image (c) Corrosion image (d) Contour image

Fig.4 Morphological processing

4 Feature Extraction of Gesture Images

The gesture binary image segmented by the elliptical model has been post-processed to reduce a lot of useless information. In order to be able to classify and recognize different gestures in the following, it is necessary to extract features, and the extracted data can reflect information such as shape and structure of the image[37-38]. In this paper, the corresponding features are extracted from the global and local parts of the image.

4.1 Hu invariant moments

The global feature extracted in this paper is the Hu invariant moment. The gray value of the image is regarded as a two-dimensional probability density distribution function. By calculating the corresponding geometric moment, the characteristics of translation, rotation and scale invariance can be obtained. This feature represents the geometric shape of the image and the calculation speed is faster[39].

For two-dimensional continuous function $f(x,y)$, the $(p+q)$ order moment is:

$$m_{pq}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^p y^q f(x,y)dxdy \qquad p,q=0,1,2,\mathrm{L} \tag{5}$$

If $f(x,y)$ is the pixel value of the digital image and the image size is $M\times N$, the integral operation needs to be converted into a sum operation, and the above formula becomes:

$$m_{pq}=\sum_{x=1}^{M}\sum_{y=1}^{N}x^p y^q f(x,y) \qquad p,q=0,1,2,\mathrm{L} \tag{6}$$

The corresponding $(p+q)$ -order center moment is:

$$\mu_{pq}=\sum_{x=1}^{M}\sum_{y=1}^{N}(x-\bar{x})^p(y-\bar{y})^q f(x,y) \qquad p,q=0,1,2,\mathrm{L} \tag{7}$$

Where $\bar{x}=\frac{m_{10}}{m_{00}},\bar{y}=\frac{m_{01}}{m_{00}}$ represents the centroid of the image.

The normalized $(p+q)$ -order central moment is:

$$\eta_{pq}=\frac{\mu_{pq}}{\mu_{00}^{\rho}},\ \ \rho=\frac{p+q}{2}+1 \tag{8}$$

The normalized first-order, second-order and third-order central moments are derived by nonlinear combination to obtain the seven invariant moment features of the image:

$$\begin{cases}
M_1=\eta_{20}+\eta_{02}\\
M_2=(\eta_{20}-\eta_{02})^2+4\eta_{11}^{\ 2}\\
M_3=(\eta_{30}-3\eta_{12})^2+(3\eta_{21}+\eta_{03})^2\\
M_4=(\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2\\
M_5=(\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})((\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2)\\
\qquad+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})(3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2)\\
M_6=(\eta_{20}-\eta_{02})((\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2)\\
\qquad+4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03})\\
M_7=(3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})((\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2)\\
\qquad-(\eta_{30}-3\eta_{12})(\eta_{21}+\eta_{03})(3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2)
\end{cases} \tag{9}$$

4.2 Hog feature

The local feature extracted in this paper is the HOG feature. The details and texture of the image are depicted by utilizing the distribution of gradients within the local regions of the image. When extracting the feature, the entire image is first divided into a plurality of non-overlapping small units. Then weighting the

gradient magnitudes of the gradient directions of the elements and combining the histograms of all the elements to describe the image. Finally, the individual units are spliced into larger blocks, which are normalized in the block to ensure the robustness to illumination. The extraction process of HOG features is as follows:

(1) Calculate the gradient and direction of the image. The original image is convoluted by the horizontal differential operator [-1,0,1] and the vertical differential operator $[-1,0,1]^T$ to obtain the horizontal gradient $Gx(x,y)$ and the vertical gradient $Gy(x,y)$, and according to this calculation gradient magnitude $G(x,y)$ and gradient direction $\alpha(x,y)$.

$$G_x(x,y) = f(x+1,y) - f(x-1,y) \tag{10}$$

$$G_y(x,y) = f(x,y+1) - f(x,y-1) \tag{11}$$

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{12}$$

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{13}$$

(2) Construct a gradient direction histogram. First, the image is divided into a plurality of cells that do not overlap, the gradient direction is divided into multiple inter cell channels. Then, the gradient magnitude of the unit pixel points is mapped into the corresponding gradient direction interval range to be accumulated, and the gradient direction histogram of the unit is obtained. That is, if it is divided into 9 intervals, the gradient amplitude of one pixel in a cell is 0.02, and the gradient direction is 20°, then the first interval is increased by 0.02, as shown in Figure 5.

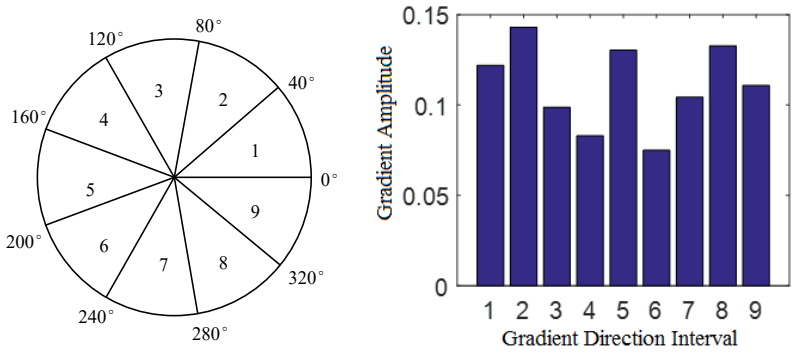


Fig.5 Gradient direction division and gradient amplitude accumulation of a unit

(3) Normalization is performed within the block to generate a feature description vector. A certain number of cells are selected to be combined into blocks that overlap each other. As shown in Figure 6, each 2×2 cell is combined into one block. Then, in order to remove the influence of illumination and shadow, l_2 -norm normalization is needed for the histogram in each block. Finally, the histograms in all blocks are combined into feature vectors to form HOG features[40].

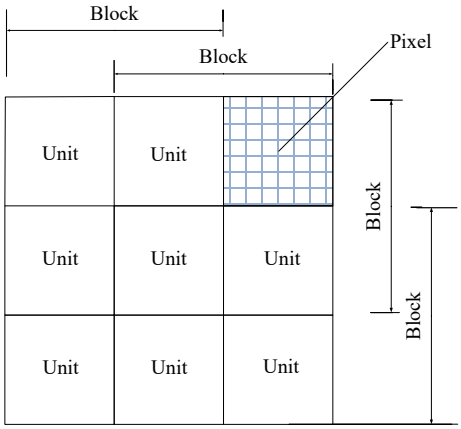


Fig.6 Schematic diagram of the structure of the unit and block

5 Dimensionality Reduction of Feature Data

Since the extracted HOG feature dimension is high, the comparison calculation is cumbersome when classifying, therefore, this paper uses principal component analysis to reduce its dimensionality[41-42]. The process of PCA feature dimension reduction is as follows:

Suppose there are n sample data in the training sample, each data has m dimensions, and a_i is the ith sample ($i = 1,2,\mathrm{L}\ ,n$). The mean of each sample:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{14}$$

Calculate the difference between each sample data and the mean:

$$d_i = a_i - \mu \tag{15}$$

The covariance matrix for all samples is:

$$C = \frac{1}{n}\sum_{i=1}^{n} d_i d_i^T = \frac{1}{n} DD^T \tag{16}$$

The eigenvalues and eigenvectors of the covariance matrix are obtained by using the singular value decomposition theorem of the matrix:

$$C = USV^T = USU^T \tag{17}$$

Where U and V are two m-order Unitary Matrices, since C is a square matrix, so U=V. The value on the diagonal of the S matrix is the feature value, and U is the corresponding feature vector. The feature values are arranged in descending order and the feature vector U is reorganized, extracting the feature vector corresponding to the first $l\ (l \le n)$ feature values to form a dimensionality reduction matrix $U_{reduce} = [u_1, u_2, ..., u_l] \in \Re^{m \times l}$, Where l is the dimension reduction dimension, the larger l is, the more eigenvectors are in U_{reduce} , the original feature data will not be lost a lot, the error will be smaller, and the size of l is determined by the contribution rate of eigenvalues, that is:

$$\eta = \frac{\sum_{i=1}^{l} S_{ii}}{\sum_{i=1}^{n} S_{ii}} \tag{18}$$

Finally, the difference between the sample and the mean is mapped to the low-dimensional space, and the dimensionality reduction of the data can be achieved, which is:

$$A_{reduce} = U^{T}_{reduce} D \tag{19}$$

6 Gesture Recognition Experiment And Result Analysis

6.1 Data acquisition and experimental environment

To verify the effectiveness of the static gesture segmentation method and the gesture recognition performance under different scale sizes, different rotation angles and different illumination conditions, it is necessary to collect data, establish a library of gesture recognition samples, and analyze the influence of various factors on gesture recognition. This article uses Kinect to collect ten types of gesture samples from different people, which means the numbers 0-9. For the convenience of subsequent experiments, the ten digits 0-9 are represented by A0-A9, as shown in Figure 7. Choose 100 samples for each type of gesture, 50 samples are used as test samples, and 50 samples remain as training samples. For the A0-A9 ten-type gesture, need to select 1000 samples.

A0 A1 A2 A3 A4

A5 A6 A7 A8 A9

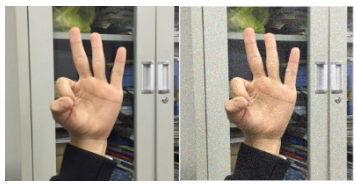
Fig.7 Ten types of gesture samples

Figure 8 is a partial sample collected in this paper. These samples were collected at different scales, different rotation angles and different lighting conditions. Figure 8 (d) is an image of added noise as a sample of the robustness of the subsequent verification algorithm. The highest frame of Kinect used in the experiment is 30FPS, the resolution is 640*480, the computer hardware configuration is CPU Intel Core i5, memory 4G, Win7 system, the software used is Visual Studio 2010 development environment, MatlabR2016a and LIBSVM software package.

(a) Different scales

(b) Different rotation angles

(c) Different lighting conditions



(d) Noise image

Fig.8 Partial sample image acquired under different conditions

6.2 Extraction experiment of Hu invariant moment and HOG feature

(1) Extraction of Hu moment invariants

The method of segmentation of static gestures is introduced in the foregoing, and gestures are divided for predefined ten types of gestures. Then post-processing, a binary contour map of ten types of gesture segmentation images can be obtained[43], as shown in Figure 9. Through calculation, the eigenvectors of 7 Hu invariant moments $M_1 - M_7$ of each gesture image can be obtained. The results of gesture A0-A9 feature extraction are shown in Table 1. Use the expression L(n) to represent the Hu moment, and H to represent the value of the Hu moment. The Hu moment can be obtained from H=L *10-n. Where n is a positive integer and the values in the table are the partial Hu moments in the 500 training samples.

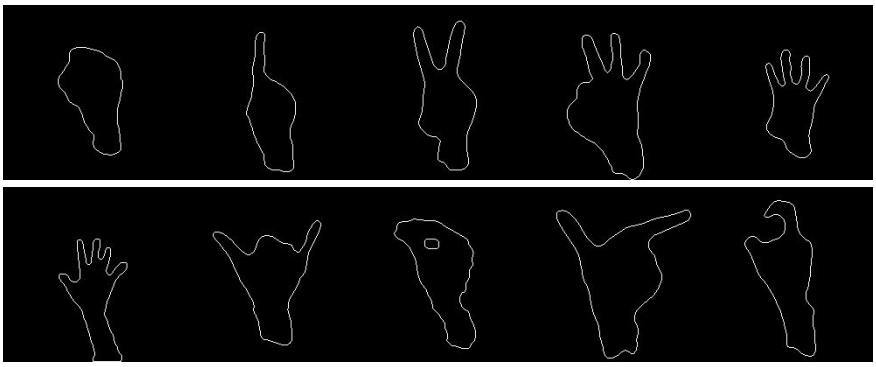


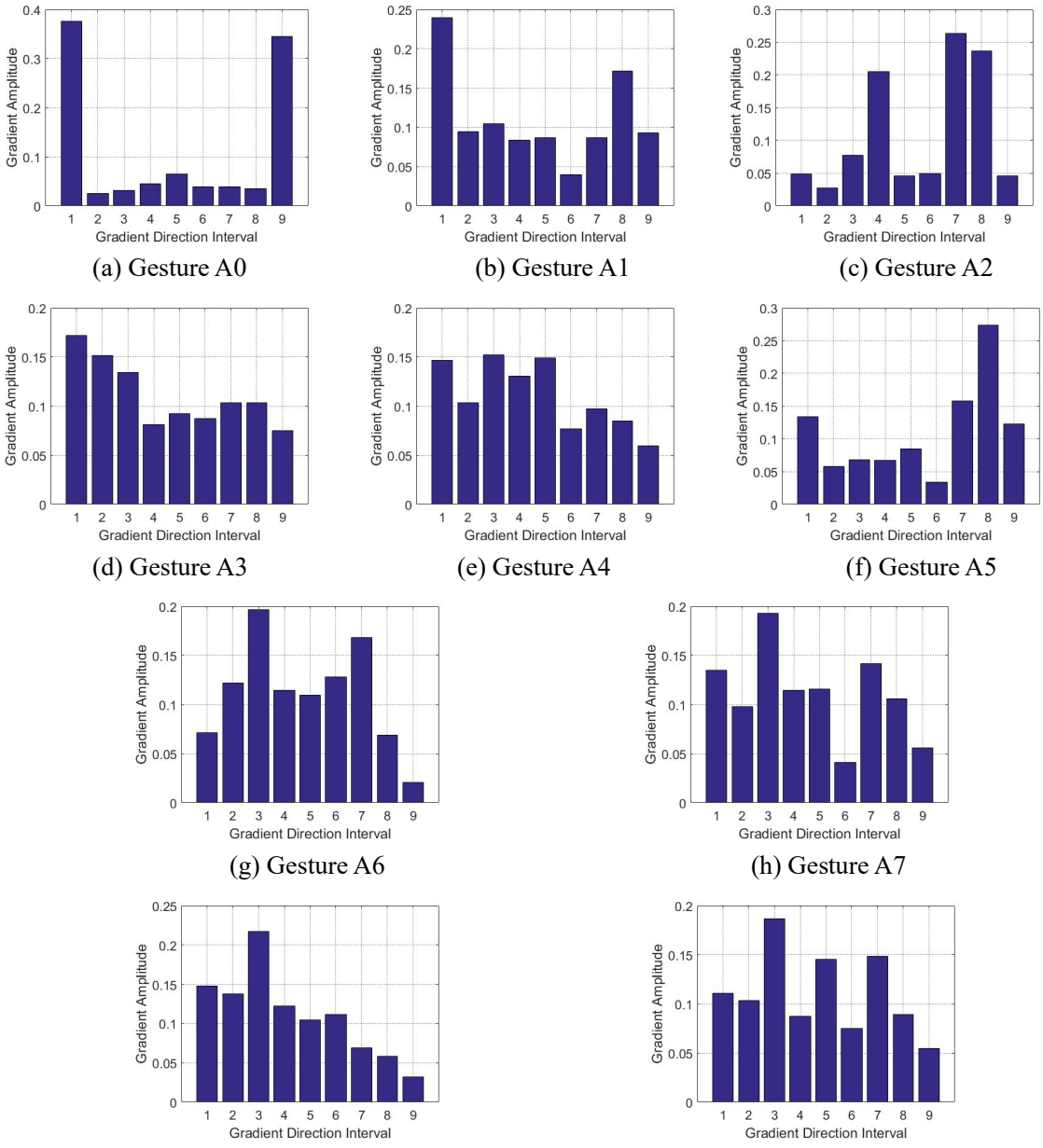
Fig.9 Binary diagram of ten types of gesture contours

Table 1 Pre-defined part of the ten-type gesture Hu invariant moment

Eigenvalues	M_1	M_2	M_3	M_4	M_5	M_6	M_7
A0	2.46(2)	1.67(4)	8.29(7)	1.96(9)	6.17(17)	2.29(11)	4.91(17)
A1	2.64(2)	4.34(4)	1.29(7)	3.39(8)	6.53(16)	5.82(10)	2.15(15)
A2	2.28(2)	2.55(4)	1.12(6)	8.17(7)	7.78(13)	1.29(8)	3.35(14)
A3	2.08(2)	1.18(4)	1.02(6)	1.16(6)	1.18(12)	1.27(8)	4.79(13)
A4	1.27(2)	3.82(5)	3.52(7)	3.45(7)	1.20(13)	2.13(9)	7.94(15)
A5	1.63(2)	1.26(4)	1.19(6)	7.54(7)	7.14(13)	8.40(9)	2.33(14)
A6	2.31(2)	2.21(5)	7.87(6)	1.19(7)	1.04(13)	5.53(10)	4.95(14)
A7	2.43(2)	2.66(4)	2.08(6)	4.40(7)	4.16(13)	6.15(9)	6.57(14)
A8	2.88(2)	5.14(5)	1.43(5)	1.94(7)	2.91(13)	1.13(9)	1.40(13)
A9	2.68(2)	3.95(4)	1.69(6)	7.09(7)	7.73(13)	1.39(8)	6.41(14)

(2) Extraction of HOG features

First, set the size of each cell of each gesture image to 16×16, and divide the gradient direction into 9 intervals, each block having a size of 32×32. Secondly, convolution operation is performed on the predefined ten types of gesture images to calculate the gradient and direction of the image. Then, the gradient amplitudes of the pixel points in the range of the gradient direction are accumulated, and a histogram of the gradient direction of each unit of each type of gesture image is obtained. Finally, using the two norms for normalization, the feature vector can be generated[44-45].



(a) Gesture A0 (b) Gesture A1 (c) Gesture A2

(d) Gesture A3 (e) Gesture A4 (f) Gesture A5

(g) Gesture A6 (h) Gesture A7

(i) Gesture A8 (j) Gesture A9

Fig.10 Predefined cell gradient direction histogram for ten types of gestures

6.3 Gesture recognition experiment results and analysis

(1) The recognition effect of the method in this paper

The extracted Hu invariant moments and HOG features are tested for their recognition effects with different weights ω_1 and ω_2 [46-48], as shown in Figure11. Where $\omega_1 + \omega_2 = 1$, when $\omega_1 = 1$, indicates that only the Hu moment recognition rate is used, and when $\omega_1 = 0$, it indicates that only the recognition rate of the HOG feature is used. When the weight of the Hu moment is $\omega_1 = 0.4$ and the weight of the HOG feature is $\omega_2 = 0.6$, the recognition rate is the highest. It can be seen from the influence of the value of A and B on the recognition performance. Different features have different effects on gesture recognition results, and the influence degree of HOG features is greater than Hu moment.

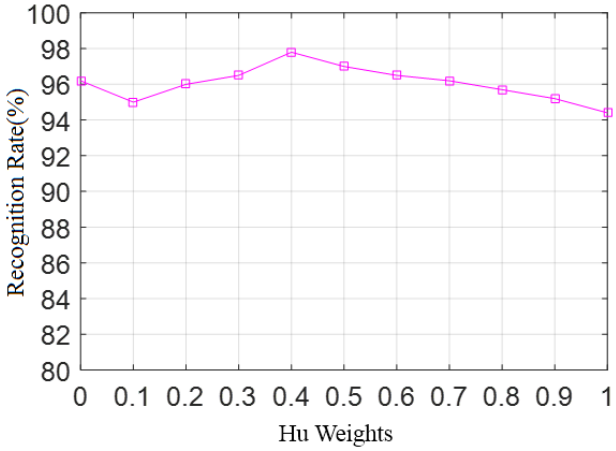


Fig.11 Effect of different weights on algorithm performance

In this paper, we choose the two best weights of Hu invariant moments and HOG features, that is $\omega_1 = 0.4, \omega_2 = 0.6$, the SVM classifier is used for classification and recognition, and the effect is shown in Figure 12. The figure visually shows the recognition rate of gestures on different collections, On the Hu invariant moment, the average recognition rate of all gestures is 94.4%. On the HOG feature, the average recognition rate of all gestures is 96.2%, on the fusion feature, the average recognition rate of all gestures is 97.8%. Obviously, the average recognition rate is higher when the two features are fused. Although the average recognition rate of Hu's invariant moments and HOG features is not much different, the overall average recognition rate of gestures is improved after combination of the two features.

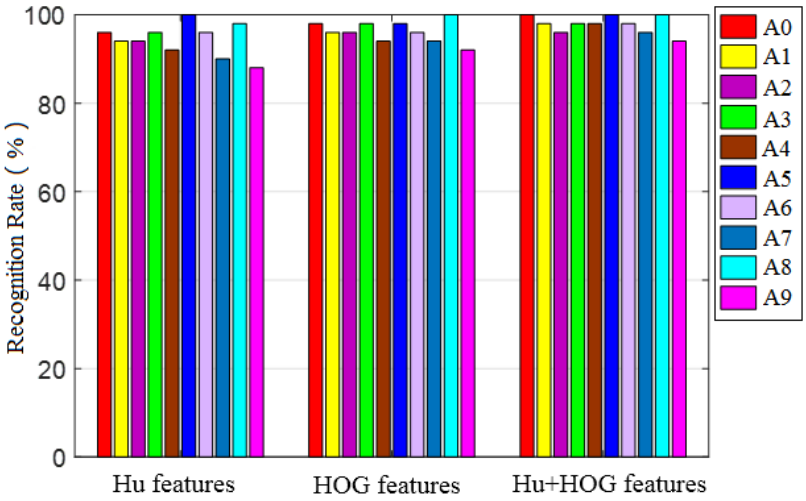


Fig.12 Recognition rate on different features

In the fusion feature, the fusion feature of gesture A0, gesture A5 and gesture A8 is relatively large, so the recognition rate reaches 100%. Gesture A9 contains a portion of a curved finger that is easily misjudged as gesture A1. Gesture A7 finger aggregation distance is basically zero contact, it is easy to misjudge it as gesture A0, the gesture outline of the gesture A6 is close to the gesture A0, so a misjudgment occurs. In the same situation, there are gesture A4 and gesture A5, gesture A3 and gesture A2. The misinterpretation of gesture A1 as gesture A0 may be due to an error in the gesture segmentation process.

Parameter Table 2 is the confusion matrix under different characteristics. a) in the Hu invariant moment, there are many gestures that are misjudged, which is the error generated when generating the binary

graph. The effect of increasing the binarization can improve the recognition rate. The recognition effect on the HOG feature in (b) is better than Hu invariant moment. In (c), the HOG feature is used to extract the local features of the gesture, and the Hu invariant moment is used as the global feature of the gesture, and the two complement each other, which improves the recognition rate. In (c), the HOG feature is used to extract the local features of the gesture, and the Hu invariant moment is used as the global feature of the gesture, and the two complement each other, which improves the recognition rate.

Table 2 Confusion matrix under different characteristics

(a) The confusion matrix of Hu invariant moments

Gesture type	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
A0	48	0	0	0	0	0	0	2	0	0
A1	1	47	0	0	0	0	0	0	0	2
A2	1	1	47	1	0	0	0	0	0	0
A3	0	0	1	48	1	0	0	0	0	0
A4	0	0	2	0	46	2	0	0	0	0
A5	0	0	0	0	0	50	0	0	0	0
A6	1	0	1	0	0	0	48	0	0	0
A7	5	0	0	0	0	0	0	45	0	0
A8	0	0	1	0	0	0	0	0	49	0
A9	2	4	0	0	0	0	0	0	0	44

(b) The confusion matrix of HOG features

Gesture type	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
A0	49	0	0	0	0	0	0	1	0	0
A1	2	48	0	0	0	0	0	0	0	0
A2	1	0	48	1	0	0	0	0	0	0
A3	0	0	1	49	0	0	0	0	0	0
A4	0	0	0	1	47	2	0	0	0	0
A5	0	0	0	0	1	49	0	0	0	0
A6	2	0	0	0	0	0	48	0	0	0
A7	1	1	0	0	0	0	0	47	0	1
A8	0	0	0	0	0	0	0	0	50	0
A9	3	1	0	0	0	0	0	0	0	46

(c) Confusion matrix on fusion features

Gesture type	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
A0	50	0	0	0	0	0	0	0	0	0
A1	1	49	0	0	0	0	0	0	0	0
A2	1	0	48	1	0	0	0	0	0	0
A3	0	0	1	49	0	0	0	0	0	0
A4	0	0	0	0	49	1	0	0	0	0
A5	0	0	0	0	0	50	0	0	0	0
A6	1	0	0	0	0	0	49	0	0	0
A7	2	0	0	0	0	0	0	48	0	0
A8	0	0	0	0	0	0	0	0	50	0
A9	0	3	0	0	0	0	0	0	0	47

(2) The robustness of this method under different conditions

The SVM algorithm combining HOG features and Hu invariant moments has certain robustness under noise conditions, rotation conditions and illumination conditions. This paper not only considers the samples of gestures in three environments, but also uses two features to describe gesture recognition from the global and local aspects, the HOG feature successfully solves the influence of illumination variation, scale size and small angle rotation on the recognition process, and combines the invariance of the rotation of the Hu moment itself, which improves the overall recognition rate of the gesture picture. Commonly used classification methods are BP neural network, K nearest neighbor[49-50], these methods were tested under the three conditions of this paper, and the recognition rates of various methods as shown in Figure 13 under different conditions were obtained. It can be seen that the method of this paper has a good recognition effect under various conditions.

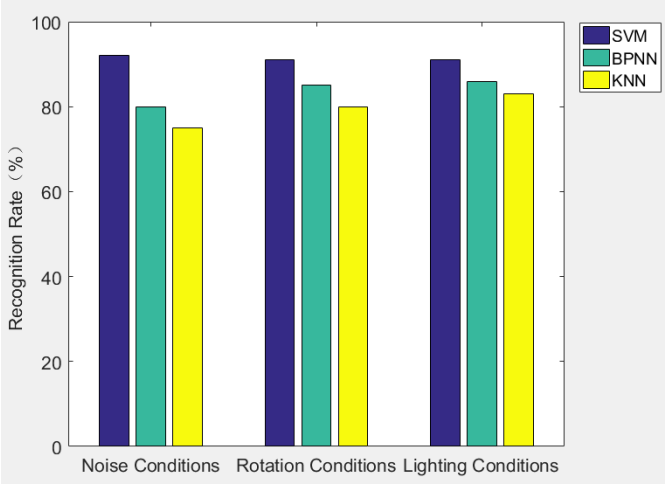


Fig.13 Comparison of recognition rates of methods under different conditions

(3) Comparison of methods in this paper with other similar methods

In [51], three expressions are added to the Hu moment as the feature of gesture recognition to match the gesture template. In [52], when the static gesture feature is selected, the concave and the perimeter area ratio of the gesture contour and the first four Hu moments are combined, and the radial kernel function is used for SVM classification.In [53], HOG features are used to identify multiple gestures using the SVM classifier. Compared with the results in the above literature, the results of the maximum recognition rate, the

minimum recognition rate and the average recognition rate are compared. As shown in Table 3, we can see that the average recognition rate of this method is higher than that of other similar methods.

Table 3 Comparison of the methods in this paper with other similar literature methods

Recognition methods	Maximum recognition rate (%)	Minimum recognition rate (%)	Average recognition rate (%)
Literature [51]	98	92	95.6
Literature [52]	99.9	85.6	95.4
Literature [53]	97.3	89.7	94.4
Method of this paper	100	94	97.8

7 Conclusion

Gesture recognition technology based on computer vision brings a new perspective of human-computer interaction. The selection of gesture segmentation methods, the application of feature extraction methods and the selection of classification algorithms will directly affect the accuracy of gesture recognition. In this paper, a static gesture segmentation method based on depth-color information and fusion features is proposed, through Kinect to collect image for color - depth segmentation, use the elliptical skin model to segment the gesture area, the feature vectors of HOG feature and seven Hu invariant moments are extracted, and two optimal weights of Hu invariant moments representing global feature and HOG feature representing local feature are selected. The recognition experiments were carried out under different conditions. Compared with the traditional classification methods, the SVM algorithm with fusion features has a relatively high recognition rate and it has better robustness. At the same time, compared with similar methods, the average recognition rate of this paper is still high, and the recognition effect is ideal.

Acknowledgments

This work was supported by grants of the National Natural Science Foundation of China (Grant Nos. 51575407, 51575338, 51575412, 51505349, 61273106) and Grants of the National Defense Pre-Research Foundation of Wuhan University of Science and Technology (GF201705).

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Chen D, Li G, Sun Y, Kong J, Jiang G, Tang H, Ju Z, Yu H, Liu H (2017) An interactive image segmentation method in hand gesture recognition. Sensors 17: 253. doi:10.3390/s17020253

[2] Sun Y, Li C, Li G, Jiang G, Jiang D, Liu H, Zheng Z, Shu W (2018) Gesture Recognition Based on Kinect and sEMG Signal Fusion. Mobile Networks and Applications 23(4): 797-805.

[3] Li G, Zhang L, Sun Y, Kong J (2018) Towards the sEMG hand: internet of things sensors and haptic feedback application. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-6293-x

[4] He Y, Li G, Sun Y, Zhao Y, Jiang G (2018) Numerical simulation-based optimization of contact stress distribution and lubrication conditions in the straight worm drive. Strength of Materials 50(11): 157-156.

[5] Jadooki S, Mohamad D, Saba T, Almazyad A, Rehman A (2017) Fused features mining for depth-based

hand gesture recognition to classify blind human communication. Neural Comput & Applic 28: 3285. https://doi.org/10.1007/s00521-016-2244-5

[6] Li G, Tang H, Sun Y, Kong J, Jiang G, Jiang D, Tao B, Xu S, Liu H (2017) Hand gesture recognition based on convolution neural network. Cluster Comput. https://doi.org/10.1007/s10586-017-1435-x

[7] Li B, Sun Y, Li G, Kong J, Jiang G, Jiang D, Tao B, Xu S, Liu H (2017) Gesture recognition based on modified adaptive orthogonal matching pursuit algorithm. Cluster Computing. https://doi.org/10.1007/s10586-017-1231-7

[8] Chang W, Li G, Kong J, Sun Y, Jiang G, Liu H (2018) Thermal Mechanical Stress Analysis of Ladle Lining with Integral Brick Joint. Archives of Metallurgy and Materials 63(2): 659–666.

[9] Fang Y, Liu H, Li G, Zhu X (2015) A multichannel surface EMG system for hand motion recognition. International Journal of Humanoid Robot 12(2). doi:10.1142/S0219843615500115

[10] Yin Q, Li G, Zhu J (2017) Research on the method of step feature extraction for EOD robot based on 2D laser radar. Discrete and Continuous Dynamical Systems-Series S 8(6): 1415–1421.

[11] Chen D, Li G, Sun Y, Jiang G, Kong J, Li J, Liu H (2017) Fusion hand gesture segmentation and extraction based on CMOS sensor and 3D sensor. International Journal of Wireless and Mobile Computing 12(3): 305-312.

[12] Sun Y, Hu J, Li G, Jiang G, Xiong H, Tao B, Zheng Z, Jiang D (2018) Gear reducer optimal design based on computer multimedia simulation. The Journal of Supercomputing. https://doi.org/10.1007/s11227-018-2255-3

[13] He L, Xiong C, Liu K, Huang J, He C, Chen W (2018) Mechatronic Design of a Synergetic Upper Limb Exoskeletal Robot and Wrench-based Assistive Control. Journal of Bionic Engineering 15: 247. https://doi.org/10.1007/s42235-018-0019-7

[14] Luo B, Sun Y, Li G, Chen D, Ju Z (2018) Decomposition Algorithm for Depth Image of Human Health Posture Based on Brain Health. Neural Computing and Applications. doi:10.1007/s00521-018-3883-5

[15] Jiang D, Li G, Sun Y, Kong J, Tao B (2018) Gesture recognition based on skeletonization algorithm and CNN with ASL database. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-018-6748-0

[16] Cheng W, Sun Y, Li G, Jiang G, Liu H (2018) Jointly network: a network based on CNN and RBM for gesture recognition. Neural Computing & Applications. https://doi.org/10.1007/s00521-018-3775-8

[17] Li J, Liu X, Ouyang G (2016) Using Relevance Feedback to Distinguish the Changes in EEG During Different Absence Seizure Phases. ClinicalEeg & Neuroscience 47(3): doi: 10.1177/1550059414548721

[18] Li G, Qu P, Kong J, Jiang G, Xie L, Gao P, Wu Z, He Y (2013) Coke oven intelligent integrated control system. Applied Mathematics & Information Sciences 7(3):1043–1050.

[19] Patvardhan C, Kumar P, Lakshmi C V (2018) Effective Color image watermarking scheme using YCbCr color space and QR code. Multimedia Tools & Applications 77(10): 12655. 77: 12655. https://doi.org/10.1007/s11042-017-4909-1

[20] Li G, Miao W, Jiang G, Fang Y, Ju Z, Liu H (2015) Intelligent control model and its simulation of flue temperature in coke oven. Discrete and Continuous Dynamical Systems - Series S 8(6): 1223–1237.

[21] Kundu A, Mazumder O, Lenka P, Bhaumik S (2017) Hand Gesture Recognition Based Omnidirectional Wheelchair Control Using IMU and EMG Sensors. Journal of Intelligent & Robotic Systems 91: 529. https://doi.org/10.1007/s10846-017-0725-0

[22] Li G, Gu Y, Kong J, Jiang G, Xie L,Wu Z,Li Z, He Y,Gao P(2013) Intelligent control of air compressor production process. Applied Mathematics & Information Sciences 7(3): 1051–1058.

[23] Fei M, Li J, Liu H (2015) Visual Tracking based on Improved Foreground Detection and Perceptual Hashing. Neurocomputing 152: 413-428.

[24] Li G, Li J, Ju Z, Sun Y, Kong J (2018) A Novel Feature Extraction Method for Machine Learning Based on Surface Electromyography from Healthy Brain. Neural Computing and Applications. doi:10.1007/s00521-018-3887-1

[25] Li G, Kong J, Jiang G, Xie L, Jiang Z, Zhao G (2012) Air-fuel ratio intelligent control in coke oven combustion process. International Journal on Information 12(11): 4487–4494.

[26] He J, Zhu X (2017) Combining Improved Gray-Level Co-Occurrence Matrix with High Density Grid for Myoelectric Control Robustness to Electrode Shift. IEEE Trans Neural Syst Rehabil Eng 99: 1539-1548.

[27] Zhu X, Liu J, Zhang D, Sheng X, Jiang N (2017) Cascaded Adaptation Framework for Fast Calibration of Myoelectric Control. IEEE Transactions on Neural Systems & Rehabilitation Engineering 25(3): 254-264.

[28] Hu C, Arvin F, Xiong C, Yue S (2017) Bio-Inspired Embedded Vision System for Autonomous Micro-Robots: The LGMD Case. IEEE Transactions on Cognitive & Developmental Systems 9(3): 241-254.

[29] Huang Y, Wang Y, Xiao L, Dong W (2014) Microfluidic serpentine antennas with designed mechanical tunability. Lab on a Chip 14(21): 4205-4212.

[30] Jiang D, Zheng Z, Li G, Sun Y, Kong J, Jiang G, Xiong H, Tao B, Xu S, Yu H, Liu H, Ju Z (2018) Gesture recognition based on binocular vision. Cluster Computing. https://doi.org/10.1007/s10586-018-1844-5

[31] He Y, Li G, Liao Y, Sun Y,Kong J, Jiang G, Jiang D, Tao B, Xu S, Liu H (2018) Gesture recognition based on an improved local sparse representation classification algorithm. Cluster Computing. https://doi.org/10.1007/s10586-017-1237-1

[32] Liao Y, Sun Y, Li G, Kong J, Jiang G, Jiang D, Cai H, Ju Z, Yu H, Liu H (2017) Simultaneous calibration: a joint optimization approach for multiple kinect and external cameras. Sensors 17(7): 1491. https://doi.org/10.3390/s17071491

[33] Miao W, Li G, Sun Y, Jiang G, Kong J, Liu H (2016) Gesture recognition based on sparse representation. International Journal of Wireless and Mobile Computing 11(4): 348–356.

[34] Li J, Wang J, Ju Z (2018) A Novel Hand Gesture Recognition based on High-level Features. International Journal of Humanoid Robotics 15(1). doi: 10.1142/S0219843617500220

[35] Ju Z, Ji X, Li J, Liu H (2017) An Integrative Framework of Human Hand Gesture Segmentation for Human-Robot Interaction. IEEE Systems Journal 11(3): 1326-1336.

[36] Li J, N.M. Allinson (2013) Building Recognition Using Local Oriented Features. IEEE Transactions on Industrial Informatics 9(3): 1697–1704.

[37] Oyedotun O K, Khashman A (2017) Deep learning in vision-based static hand gesture recognition. Neural Computing & Applications 28: 3941. https://doi.org/10.1007/s00521-016-2294-8

[38] Miao W, Li G, Jiang G, Fang Y, Ju Z, Liu H (2015) Optimal grasp planning of multi-fingered robotic hands: a review. Applied & Computer Mathematics 14(3): 238–247.

[39] Liu H, Wu J, Fan S, Jin M, Fan C (2018) Integrated virtual impedance control based pose correction for a simultaneous three-fingered end-effector. Industrial Robot 45(8). doi:10.1108/IR-09-2017-0173

[40] Li Z, Wang B, Liu H (2016) Target Capturing Control for Space Robots with Unknown Mass Properties: A Self-Tuning Method Based on Gyros and Cameras. Sensors 16(9):1383. doi: 10.3390/s16091383

[41] Cui M, Prasad S (2015) Class-dependent sparse representation classifier for robust hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing 53(5): 2683-2695.

[42] Liu H, Yang D, Jiang L, Fan S (2014) Development of a multi-DOF prosthetic hand with intrinsic actuation, intuitive control and sensory feedback. Industrial Robot An International Journal 41(4):381-392.

[43] Shull P B, Zhu X, Cutkosky M R (2017) Continuous Movement Tracking Performance for Predictable

and Unpredictable Tasks with Vibrotactile Feedback. IEEE Transactions on Haptics 10(4): 466-475.
[44] Huang Y, Dong W, Huang T, Wang Y, Xiao L (2015) Self-similar design for stretchable wireless LC strain sensors. Sensors & Actuators A Physical 224. doi: 10.1016/j.sna.2015.01.004
[45] Qi J, Jiang G, Li G, Sun Y, Tao B (2018) Surface EMG Hand gesture recognition system based on PCA and GRNN. Neural Computing and Applications. doi:10.1007/s00521-018-3885-3
[46] Bellocerezo R, Bianconi F, Fernández A, González E, Maria FD (2016) Experimental comparison of color spaces for material classification. Journal of Electronic Imaging 25(6). doi: 10.1117/1.JEI.25.6.061406
[47] Chai G, Zhang D, Zhu X (2017) Developing Non-Somatotopic Phantom Finger Sensation to Comparable Levels of Somatotopic Sensation through User Training With Electrotactile Stimulation. IEEE Transactions on Neural Systems & Rehabilitation Engineering 25(5): 469-480.
[48] Li C, Li G, Jiang G, Chen D, Liu H (2018) Surface EMG data aggregation processing for intelligent prosthetic action recognition. Neural Computing and Applications. doi:10.1007/s00521-018-3909-z
[49] Satapathy S, Sri Madhava Raja N, Rajinikanth V, Ashour A (2018) Multi-level image thresholding using Otsu and chaotic bat algorithm. Neural Comput & Applic 29: 1285. https://doi.org/10.1007/s00521-016-2645-5
[50] Dong W, Gu G, Zhu X, Dong X (2015) Solving the Boundary Value Problem of an Under-Actuated Quadrotor with Subspace Stabilization Approach. Journal of Intelligent & Robotic Systems 80: 299. https://doi.org/10.1007/s10846-014-0161-3
[51] Wu H, Huang Y, Xu F, Duan Y, Yin Z (2016) Energy Harvesters for Wearable and Stretchable Electronics: From Flexibility to Stretchability. Advanced Materials 28(45). doi: 10.1002/adma.201602251
[52] Radman A, Zainal N, Suandi S A (2017) Automated segmentation of iris images acquired in an unconstrained environment using HOG-SVM and GrowCut. Digital Signal Processing 64: 60-70.
[53] Jebril N A, Al-Zoubi H R, Al-Haija Q A (2018) Recognition of Handwritten Arabic Characters using Histograms of Oriented Gradient (HOG). Pattern Recognition & Image Analysis 28: 321. https://doi.org/10.1134/S1054661818020141