International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.12 (2016), pp.21-34
http://dx.doi.org/10.14257/ijmue.2016.11.12.03

Dynamic Hand Gesture Trajectory Recognition Based on Block Feature and Skin-Color Clustering

Zhang Qiu-yu, Lv Lu, Lu Jun-chi, Zhang Mo-yi and Duan Hong-xiang

School of Computer and Communication, Lanzhou University of Technology, Lanzhou, 730050, China
zhangqylz@163.com, lvlu901219@126.com, lujunchi9@163.com

Abstract

In recent years, dynamic hand gesture recognition has been a research hotspot of human-computer interaction. Since most existing algorithms contain problems with high computational complexity, poor real-time performance and low recognition rate, which cannot satisfy the need of many practical applications. Moreover, key frames obtained by inter-frame difference degree algorithm contain less information, which leads to less identified species and lower recognition rate. To solve these problems, we present a dynamic hand gesture trajectory recognition method based on the theory of block feature to extract key frames and the skin-color clustering's hand gesture segmentation. Firstly, this method extracts block feature of degree of difference between frames in hand gesture sequence to select key frames accurately. Secondly, the method based on skin-color clustering is applied to obtain the area of hand gesture after segmenting hand gestures from images. Finally, hidden Markov model (HMM), in which the angle data of hand gesture trajectories are input, is used for modeling and identifying dynamic hand gestures. Experimental results show that the method of key-frame extraction is used to obtain information of dynamic hand gestures accurately, which would improve the recognition rate of dynamic hand gesture recognition and, at the same time, can guarantee the real-time of hand gesture recognition system. The average recognition rate is up to 86.67%, and the average time efficiency is 0.39s.

Keywords*: Dynamic hand gesture recognition; Hand gesture trajectory; Key frame; Block feature; Skin-color clustering*

1. Introduction

With the rapid development of technology information, embedding computer into application environment duly has become a new research area of Human Computer Interaction. Traditional input devices, such as keyboard and mouse, are not simple and intuitionistic. However, hand gesture, a kind of simple, visualized and convenient Human Computer Interaction, can respond to complex interaction system rapidly [1-2]. Particularly, dynamic hand gesture recognition based on vision has become a hot research area. Hand gesture recognition is also widely applied in sign language recognition, virtual reality, video games. Compared to static hand gesture recognition, dynamic hand gesture recognition receives more extensive use in real world. Nevertheless, dynamic hand gesture consists of a series of different hand gestures of gesture sequences and it takes a lot of time to recognize each kind of hand gesture, which makes the velocity of hand gesture recognition unavailable [3-4].

Hand gesture recognition based on vision is a main research area. Ren *et al.* [5] proposed an approach based on spatio-temporal appearance modeling and dynamic space time warping algorithm to recognize hand gestures with complex background. Average recognition rate is 97%, however, the average time of segmentation, parameters extraction and recognition are 1.1s, 0.9s and 0.07s. Bao *et al.* [6] proposed an approach based on

ISSN: 1975-0080 IJMUE

SURF to track hand gesture trajectory. They built a dynamic hand gesture model after time warping and recognized dynamic hand gestures with data stream clustering method based on correlation analysis. Recognition rates of training set and test set are 87.1% and 84.6%. Pang and Ding [7] proposed to use divergence, vortices and hand movement direction vector as features and recognize hand gesture with traditional hidden Markov model (HMM). The recognition rate was improved, but time complexity was still relatively high. On the basis of von Miser and Fiser mixture theory, Beh *et al.* [8] proposed to use probability density function as the input of HMM model to build a spatio-temporal data model. This approach can eliminate the effect of arms on the experimental results and ignore the distance between the cameras. Wang *et al.* [9] proposed to use invariant curve distance and direction vector as features, use the threshold model to generate the optimal threshold and HMM to recognize dynamic hand gestures accurately. Wang *et al.* [10] proposed to quantify the directional feature of hand gestures and code after preprocessing the feature. They improved discrete HMM algorithm to recognize dynamic hand gestures. This approach has high recognition rate, but it takes long time to recognize, which makes the real-time capability unavailable.

In this work, aiming at the problems that existing dynamic hand gesture recognition algorithm has high computational complexity and low real-time performance, this paper studies key frame extraction algorithm and feature extraction of hand gesture trajectory through the establishment of a dynamic hand gesture recognition system. To solve the problem that the key frame image contained in the existing frame difference algorithm resulting in the lack of recognition types and low recognition rate, this paper presents a dynamic hand gesture recognition approach based on block feature extraction and color clustering. Firstly, the image difference is extracted by using the method of block, and the key frame sequence is selected. Secondly, the method of skin-color clustering is used to segment the image of the key frame sequence. Finally, using the HMM modeling and recognize of dynamic gestures. Experimental results show that the proposed method can effectively meet the real-time performance of the dynamic hand gesture system and guarantee the accuracy of the algorithm.

The rest of this paper is organized as follows. Section 2 describes the preliminaries. A detailed dynamic hand gesture trajectory recognition method is described in Section 3. Subsequently, Section 4 gives the experimental results as compared with other related methods. Finally, we conclude our paper in Section 5.

2. The Preliminaries

At present, the key frame extraction methods can be summarized as the following four kinds:

(1) Key frame extraction method based on shot boundary [11]. The first frame, the middle or the end frame of each shot are chosen as the key frames. This method is concise in design, small in computation, and suitable for switching scenes with simple contents or fixed scenes. But for complex and the variety of scenes, extracted shots are often unable to represent the information of the shots accurately.

(2) Key frame extraction method based on motion analysis [12]. On the basis of optical flow operation, the key frame is selected according to the structure of the lens. However, this method has a huge amount of computation, the real-time performance is poor, and the local minimum value is not necessarily accurate.

(3) Key frame extraction method based on visual content [13]. The key frames are extracted by the color, texture and other visual information of each frame. When the information is significantly changed, the current frame is the key frame. Main idea: Firstly, regard the first frame of the shot as the key frame. Then, the difference between the previous key frame and the rest frame is calculated. If the difference is greater than a certain threshold, another key frame is selected. This method can select the appropriate

number of key frames according to the change of content, but the selected frame is not necessarily representative of meaning, and it is easy to select too many key frames.

(4) Key frame extraction method based on clustering analysis [14]. Clustering method can effectively capture the visual content of the video camera, but it cannot save the time sequence and the dynamic information of the image frames. This method is the main technology of key frame extraction. Main idea: Firstly, determine an initial cluster center. Secondly, according to the distance between the current frame and clustering center to judge the current frame is classified as class or as a new clustering center. After clustering the frame, the closest frame is taken as the key frame.

3. The Proposed Schemes

3.1. Extraction Dynamic Hand Gesture Trajectory Recognition Procedure

As is shown in Figure 1, this paper studies dynamic hand gesture recognition based on two aspects: the key frame extraction based on block feature, hand gesture segmentation based on the skin-color clustering.

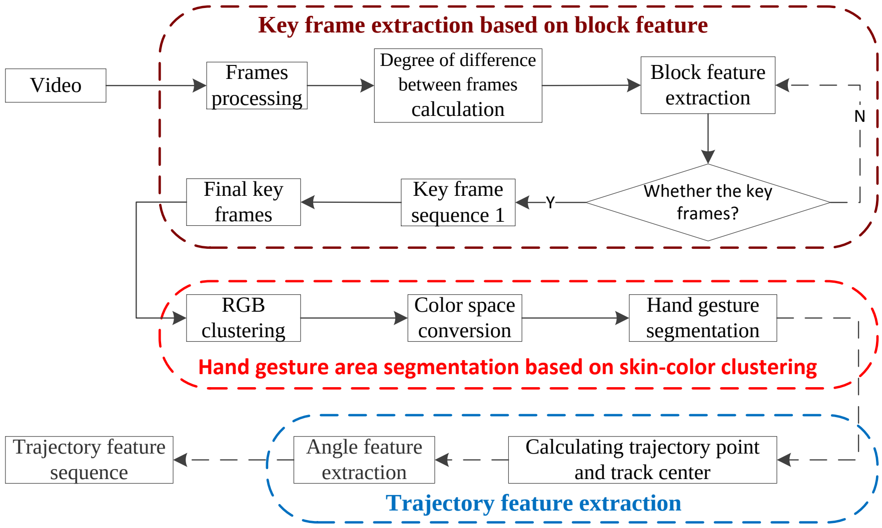


Figure 1. The Flow Chart of Dynamic Hand Gesture Trajectory Recognition

(1) Firstly, calculate the difference of the image of the dynamic hand gesture sequence, compare the block features of difference of extracted frames and obtain the accurate key frame sequence.

(2) A hand gesture segmentation method based on skin-color clustering is used to segment the image of key frames. By computing the trajectory point sequence, the point of the trajectory is extracted as the feature of hand gesture trajectory, and the input HMM algorithm is used for modeling and recognition.

3.2. Key Frame Extraction Based on Block Feature

This paper uses the key frame extraction method based on the block feature to select the key frame of the image frame sequence. The basic idea is to calculate the difference of all the images in the hand gesture image sequence, and extract the block features of the frame difference. Then the maximum M value is selected as the key frame, and the final key frame sequence is obtained through time warping. Steps are as follows:

Step 1: Convert color space. Convert RGB value of each pixel to gray value, calculation formula is as follows:

$$G_n(i,j) = R_1 \times r_n(i,j) + G_1 \times g_n(i,j) + B_1 \times b_n(i,j) \tag{1}$$

Where R_1=0.299, G_1=0.587, B_1=0.114, n=1, 2, …, N, $r_n(i, j)$, $g_n(i, j)$ and $b_n(i, j)$ are red, green and blue components of (i, j) pixel point of the n-th frame image.

Step 2: Calculate the difference and extract frame image difference of block features. Difference calculation formula is as follows:

$$Diff_n = \sum_i \sum_j C_n(i,j) \tag{2}$$

Where,

$$C_n(i,j) = \begin{cases} 1 & \sum_{15(i-1)+1}^{15i} \left(\sum_{20(j-1)+1}^{20j} A_n(i,j) \right) \\ 0 & others \end{cases} \tag{3}$$

Formula (3) can be understood as: a (160×120) difference matrix is divided into 8×8 blocks to extract features, as follows:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,120} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,120} \\ \vdots & \vdots & \ddots & \vdots \\ a_{160,1} & a_{160,2} & \cdots & a_{160,120} \end{bmatrix} \tag{4}$$

To convert

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,8} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,8} \\ \vdots & \vdots & \ddots & \vdots \\ A_{8,1} & A_{8,2} & \cdots & A_{8,8} \end{bmatrix} \tag{5}$$

Be $A_{1,1} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,15} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,15} \\ \vdots & \vdots & \ddots & \vdots \\ a_{20,1} & a_{20,2} & \cdots & a_{20,15} \end{bmatrix}$, $A_{1,2} = \begin{bmatrix} a_{1,16} & a_{1,17} & \cdots & a_{1,30} \\ a_{2,16} & a_{2,17} & \cdots & a_{2,30} \\ \vdots & \vdots & \ddots & \vdots \\ a_{20,16} & a_{20,17} & \cdots & a_{20,30} \end{bmatrix}$, …… and so on, in

$$A_{n+1}(i,j) = P_{n+1}(i,j) - P_n(i,j) \tag{6}$$

$$P_n(i,j) = G_{n+1}(i,j) - G_n(i,j) \tag{7}$$

Where n=2, 3, …, N-1, $G_n(i, j)$ is the value of (i, j) pixel point of n-th difference matrix.

Step 3: The maximum value of the difference is selected as the key frame sequence according to (8), all blue "○" marked points in Figure 2. Then, select all the extreme points and obtain the final key frames, all red "☆" marked frames in Figure 2.

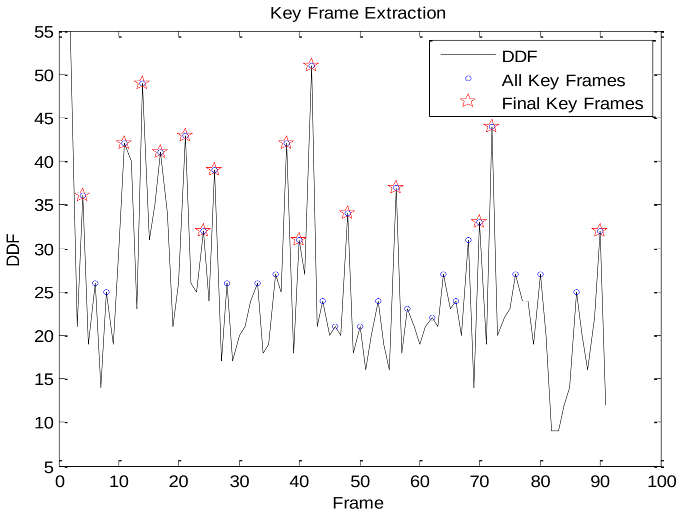


Figure 2. Selection of Key Frame Based on Block

$$e(i)=\begin{cases}1 & diff''(i)\cap diff'(i)=0\\ 0 & others\end{cases} \tag{8}$$

Step 4: By using the time warping algorithm, we can get the key frames of video sequences.

Series of hand gesture images are obtained by the extraction of the key frame. After analyzing Figure 2, it is obvious to find out that the selected key frame sequence is more uniform in Figure 2. In order to build HMM model of image sequences and recognize it, we need to segment the target region and extract the feature vectors from the target area.

3.3. Hand Gesture Segmentation Based on Skin-Color Clustering

Illumination conditions in the real environment, including light direction, brightness, color, will influence the image greatly. Therefore, the illumination compensation is necessary before image processing, which can reduce the impact of a certain light.

(1) The commonly used method of illumination compensation is Gray World, and concrete algorithm steps are as follows:

Step 1: Calculate the average *avR*, *avG*, *avB* of *R*, *G*, *B* three components of each pixel in the image, and obtain the average gray value *avGray* of the image, formula is as follows:

$$avGray=(avR+avG+avB)/3 \tag{9}$$

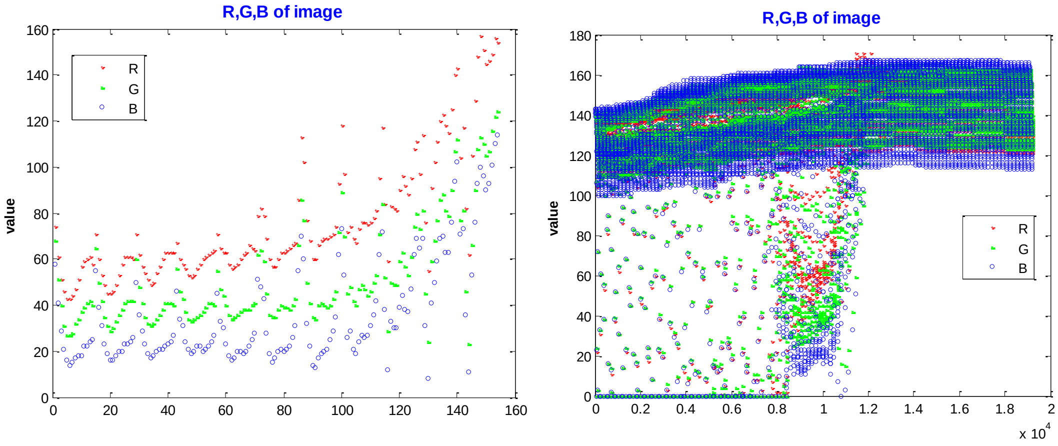
Step 2: Use the average gray value to adjust *R*, *G* and *B* of each pixel in the original image. Formulas are as follows:

$$\begin{aligned}C(R)&=C(R)\cdot avGray/avR\\ C(G)&=C(G)\cdot avGray/avG\\ C(B)&=C(B)\cdot avGray/avB\end{aligned} \tag{10}$$

Step 3: Range *C*(*R*), *C* (*G*), *C* (*B*) to 0 ~ 255, if the value is out of range, we set it to 255. The color offset in the image can be eliminated effectively in this way so that each pixel in the different color distribution tends to balance and plays a good role in light compensation.

(2) In RGB color space, *R*, *G* and *B* all have different degrees of brightness information and some correlations. It is necessary to normalize the luminance component in RGB space, but this method can only eliminate the relative luminance component in *R*, *G* and *B*. It can be known that the adaptive method is not good for skin-color detection in the

color space. After the experiment, the distribution of three components of the color and the color of the whole image is obtained, as shown in Figure 3.



(a) RGB Distribution of Skin-color Region (b) RGB Distribution of The Whole Image

Figure 3. RGB Distribution Diagram

Processing procedures are as follows:

Step 1: From Figure 3, we can see that the skin-color region of the *R*>*G*>*B*. We can cluster skin-color region according to the characteristics. Clustering formula is as follows:

$$r(c)=\begin{cases}0 & R>G\geq B\\ 1 & G\geq R>B\\ 2 & G>B\geq R\\ 3 & B\geq G>R\\ 4 & B>R\geq G\\ 5 & R\geq B>G\end{cases} \tag{11}$$

The clustering results of all images show that the approximate area of the hand gesture can be selected. The obtained area is as shown in Figure 4.

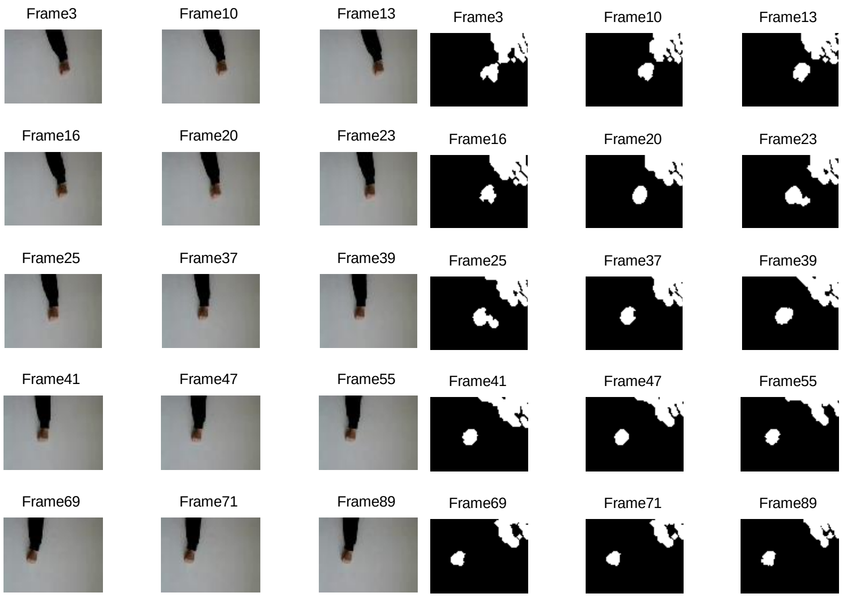


Figure 4. Hand Gesture Segmentation Based on RGB Clustering

From Figure 4, it can be seen that only RGB color clustering cannot successfully segment the hand gesture region, so it is necessary to deal with the hand gesture image.

Step 2: The RGB value of the image is converted to HSV value. In the HSV color space, the H values of the pixels in the skin-color image and the whole image are collected, and the spatial distribution of the skin-color points H is shown in Figure 5.

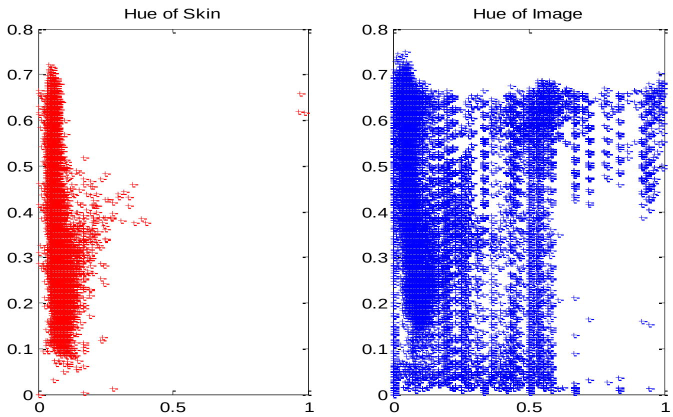


Figure 5. *H* component Distribution of Skin-color and The Whole Image

From Figure 5, it can be seen that the distribution range of H component of skin-color is [0, 0.13] and the distribution range of S component of skin-color is [0.08, 0.8]. Formula is as follows:

$$B(i,j)=\begin{cases}1 & H(i,j)\in[0,0.13]\cap S\in[0.08,0.8]\\ 0 & \quad\quad\quad\quad others\end{cases} \tag{12}$$

Based on the combination of formula above and clustering formula, we can segment the hand gesture region from the image. The results obtained are shown in Figure 6.

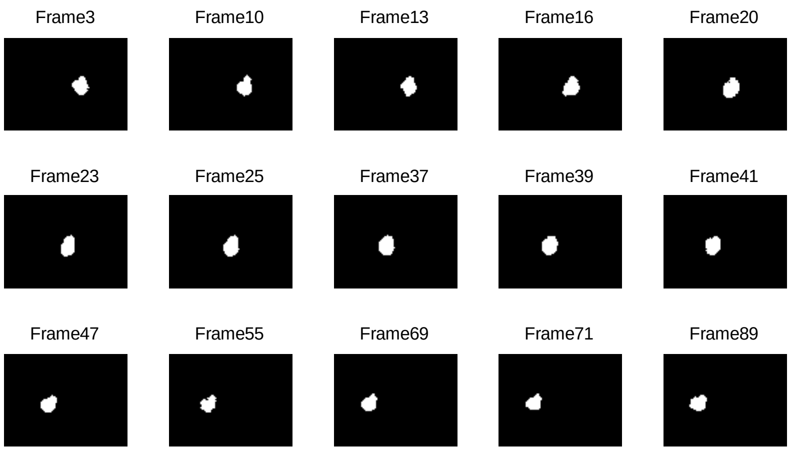


Figure 6. Final Hand Gesture Segmentation Results

Step 3: After the segmentation, the image is processed by erode and dilate operations. The edge of the hand gesture region is in a zigzag contour, so it is necessary to smooth hand gesture region with image erosion and expansion.

(3) When extract hand gesture region in YCbCr color space, if the mean value of two classes is very similar in segmentation samples, it shows that there is a certain overlap between the two classes. By using the method of pre-splitting, the average value of the class can be effectively separated, so that the error can be minimized. The classification procedures are as follows:

Step 1: Firstly, a pixel point set H is set up for the image:

$$H = \{v_m | m \in [1, N]\} \tag{13}$$

Step 2: Then randomly select a sample v_j from the point set H as the reference value of the mean value λ_1 of the K_1 class; The initial number of classes is 1, and then all of the pixel points are calculated, and finally get the most distant point g_m of the class.

$$g_m = \underset{g_j \in K_1}{\arg\max}\left(d\left(g_j, \lambda_1\right)\right) \tag{14}$$

In (14), the distance between pixel point g_j and class center λ_1 is represented. Add a class center λ_2=g_m. The number of samples increases to 2.

Step 3: According to (15).

$$i = \underset{m \in [1,2]}{\arg\min}\left(Dist\left(g_j, K_m\right)\right) \tag{15}$$

Re-cluster all pixel points. We select the closest distance to cluster according to the distance between pixels and centers:

$$H^m = \{K_i\left(i \in [1, m]\right)\} \tag{16}$$

After classifying, mean μ_i and variance σ_i^2 are estimated again. Repeat **Step 2** and keep traversal operation of set of all classes $K_i(i \in [1, m])$. Classification section is added to classify samples and classifying stops until m>=M (M is the number of classification). Finally, this algorithm clusters sample set H into M classes.

Traditional K-means clustering algorithm is improved in this paper. The split K-means clustering algorithm can effectively reduce the noise of the sample and reduce the error caused by the classification, which accelerates the rate of clustering, avoids the dead cycle and obtains better results. The method does not need to consider all kinds of skin-color threshold of pixels. Skin-color region and background region are separated by image clustering method. Compared with the traditional segmentation method based on skin-color, this method can effectively improve the skin-color segmentation effect.

3.4. Trajectory Feature Extraction

Steps of trajectory feature extraction are as follows.

Step 1: Calculate the center of hand gesture region of each image, which means centroid point (x_n, y_n). Formula is as follows:

$$\left(x_n, y_n\right) = \left(\frac{\sum_i \sum_j x \cdot b_n(i, j)}{\sum_i \sum_j b_n(i, j)}, \frac{\sum_i \sum_j y \cdot b_n(i, j)}{\sum_i \sum_j b_n(i, j)}\right) \tag{17}$$

Step 2: Calculate the trajectory center (x_0, y_0) of trajectory points {(x_1, y_1), (x_2, y_2), … , (x_M, y_M) }. Formula is as follows:

$$\left(x_0, y_0\right) = \left(\frac{1}{T}\sum_{t=1}^{T} x_t, \frac{1}{T}\sum_{t=1}^{T} y_t\right) \tag{18}$$

Step 3: Calculate the vector of each trajectory point $X_n(x_n, y_n)$ and trajectory center $Q(x_0, y_0)$. Formula is as follows:

$$\overrightarrow{QX_n} = (x_n - x_0, y_n - y_0) \tag{19}$$

Calculate the angle that x axis positive direction and connecting line of remaining trajectory points and trajectory centers formed. Formula is as follows:

$$s_i = \begin{cases} c & c = d \\ d & d = 180 + c \\ c+180 & ((c+d) = 180)\, and\, (c \neq 90) \\ c+360 & (c+d=0)\, and\, (c \neq d) \end{cases} \tag{20}$$

where,

$$c = \arctan\frac{y_n - y_0}{x_n - x_0}, \quad d = \arccos\frac{x_n - x_0}{\sqrt{(x_n - x_0)^2 + (y_n - y_0)^2}} \tag{21}$$

Step 4: Because the angle degree ranges 0°~360°, it needs structured angle. Angle structured graph is shown in Figure 7, formula is shown in (22).

$$Ang_n = \frac{s_n}{22.5} \tag{22}$$

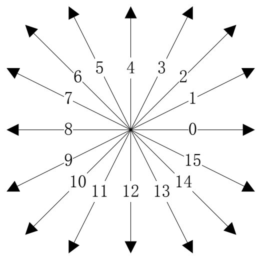


Figure 7. Angle Structured Graph

Step 5: The obtained angle sequence vector ***Ang***=(Ang_1, Ang_2,…, Ang_N) is used for the feature vector. Input HMM model and train it.

3.5. Hand Gesture Recognition Based on HMM

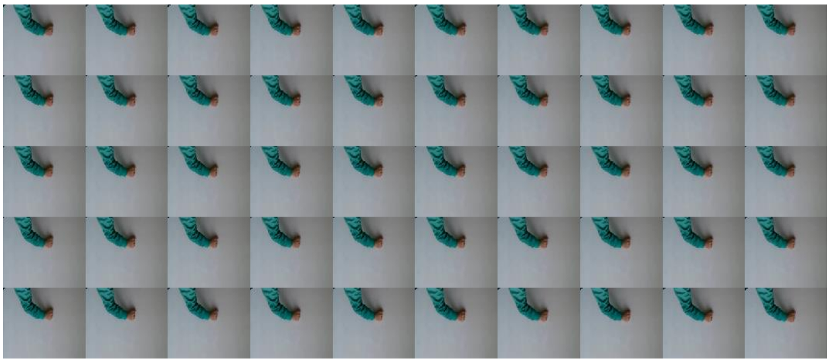
HMM is composed of first-order Markov process and observed random process. They are used to describe the relationship among state transition, state and the observation sequence. Where (1) first-order Markov process is described by state transition probability matrix ***A*** and Initial state distribution π, which can describe the state transition. (2) Observation random process is described by state output probability distribution matrix ***B***, which can describe the relationship between state and observation sequence.

HMM model has three main problems, including Assessment, decoding and learning. Forward backward algorithm, Viterbi algorithm [15] and Baum-Welch algorithm [16] can be used to solve these problems. This paper uses the mode from left to right as the basis of the model. In learning procedure, the more number of model states selected, the better results will be obtained. Specific procedure is introduced in [17].

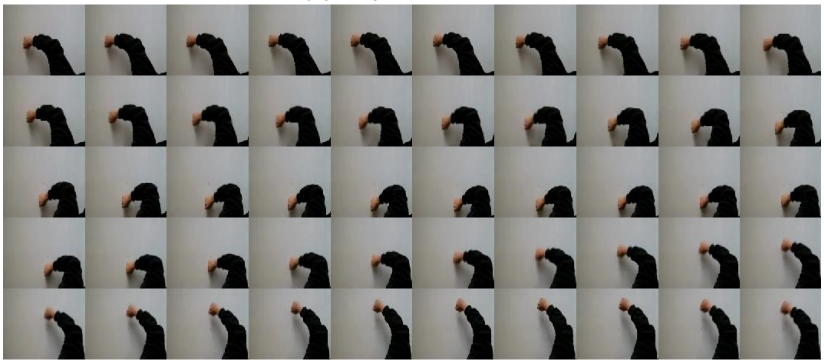
4. Experimental Results and Analysis

Experiment platform: Hardware environment: Inter (R) Core(TM) i3-2120, 4G, 3.30GHz. Software environment: Win7, MATLAB R2013a. Experiment data set of dynamic hand gesture video library is composed of number "0-9" of dynamic hand gestures. Three experimenters show 10 different hand gestures and repeat 10 times. We can get 300 hand gesture samples (3×10×10). The duration of each hand gesture is different, and it is about 3~12 *s* (Frame rate δ=25 *fps*). To simulate the application of close distance hand gesture recognition, we make sure that the distance between hand gesture and camera is 30-80 *cm*. The size of each 24 true color image is 160×120.

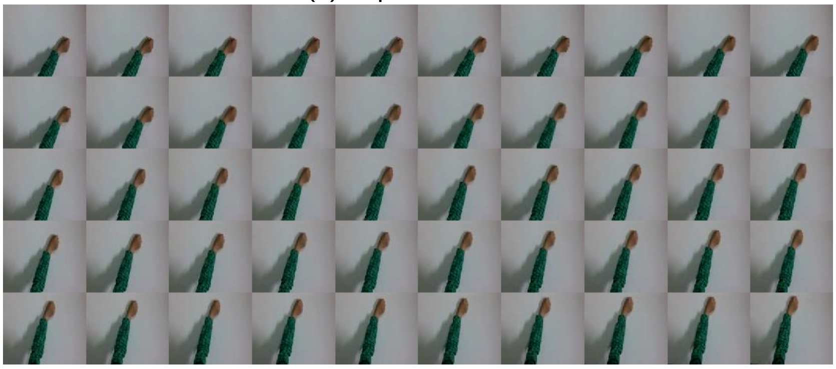
Figure 8 shows the 1-50 frame images of dynamic hand gesture “7” in the hand gesture video library. Figure 8(a), Figure 8(b) and Figure 8(c) show the 1-50 frame images of experimenter A, B and C.



(a) Experimenter A



(b) Experimenter B



(c) Experimenter C

Figure 8. 1-50 Frame Images of Dynamic Hand Gesture “7”

4.1. Comparison of HSV Clustering and YCbCr Clustering

In experiment, hand gesture set of each of the participants is trained and recognized by using HSV clustering and YCbCr clustering. First 6 hand gestures of each participant’s hand gestures are used for training set, which means that there are 60 (6×10) samples in the training set. Test set is composed of all samples of each participant’s hand gesture set, which means that there are 100 samples in each test set. Dynamic hand gesture training library and test library are tested with the method of HSV clustering and YCbCr clustering. Experiment results are shown in Table 1.

Table 1. Recognition Results Based on HSV Clustering and YCbCr Clustering from Different Participants

Methods	Average sequence length	Operation speed (*s*)		Recognition rate (%)	
		HSV	YCbCr	HSV	YCbCr
Experimenter A	138	0.399596	0.458135	88	81
Experimenter B	100	0.351684	0.399676	100	98
Experimenter C	149	0.422069	0.465978	88	90

Through the analysis of the data in Table 1, it can be seen:

(1) The average sequence length of the experimenter B was the shortest and the time he needs is the shortest. The average sequence length of the experimenter C was the longest and the time he needs is the longest. It can be concluded that the recognition time is longer when the sample is recognized with the increase of the average sequence length of the sample and the same segmentation method. They are proportional to each other.

(2) The recognition rate of the experimenter B and the experimenter C with two methods are basically the same. For each participant, the recognition rate of the two methods based on HSV clustering and YCbCr clustering are close in general, which means the results of the two methods are similar, so the HSV clustering method is selected for the following calculation.

(3) Among the two methods, the recognition rate of the experimenter A is very different. It can be predicted that the YCbCr clustering method is not robust to the environment of the experimenter A. In order to ensure the stability of the experiment, it is better to select the HSV clustering algorithm.

4.2. Mixed Sample Recognition Rate

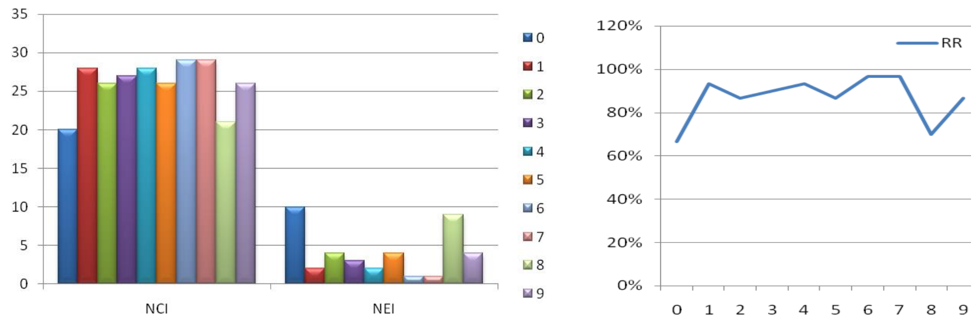
In the experiment, all the samples of the experimenter A, B and C are selected for the experiment, which has 30 samples of each dynamic hand gesture and a total of 10 hand gestures. We select the first 6 frames of each of the participants for training. Training sample has 180 ($3\times6\times10$) images. All 300 samples are used to recognize. In experiment results, the number of wrong recognition samples is 40 and the recognition rate is 86.67%. The average operating speed of each sample is 0.393446s. Specific recognition results are shown in Table 2.

It can be seen from Table 2 that the recognition rate of the mixed sample is lower than the experiment mentioned in Section 4.1. For different participants, the operating speed and the recognition rate are different.

Table 2. Recognition Results of Mixed Sample

-	Operation speed (s)	Recognition rate (%)
Experimenter A	0.399596	88
Experimenter B	0.351684	100
Experimenter C	0.422069	88
All samples	0.393446	86.67

Train and recognize dynamic hand gesture set “0-9” of all samples. Experiments results are shown in Figure 9. Figure 9(a) shows the number of correct recognition of errors for each hand gesture. Blue presents the number of correction and Red presents the number of errors. Figure 9(b) is recognition line chart of each hand gesture “0-9”.



(a) The Number of Correction and Errors (b) Recognition Rate of Hand Gesture "0-9"

Figure 9. Recognition Results of Hand Gesture "0-9" Based on HSV Clustering

It can be seen in the Figure 9 that the error recognition rate of "0" and "8" is higher than other hand gesture. The main reasons for this difference are:

(1) As is shown in Figure 10, different people have different habits of showing "0" and "8". They are more random, so the robustness of the method is very bad. It can be concluded that there are some differences in the habits and speed of individual hand gestures, which leads to a dramatic change in the rate of recognition.

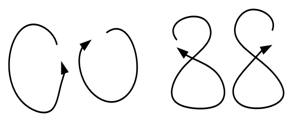


Figure 10. Different Writing Ways of "0" and "8"

(2) The angle feature that HMM training exacts is relatively weak. It will not accurately generalize the motion trajectory and direction of hand gestures, which will lead a high recognition rate for the hand gesture of a certain individual.

It is not very accurate to generalize the motion trajectory and direction of hand gestures. It cannot be applied to all situations and it is not practical.

In addition to the main reasons mentioned above, the effects of illumination and complex backgrounds make the recognition rate unsatisfied.

4.3. Comparison with Existing Methods

In order to verify the accuracy and real-time performance of the key frame selection method, operating speed and recognition rate of algorithm is statistical in this paper. It is compared with the method in [17]. Specific results are shown in Table 3. According to Table 3, the proposed method has the same experiment platform, experiment environment and hand gesture library. The recognition rate of the proposed method reduces by 1%, but the operating speed of the proposed method increases by 14%, which illustrates that this paper improves the method in [17] successfully and ensures the recognition rate. The operating speed of system is improved and the real-time recognition of dynamic hand gesture is a step forward.

Table 3. Comparison Results of Existing Recognition Method

	The proposed method	Ref. [17]	Ref. [6]	Ref. [5]
Platform frequency	3.3 GHz	3.3 GHz	2.2 GHz	600 MHz
Frame rate (frame/*s*)	25	25	8~16	10
Recognition rate(%)	86.67	87.67	84.6	91.7
Operating speed (*s*)	0.393446	0.457524	1~3	≥2.07

5. Conclusions

This paper presents a dynamic hand gestures trajectory recognition based on block feature and skin-color clustering. On the basis of key frame of dynamic hand gesture trajectory recognition procedure, the proposed method has solved the problem that few species of recognition and low recognition rate caused by lacking information of frame difference algorithm. The central idea of the proposed method is combining key frame extraction method of block feature and skin-color clustering segmentation method. Firstly, this paper calculates frame difference and extracts block feature. Compared with block feature, key frame sequence is extracted accurately and little difference of image is ignored. Secondly, skin-color clustering method is used to segment hand gesture region of key frame sequence images. Finally, HMM is used to build a dynamic hand gesture model and recognize. Experiment results illustrate that the proposed method is effective, reliable and robust to complex illumination. It also meets the demand of real-time of dynamic hand gesture trajectory recognition.

How to simply and accurately extract hand gesture from complex backgrounds will be the focus of our research in the future. HMM algorithm will be improved to build hand gesture model accurately.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (No. 61363078), the Youth Science and Technology Foundation of Gansu Province (No. 148RJYA001). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

[1] K. Z. Mei, L. Xu, B. L. Li, B. Lin and F. Wang, "A real-time hand detection system based on multi-feature", Neurocomputing, vol. 158, **(2015)**, pp. 184-193.
[2] S. S. Rautaray and A. Agrawal, "Real time multiple hand gesture recognition system for human computer interaction", International Journal of Intelligent Systems and Applications, vol. 4, no. 5, **(2012)**, pp. 56-64.
[3] H. Li and M. Greenspan, "Model-based segmentation and recognition of dynamic gestures in continuous video streams", Pattern Recognition, vol. 44, no. 8, **(2011)**, pp. 1614-1628.
[4] Y. Song, D. Demirdjian and R. Davis, "Continuous body and hand gesture recognition for natural human-computer interaction", ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 2, no. 1, **(2012)**, pp. 5:1-5:28.
[5] H. B. Ren, Y. X. Zhu, G. Y. Xu, X. Y. Lin and X. P. Zhang, "Spatio-temporal appearance modeling and recognition of continuous dynamic hand gestures", Chinese Journal of Computers, vol. 23, no. 8, **(2000)**, pp. 824-828.
[6] J. T. Bao, A. G. Song, Y. Guo and H. R. Tang, "Dynamic hand gesture recognition based on SURF tracking", Robot, vol. 33, no. 4, **(2011)**, pp. 482-489.
[7] H. Pang and Y. Ding, "Dynamic hand gesture recognition using kinematic features based on Hidden Markov Model", Proceedings of the 2nd International Conference on Green Communications and Networks(GCN2012), Springer Berlin Heidelberg, Chongqing, China, vol. 5, **(2013)**, pp. 255-262.
[8] J. Beh, D. K. Han and Durasiwami, "Hidden Markov Model on a unit hypersphere space for gesture trajectory recognition", Pattern Recognition Letters, vol. 36, **(2014)**, pp. 144-153.
[9] X. Wang, M. Xia and H. Cai, "Hidden-Markov-Models-based dynamic hand gesture recognition",

Mathematical Problems in Engineering, vol. 2012, **(2012)**, pp. 1-11.
[10] Q. Wang, Y. R. Xu and X. Bai, "Dynamic gesture recognition using 3D trajectory", Proceedings of the 4th IEEE International Conference on Information Science and Technology, Shenzhen, China, **(2014)**, pp. 598-601.
[11] Y. Rui, T. S. Huang and S. Mehrotra, "Exploring video structure beyond the shots", Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Austin, TX, **(1998)**, pp. 237-240.
[12] W. Wolf, "Key frame selection by motion analysis", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, GA, vol. 2, **(1996)**, pp. 1228-1231.
[13] H. J. Zhang, J. Wu and D. Zhong, "An integrated system for content-based video retrieval and browsing", Pattern Recognition, vol. 30, no. 4, **(1997)**, pp. 643-658.
[14] A. Hanjalic and H. J. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis", IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 8, **(1999)**, pp. 1280-1289.
[15] M. Elmezain and A. Al-Hamadi, "Gesture recognition for alphabets from hand motion trajectory using Hidden Markov Models", Proceedings of the IEEE International Symposium on Signal Processing and Information Technology, Giza, **(2007)**, pp. 1192-1197.
[16] Y. Zhang, P. Liu and F. K. Soong, "Minimum error discriminative training for radical-based online Chinese handwriting recognition", Proceedings of the 9th IEEE International Conference on Document Analysis and Recognition, Parana, vol. 1, **(2007)**, pp.53-57.
[17] Q. Y. Zhang, L. Lv, M. Y. Zhang, H. X. Duan and J. C. Lu, "A dynamic gesture trajectory recognition based on key frame extraction and HMM", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, no. 6, **(2015)**, pp. 91-106.

