International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2173132

Real-Time Hand Gesture Recognition

Pranjali Manmode*1, Rupali Saha2, Manisha N. Amnerkar3
[1]M-Tech Student, Computer Science & Engineering, RTMNU, Nagpur, Maharashtra, India
[23]Assistant Professor, Computer Science & Engineering, RTMNU, Nagpur, Maharashtra, India

Article Info
Volume 7, Issue 3
Page Number: 618-624

Publication Issue :
May-June-2021

Article History
Accepted : 10 June 2021
Published : 16 June 2021

ABSTRACT

With the rapid development of computer vision, the demand for interaction between humans and machines is becoming more and more extensive. Since hand gestures can express enriched information, hand gesture recognition is widely used in robot control, intelligent furniture, and other aspects. The paper realizes the segmentation of hand gestures by establishing the skin color model and AdaBoost classifier based on haar according to the particularity of skin color for hand gestures and the denaturation of hand gestures with one frame of video being cut for analysis. In this regard, the human hand is segmented from a complicated background. The camshaft algorithm also realizes real-time hand gesture tracking. Then, the area of hand gestures detected in real-time is recognized by a convolutional neural network to discover the recognition of 10 common digits. Experiments show 98.3% accuracy.

Keywords : Hand Gesture Segmentation, Hand Gesture Tracking, Hand Gesture Recognition, Neural Network

I. INTRODUCTION

With the development of interaction between human and machine, the interaction between computer and human is becoming more and more frequent. Among them, hand gestures are commonly used in this aspect. Since there are various hand gestures and enriched information contained in them, recognition of hand gestures has been greatly used in many fields, such as UAV, somatosensory game, sign language recognition, and so on [1][2]. In this regard, it is of great significance to study hand gesture recognition.

The interaction system in the paper is further composed of three parts as hand gesture segmentation, hand gesture tracking, and hand gesture recognition. In terms of hand gesture segmentation, it is realized by cutting the relevant specific hand gesture from one frame of video, which is also the first step for hand gesture recognition. It mainly includes the types based on skin color, edge detection, motion information, and statistical template, which have different advantages and disadvantages. The paper adopts a fusion algorithm to realize the hand gesture segmentation in a complicated environment.

Regarding hand gesture tracking, it is about real-time location and hand gesture tracking in the video according to some features, so it is the critical step for hand gesture recognition. Hand gesture tracking ensures that the targeted hand gestures are not lost and kept in real-time monitoring.

Currently, the algorithm for hand gesture tracking, which is applied widely, includes meanshift[3], Kalman filtering and optical flow algorithm, etc. According to the requirements for real-time tracing and accuracy in the process, the paper adopts the cam shift algorithm, the improved mean-shift version.

Hand gestures ranging from 1-10 in the complicated background are adopted in the experiment, with 1600 pictures collected for the training set, so there are 16000 pictures of hand gestures. Then, 4000 images of hand gestures, 400 photos for each type are detected for the test set. The paper adopts the LeNet-5 network to recognize hand gestures within a specific area and realize the classification of 10 types of digits of hand gestures ranging from 1-10. The main structure of an interactive system is shown in Fig.1.

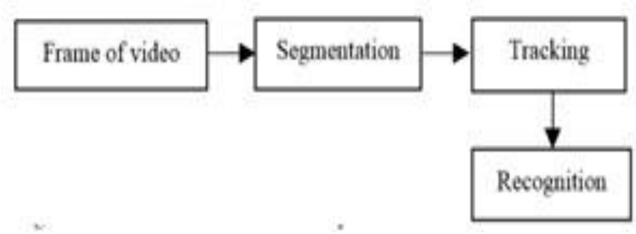

Fig.1 Main structure of an interactive system

II. HAND GESTURE SEGMENTATION

Presently, there are many ways to realize hand gesture segmentation. Based on the segmentation of skin color model, the skin color model is established to realize the hand gesture segmentation according to the difference between skin color of hand gestures and external environment and the model is not affected by the hand postures, but it is not able to exclude the objects which are similar to the skin color, such as human face and so on; the hand gesture segmentation based on edge detection[4] can segment the hand gestures according to the discontinuity of gray value in the margin area of the image region, but it is easy to be interrupted by the noise, and it has strict requirements for the background; the hand gesture segmentation based on movement information, including frame difference method and background difference method and so on[5] adopts the information of movement of hand gestures to segment hand gestures on the premise of static of background. The effect is good in a static environment while not well-performed in the dynamic experience; the segmentation method of hand gesture based on statistical template matching can rapidly identify the hand area and non-hand area by using training classifier of gesture template feature, but it can only recognize one or more hand gestures; it can not satisfy our demands. The hand gesture segmentation in the pre- paper processes the images and establishes a Gaussian mixture model according to the skin colors; moreover, it also segments hand gestures by combining with AdaBoost classifier based on Haar features

A. Gaussian mixture model of skin color

Gaussian mixture model of skin color is the parameterized distribution model of skin color [6].

Initial value of the parameter.

E step: according to the current model parameter, calculate the posterior weight expectation.

$$w^{(t)} = \arg\max l\left(X, \theta^{(t)} w^{(t-1)}\right) \tag{1}$$

3) M step: calculate the model parameter in the new round of iteration.

$$\theta^{(t+1)} = \arg\max l\left(X, \theta^{(t+1)} \theta^{(t)}\right) \tag{2}$$

4) Loop the second and third steps until detection convergence.

Suppose Gaussian mixture function of skin color:

$$f(x) = \sum \alpha_i N(x \mid t_i, \sum i) \tag{3}$$

N is Gaussian mixture function of skin color with multiple dimensions, t_i is the mean vector, $\sum i$ is the covariance matrix. In the training process, the posterior weight expectation can be calculated and obtained after E step.

$$w_j^i = P(z^{(i)} = j \mid x^{(i)}; \alpha, t, \Sigma) \qquad (4)$$

After M step, the maximum likelihood value of training sample is calculated and obtained.

$$l(a,t,\Sigma) = \sum_{i=1}^{n} \sum_{z^{(i)}=1}^{K} \log p\left(x^{(i)}, z^{(i)}; \alpha, t, \Sigma\right) \qquad (5)$$

$$\mathrm{C}(\mathrm{i},\mathrm{j}) = log \frac{\frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left(-0.5\left(x^i - t_j\right)^T \sum_j^{t-1}\left(x^i - t_j\right)\right)\alpha_j^{(i)}}{w_j^{(i)}} \qquad (6)$$

Fix the parameters $\alpha_j, \sum i$, then take the derivative of t_i with likelihood.

$$\nabla_{t_q} = \sum_{i=1}^{n} w_q^{(i)} \left(\sum_q^{-l} x^{(i)} - \sum_q^{-l} t_q \right) \qquad (7)$$

Obtain the update of mean parameters.

$$t_q = \left(\sum_{i=1}^{n} w_q^{(i)} x^{(i)} \right) / \left(\sum_{i=1}^{n} w_q^{(i)} \right) \qquad (8)$$

The Gaussian mixture model of skin color is used to segment the hand gesture, the effect is shown in the Figure 2.

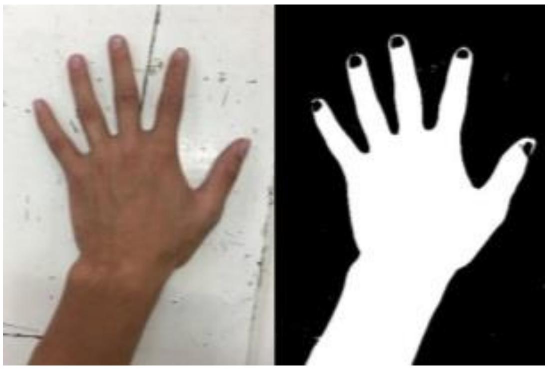

Fig 2. Skin color segmentation.

B. Hand gesture segmentation based on model features

The hand gesture segmentation based on the skin color model is susceptible to being interrupted by objects with a similar color of hands, such as human face and so on. In order to overcome the above shortcomings, the hand gesture segmentation based on model features is adopted after the detection of skin color. Then, the hand gestures features are extracted by plenty of hand gestures, and the classifiers are trained using these features, which is conducive to distinguishing the hand area and non-hand area. The paper adopts the AdaBoost classifier based on the Haar feature.

C. Haar feature

Haar feature reflects the image grayscale value change. The black and white rectangle areas are used to compose the feature model. The pixel sums under white spots are subtracted from the pixel sums under the black sites in the feature model. And express the feature value of objects. As shown in Fig.3, A and B represent the margin feature; C presents a linear feature while D represents a diagonal feature. Lienhart R.et al.[7] expand the above essential elements and form the expanded rectangle feature.

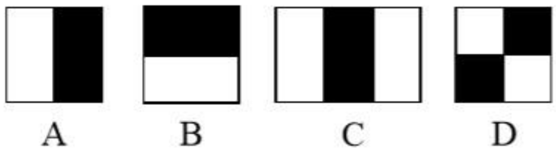

Fig.3 Basic haar feature

If all the rectangle feature areas are traversed when calculating Haar feature value, it will cause much-repeated calculation and waste a lot of time. Two integral images can rapidly calculate the rectangle feature, and its primary idea is to convert the original image into an integral vision. When calculating the sum of a pixel within one specific area, just the values of four angular points of the matrix area in the critical image are indexed, then the simple addition and minus the calculation is required for obtaining the Haar feature value. In the whole process, the images are just detected for one time, which greatly enhances the efficiency of calculation. Formula definition of elements in integral image:

$$ii(x,y) = \sum_{k \le x, l \le y} I(k,l) \tag{9}$$

the elements in the integral image are the sum of all the pixel values in the left upper corner of the original image. s(i,j) represents the accumulated sum in the row direction, initialized s(i,-1) is 0; ii(i,j) represents the results of integral images, initialized ii(-1,j) is 0; after scanning the image, the following formula can be used to realize iterative operation:

$$s(i,j) = s(i,j-1) + I(i,j) \tag{10}$$

$$ii(i,j) = ii(i-1,j) + s(i,j) \tag{11}$$

Scan the row of the image by row, and then the integral image ii is constructed, which can be used to calculate the Haar feature value rapidly.

b) AdaBoost classifier

AdaBoost is an iterated learning method, a robust learning algorithm upgraded from a group of weak learning algorithms.

The steps are as the following:

1. Create the training dataset, regard the pictures containing hand gestures as the positive samples, the background environment without hand gestures as the negative samples. To ensure the accuracy, 1000 pictures of positive models and 2000 pictures of negative samples are selected.
2. Calculate the feature values of these pictures, as well as the integral image.
3. Build multiple weak classifiers according to the feature values generated.
4. Adopt AdaBoost to train the weak classifier into the robust classifier.
5. Test the built robust classifier and make some adjustments accordingly.

C. Hand gesture segmentation

A gaussian mixture model of skin color and AdaBoost classifiers based on the Haar feature are introduced, respectively. The Hand gesture segmentation based on skin color has the advantages of strong adaptability, not being affected by posture and rotation. However, the Gaussian mixture model of skin color is quickly interrupted by objects with similar skin color, such as the human face, arm, etc. The accuracy and robustness of the AdaBoost classifier based on the Haar feature are good, and they have good performance even in a complicated environment. Still, the classifier can only recognize the hand gestures with a specific model, which is primarily affected by hand postures and rotation.

In order to enhance the accuracy of hand gesture segmentation, the model of skin color and the classifier can be combined. Firstly, the Gaussian mixture model of skin color is used to segment the area, possibly being the hand's area from the images. Then, the AdaBoost classifier is used to locate the hand gesture position. The specific flow is shown in Figure 4:

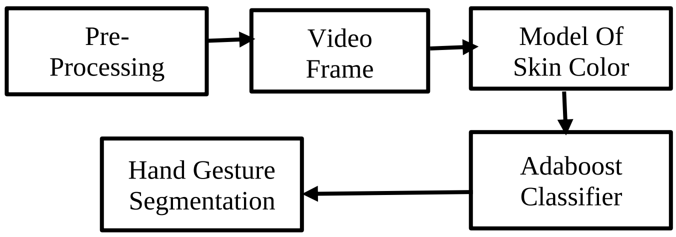

Fig. 4 : Flow Chart Ofhand Gesture Segmentation

The images of hand gestures are segmented according to the above flow; the following image shows the effect.

Fig 5. Hand gesture detection

III. HAND GESTURE TRACKING

The paper adopts the CamShift algorithm to track the targeted hand gestures with good robustness of the deformation of the targets. Moreover, since the algorithm is relatively simple, the hand gesture position can be detected in real-time.

A. Principle of CamShift algorithm

Camshaft algorithm is the continuous and adaptive target tracking algorithm and the improved version of the mean-shift algorithm.

The core of CamShift is MeanShift. On the premise of distance similarity measurement, the image gray similarity measurement is added. The main steps of CamShift are shown in the following:

1. In order to reduce the effect of lighting on hand gestures, one frame of video is transferred from RGB space to the HSV space; then the histogram statistics is carried out for H weight to calculate the color histogram HB ;
2. According to HB, make a reverse projection for the previous frame of the image to obtain probability distribution projection image of the last frame;
3. Operate the MeanShift algorithm.

Camshaft algorithm can monitor the hand gesture position in real-time and obtain the hand gesture area.

IV. HAND GESTURE RECOGNITION

According to hand gesture segmentation and hand gesture tracking, we can detect hand gestures in real-time. We adopt the classical LeNet-5 neural network[8] to recognize hand gestures. Date set of hand gestures includes ten categories of hand gestures ranging from 1 to 10 in the indoor environment, with 2000 pictures for each motion.

A. Principle of a convolutional neural network[9]

a) Convolutional layer

The convolutional layer realizes the convolutional process between the feature map in the last layer and one convolution kernel. The formula is:

$$\begin{cases} x_j^l = f\left(u_j^i\right) \\ u_j^l = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \end{cases}$$

b) Pooling layer

Pooling layers in CNNs summarize the outputs of neighboring groups of neurons in the same kernel map[10].

The pooling layer collects samples for the feature maps and segments them into multiple n' n image blocks that are not overlapped with each other, then carries out summation, the average or maximum value for each image block.

$$u^l = w^l x^{l-1} + b^l \qquad (13)$$

down(·) represents the down-sampling function.

c) Fully-connected layer and softmax function

After the convolutional layer and pooling layer, the feature maps of 2D images are connected as the 1D feature maps used to input a fully connected network. The output of the last fully connected layer is fed to a softmax which produces a distribution over the 10 class labels.

$$u^l = w^l x^{l-1} + b^l \tag{14}$$

In the formula, wl represents the weight of the fully connected layer, xl -1 represents the output value of the last layer, bl represents the bias, ul represents the output of the fully connected layer.

B. Network structure

The structure of LeNet-5 includes a convolutional layer, pooling layer, fully connected layer, and Softmax layer.

C. Experiments

The date for an experiment in the paper includes 10 categories of hand gestures ranging from 1 to 10 in the indoor environment, with 2000 pictures for each gesture.

Set the maximum steps for implementation as 10000 in total, set the batch as 32, set the fixed learning rate as 0.001, the second regularization parameter as 0.00004. then the loss curve and accuracy are shown in the following content:

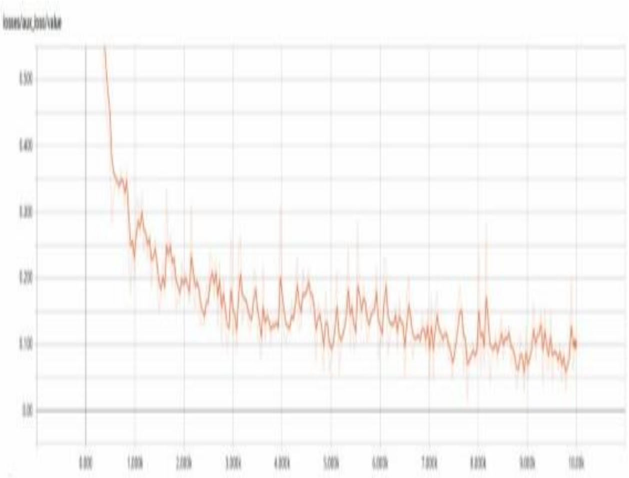

Fig. 6 Loss curve

The average accuracy of hand gesture recognition is 98.3%, the recognition rate for numbers 7 and 9 is not high because their hand gestures are complicated. Moreover, the network is unable to obtain the 3D information of hand gestures.

Real Figure	Hand gesture recognition 1	2	3	4	5	6	7	8	9	10	accuracy
1	395	3	0	0	0	0	1	1	0	0	98.75%
2	4	390	6	0	0	0	0	0	0	0	97.5%
3	0	3	394	3	0	0	0	0	0	0	98.5%
4	0	0	2	394	6	0	0	0	0	0	98.5%
5	0	0	0	2	398	0	0	0	0	0	99.5%
6	0	2	0	0	0	398	0	0	0	0	99.5%
7	7	3	0	0	0	0	380	0	4	6	95%
8	1	0	0	0	0	0	0	399	0	0	99.75%
9	1	0	0	0	0	0	7	0	386	6	96.5%
10	0	0	0	0	0	0	1	0	1	398	99.5%

Fig.7 Results of categories of hand gestures

V. CONCLUSION

The paper researches a set of overall flows for hand gesture recognition. Using AdaBoost classifier based on Haar feature, hand gesture segmentation realizes the acquisition of hand gesture area in a complicated environment. Using camshaft algorithm for hand gesture tracking according to the movement of hand gestures and features of deformation ensures to acquire the hand gesture area in real time, finally, the hand gesture area is classified by a convolution neural network.

VI. REFERENCES

[1]. SUN Li-juan, ZHANG Li-cai, GUO Cai-long. Technologies of Hand Gesture Recognition Based on Vision [J]. Computer Technology and Development, 2008, 18 (10) :214-216.

[2]. YI Jing-guo, CHENG Jiang-hua, KU Xi-shu. Review of Gestures Recognition Based on Vision [J]. Computer Science,2016,43(6A):103-108.

[3]. Gao long, Research on Static Gesture Recognition Algorithm Based on Neural Network [D]. Ning Xia University,2017.

[4]. Li guang-hua, Gesture Recogintion Technology Research and Application Based on Computer

Vision[D]. University of Electronic Science and Technology of China,2014.
[5]. Comaniciu D, Ramesh V, Meer P. Real-Time Tracking of Non-Rigid Objects Using Mean Shift[C]. Computer Vision and Pattern Recognition,2000.Proceedings.IEEE Conference on.IEEE,2000:2142.
[6]. Liu shi-lei, The study concerning human-computer interaction used in the manual segmentation and identification of key techniques[D]. Shan Dong University,2017.
[7]. R Lienhart,J Maydt. An extended set of haar-like features for rapid object detection[C].2002 International Conference on IEEE,2002,1:900-903.
[8]. CHEN Y N,HAN C C, WANG C T,et al.The application of a convolutional neural network on face and license plate detection[C] 18th international conference on pattern recognition.Hong Kong,China:IEEE,2006:552-555.
[9]. CIRESAN D C, MEIER U, MASCI J, et al. Flexible, high performance convolutional neural networks for image classification[C] IJCAI'11: Proceedings of the Twenty-Second International JointConference on Artificial Intelligence. Menlo Park, CA: AAAI Press,2011: 1237-1242.
[10]. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision[C]ISCAS.2010,2010:253-256.

Cite this article as :

Pranjali Manmode, Rupali Saha, Manisha N. Amnerkar, "Real-Time Hand Gesture Recognition", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 3, pp. 618-624, May-June 2021. Available at doi : https://doi.org/10.32628/CSEIT2173132
Journal URL : https://ijsrcseit.com/CSEIT2173132