A Static Gesture Recognition Method Based on Elliptical Skin Model and Deep Learning

Dongbo Zhao [1, a], Hui Li [2]

[1]School of Electronic Engineering, Xi'an Aeronautical University, Xi'an, 710077, China;

[2]School of Electronic Information, Northwestern Polytechnic University, Xi'an, 710129, China.

[a]alien_dffy@163.com

Abstract

The gesture recognition based on computer vision is a hotspot in human computer interaction. So that a static gesture recognition method based on ellipse skin model and deep learning was proposed. The skin area of hand was detected by ellipse skin model at first, which was faster and more accurate than other methods and then the samples of gesture picture were collected, classifying by its labels. The gesture database was trained and predicted off-line by the deep learning frame(Tiny-DNN) last, which solved the problem of palm tilt angle, and had strong robustness and high accuracy. Experimental results show that the algorithm can recognize different gestures in conditions of complex environment by simple video device.

Keywords

Elliptical skin model; deep learning; static gesture recognition.

1. Introduction

Gesture recognition is widely used in the fields of artificial interaction and intelligent recognition. It is a hot topic in pattern recognition. Many scholars have put forward a new algorithm [1] in this field. At present, gesture recognition is roughly divided into two categories: 1) hand gesture recognition based on equipment, common Kinect based gesture recognition [2] based on Microsoft and gesture recognition [3][4][5] based on other devices such as special depth cameras. This kind of algorithm can be achieved because of large company hardware support and commercial application. The recognition rate is high, but the cost is high; 2) based on the common camera gesture recognition, this method is simple, complex scene and low recognition rate because of the camera and other devices. Many researchers can only study [6][7][8] under the ideal test conditions. Gesture recognition in complex scenes is not ideal.

Gesture recognition process can be divided into gesture image acquisition, palm detection, gesture feature extraction, gesture recognition and other steps. At present, palm detection methods are based on YCrCb color space ellipse skin model palm detection [9], Haar and Adaboost palm detection [10]. Among them, the palmar detection based on the elliptical skin model has the advantages of fast speed and simple realization, but when there is a color similar to the skin, it often causes interference. The palm detection based on Haar and Adaboost has shortcomings such as inaccuracy and slow speed. Gesture recognition is based on a variety of classifier based methods, [2][11][12], [13] based on dynamic modeling, and a template matching based method, [14]. The accuracy of recognition depends on the selected palmar features. When the palm features can adapt to the change of the palm's tilt and size, the better results can be achieved.

In recent years, with the in-depth study of deep learning, the concept of deep learning has been introduced into many image domains, in order to improve the robustness and accuracy of [15]. Therefore, a static gesture recognition algorithm based on ellipse model and deep learning is proposed in this paper. First, the ellipse model skin detection is used to locate the general position of the palm, then the detected palm template is extracted, and the depth learning is used to train and predict. By collecting enough samples to increase the diversity of the training picture, it can effectively solve the problem of slanting in the palm recognition and the block deformation, and can achieve a higher recognition effect.

2. Palm Test

Palm detection is an important process of gesture recognition. Accurately locating the area in which the palm is located can greatly improve the accuracy of recognition. The palm skin detection method based on ellipse skin detection model is adopted in this paper.

2.1 Ellipse Skin Model

In the elliptical skin model, the color of the skin color after the nonlinear transformation is obviously elliptical in the image, and the distance between the two chromaticity components can be matched with the formula (1) (2).

$$\frac{(x-ec_x)^2}{a^2}+\frac{(y-ec_y)^2}{b^2}=1 \tag{1}$$

$$\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} C_b^{'}-c_x \\ C_r^{'}-c_y \end{bmatrix} \tag{2}$$

In the experiment, we can adjust the parameters according to the test results, and then $c_x=114.38$, $c_y=160.02$, $ec_x=1.60$, $ec_y=2.41$, $\theta=2.53$, $a=25.39$, $b=14.03$, we can get an ideal skin mask image.

2.2 Palm Extraction

When using the ellipse skin model to detect the skin area in the picture, some interference areas need to be removed. In the process of skin detection, the palm of the palm occupies the largest area, so the largest area in the picture is extracted, which is the area where the hand is located in the gesture recognition. To extract the palm process, after the ellipse skin model is detected, the contour of the palm is extracted, all the outlines are traversed, the maximum contour is searched, and the area of the largest area is obtained, which is the area of the hand gesture. Figure 1 uses the ellipse model to detect the skin region and extract the gesture area.

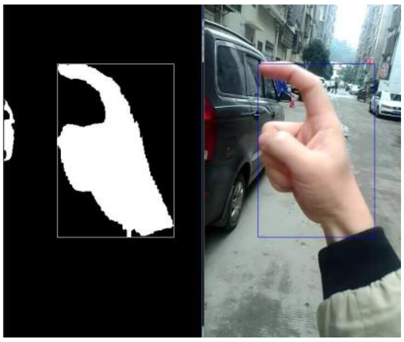

Figure 1. Elliptical skin model detection and extraction of gesture area

3. Deep Learning

After the concept of deep learning was put forward by Professor Geoffery Hinton of University of Toronto, various major companies have adopted various deep learning frameworks. There are three main current deep learning models, namely convolution neural network model, stack self coding network model and deep trust network model. The convolution neural network model is essentially a mapping relation of input to output. It can learn this mapping from a large number of data, but it does not need precise mathematical expressions. As long as it is trained with the known pattern to the convolution neural network, the neural network can have this mapping ability [16]. The convolution neural network training is divided into two stages:

1) The forward propagation stage. A sample is extracted from the training sample and will be input into the network. The information is transferred from the input layer to the output layer by step by step, and the corresponding actual output is calculated.

$$O_p = F_n\left(\cdots\left(F_2\left(F_1\left(XW_1\right)W_2\right)\cdots\right)W_n\right) \tag{3}$$

2) The backward propagation stage. It is also called the phase of error propagation.

The difference between the actual output and the ideal output is calculated.

$$E_P = \frac{1}{2}\sum\nolimits_j \left(y_{pj} - o_{pj}\right)^2 \tag{4}$$

The weight matrix is adjusted according to the method of minimizing error.

In this paper, we use the deep learning library Tiny-DNN based on convolution neural network model. The library input many training samples, through the forward propagation of the middle layer, and then continue to propagate backward, get the minimum error, repeated cycle, get the best effect. The training process is shown in Figure 2

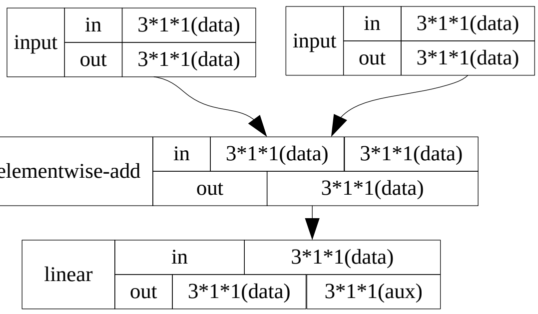

Figure 2. The flow chart of Tiny-DNN framework training

3.1 Make And Train Hand Gesture Recognition Samples

3.1.1 Production Samples

Using an imaging device to shoot a variety of different gestures or videos. When each gesture is taken, a number of rotation angle gestures are made, the sample pictures of each gesture are classified and put into the corresponding named folders, using the elliptical skin to detect the model, and the detected gesture mask images are put into the image. The location of another folder is shown in Figure 3. Check the sample pictures in the folder and manually remove the samples that do not meet the requirements. The algorithm has produced 7 sets of gesture recognition samples from 3 to 9. If you need to recognize other gestures, you can take photos and add them yourself. Figure 4 is an example of a variety of different gesture samples.

3.1.2 Training Sample

Each gesture mask sample is put in the corresponding folder, then the deep learning Tiny-DNN framework is used to read all the pictures according to the corresponding numbers. The minimum collision space is set and the number of collisions is trained. The more the number of collisions, the stronger the self learning ability and the more accurate the training results are.

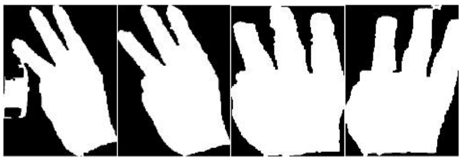

Figure 3. Sample of gestures with different rotation angles (with gesture 3 as an example)

Figure 4. Samples of different gesture sets (from 3 to 9)

3.2 Prediction And Statistical Results

Use imaging equipment to capture test sets and put them in corresponding lives.

In the name folder, the palm is detected according to the skin model of the ellipse. The mask image of the palm is put into the depth learning prediction frame, and the category of the palm is judged, and the statistical time and the correct rate are carried out. The concrete results are given and analyzed in the experiment.

4. Experiment

In order to verify the algorithm compared with other static picture hands the potential recognition algorithm is more advantageous. In this paper, we test the hand gesture recognition data set in its own, select the document [14] based on the Hu matrix gesture recognition, the document [11] is based on the multi model UAV gesture recognition, the literature [6] is based on the HOG and the related gesture recognition, and the literature [7] is based on the gesture recognition of the positive phase space. This algorithm uses hand gesture recognition data to shoot in the field, and the scene is very complicated. A total of 1141 data test sets of gestures from 3 to 9 are collected, each gesture has a slight angle of rotation, and then the average time and the correct rate used for each algorithm are counted. This experiment adopts the 64 bit operation system in win10, and develops software using VC2015's Win32 bit and Tiny-DNN library programming. Processor bit Intel i3-6000, the main frequency is 3.70GHz, not using multi-core parallel and GPU acceleration technology.

4.1 Palm Test

Palm detection is more adequate in outdoor light. A number of different gesture videos are captured by ordinary collection, and then each video is processed into a picture set according to different gestures into the corresponding folder. The ellipse skin detection model is used to get the drowning mode of gesture recognition, and the result is shown in Figure 5.

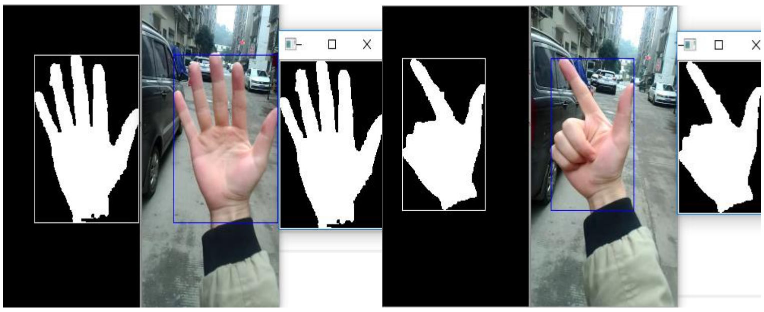

Figure 5. The effect of different palm testing

Then the correct rate of palm testing was calculated. In the experiment, 1141 pictures were tested. The results showed that 1141 areas were correctly detected in the palm of the palm, and the correct rate of palm test was 100%. From Figure 5, it can be seen that the ellipse skin detection model has good detection performance. At the same time, the detection principle of the model is very simple, so it is easy to implement, and the speed of detection is faster.

4.2 Time Comparison

The algorithm mentioned in the literature mentioned above is tested with the training set obtained by oneself, and statistics the time of each algorithm to locate palm and recognize gesture. All algorithms are carried out under the same experimental hardware and software. The average recognition time statistics of each algorithm are shown in Table 1.

Table 1. Time comparison of various algorithms

algorithm	Document [14]	Document [11]	Document [6]	Document [7]	This paper
times(ms)	23	46	86	29	59

As can be seen from table 1, machine learning and deep learning are the least used. Using machine learning recognition is less than using depth learning, and the document [6] uses more HOG features. Deep learning is more complex than theory, and it is the most important time to achieve it.

4.3 Correct Rate Comparison

As compared with the time comparison method, the correct rate of various algorithms is obtained under the same condition, the correct rate of each algorithm is counted, and the average value is retaken, as shown in Table 2.

Table 2. Comparison of the accuracy of various algorithms

algorithm	Document [14]	Document [11]	Document [6]	Document [7]	This paper
Correct rate (%)	62	78	68	82	97

As can be seen from table 2, document [14] uses traditional Hu contour features to recognize gestures, which can not resist rotation and deformation, and has the lowest detection rate. Literature [11], such as SVM, Bias and other machine learning methods, improve the recognition rate, but the correct rate of machine learning has a great relationship with the selected features. HOG and other features can not effectively deal with the angle of gesture images, so it can not achieve a very high recognition rate. In this paper, the ellipse skin detection model is used to obtain a very high detection rate of the palm. Because of the deep learning, the correct characteristics are obtained by continuous self learning and adjusting the parameters, so that the correct rate is greatly improved. The accuracy of deep learning depends on training samples, and the accuracy of recognition can be further improved by selecting enough and correct samples.

4.4 The Correct Rate of Hand Gesture Recognition in This Algorithm

In order to improve the accuracy of the algorithm, it is necessary to make clear the correct rate of each gesture, so as to process the training samples of the gesture, and remove some of the samples that do not meet the requirements out of the training set, and increase some common samples.

Table 3. Comparison of the accuracy of different gestures

gesture	3	4	5	6	7	8	9
Correct rate (%)	97.6	99.9	84.8	100	99.9	100	100

As you can see from table 3, when the gesture is 5, the recognition accuracy is low, because the small finger or thumb is out of the screen when the angle of the gesture 5 is too large, resulting in a wrong classification. The method is to increase the distance between the palm and the camera, let the palm of the palm in the camera range, while the training of the gesture recognition does not meet the requirements of the sample, and increases some samples of different angles, thus improving the accuracy of recognition.

5. Conclusion

In this paper, a gesture recognition algorithm based on the ellipse skin detection model and depth learning is proposed, which makes use of depth learning for gesture recognition, which greatly improves the accuracy and robustness. In traditional methods, though detecting the Hu characteristics of the skin region, although the speed is fast, the application scenario is simple and the accuracy rate is not high. Some method of gesture recognition based on machine learning is not very high in complex scenes because the selected features can not adapt to the left and right tilt of gestures and the change of palm shape. The method used in this paper can well solve the impact of angle, shape and other aspects on recognition, and has good recognition effect. It is stable and efficient in the system with low time requirement and complex background.

References

[1] J. Ma and Y. Du. Study on the evaluation method of in-vehicle gesture control [J]. 2017 3rd IEEE International Conference on Control Science and Systems Engineering (ICCSSE), Beijing, China, 2017, pp. 145-148.Method of in-vehicle gesture control [J]. 2017 3rd IEEE International Conference on Control Science.

[2] D. Nguyen and H. S. Le. Kinect Gesture Recognition: SVM vs. RVM [J]. 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, 2015, pp. 395-400.Recognition: SVM vs. RVM [J]. 2015 Seventh International Conference on Knowledge and Systems Systems (E), vs.,

[3] Cao Chuqing, Li Ruifeng, Zhao Lijun. Based on depth image. Gesture recognition method of technology [J]. Computer Engineering 2012,22 (8): 16-21.

[4] Lu Lei, Zhang Jinling, Zhu Yingjie, Liu Hong. Data glove static gesture recognition method [J]. Computational assistant design and graphics Journal.2015,27 (12): 2410-2418

[5] R. Augustauskas and A. Lipnickas, Robust hand detection using arm segmentation from depth data and static palm gesture recognition [J]. 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, Romania, 2017, pp. 664-667.Hand detection using arm segmentation from depth data and static palm gesture.

[6] V. Sombandith, A. Walairacht and S. Walairacht. Hand gesture recognition for Lao alphabet sign language using HOG and correlation [J].2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 2017, pp. 649-651Walairacht. Hand gesture recognition for Lao alphabet sign language using HOG and.

[7] B. B. Gatto, E. M. DOS Santos and W. S. D. Silva. Orthogonal Hankel Subspaces for Applications in Gesture Recognition [J]. 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Niterói, Brazil, 2017, pp. 429-435.Silva. Orthogonal Hankel Subspaces for Applications in Gesture Recognition [J]. 2017 30th SIBGRAPI Conference.

[8] Joshi, S. Ghosh, M. Betke, S. Sclaroff and H. Pfister. Personalizing Gesture Recognition Using Hierarchical Bayesian Neural Networks [J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 455-464.Pfister. Personalizing Gesture Recognition Using Hierarchical Bayesian Neural Networks [J]. 2017 IEEE Conference.

[9] X. Tan, Yueqiang Jin, Guizhen Feng and Xiao Jiang. Gesture segmentation based on YCrCb ellipse skin model and background subtraction [J]. 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), Nanjing, 2015, pp. 322-326.Xiao Jiang. Gesture segmentation based on YCrCb ellipse skin model and background.

[10] Y. Yu, S. Bi, Y. Mo and Weiheng Qiu. Real-time gesture recognition system based on Camshift algorithm and Haar-like feature [J]. 2016 IEEE International Conference on Cyber Technology in

Automation, Control, and Intelligent Systems (CYBER), Chengdu, 2016, pp. 337-342.Real-time gesture recognition system based on Camshift algorithm and Haar-like feature [J]. IEEE.

[11] Y. Ma et al. Hand gesture recognition with convolutional neural networks for the multimodal UAV control [J]. 2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS), Linköping, Sweden, 2017, pp. 198-203.Convolutional neural networks for the multimodal UAV control [J]. 2017 Workshop on Research.

[12] Joshi, S. Ghosh, M. Betke, S. Sclaroff and H. Pfister. Personalizing Gesture Recognition Using Hierarchical Bayesian Neural Networks [J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 455-464.Pfister. Personalizing Gesture Recognition Using Hierarchical Bayesian Neural Networks [J]. 2017 IEEE Conference.

[13] Xiao Ling, Li Renfa, van van Tsai. Based on self learning sparsity. Dynamic gesture recognition method expressed [J]. Journal of communication, 2013, 34 (6): 128-135.

[14] Zhao Jian, Zhang Dongquan. Digital gesture based on OpenCV. Recognition algorithm [J]. computer application.2013,33 (S2): 193-196.

[15] G. Hinton et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups [J]. in IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82-97, Nov. 2012.Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups[J].

[16] Liu Jianwei, Liu Yuan, Luo Xionglin. Deep learning research. Exhibition [J]. computer application research.2014,31 (7): 1921-1931