Journal of Physics: Conference Series

Hand Gesture Recognition with Skin Detection and Deep Learning Method

Hand Gesture Recognition with Skin Detection and Deep Learning Method

Hanwen Huang[1], Yanwen Chong[2*], Congchong Nie[2], Shaoming Pan[2]

[1] College of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

[2] State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

*corresponding author: ywchong@whu.edu.cn

Abstract. Gesture recognition, although has been exploring for many years, is still a challenging problem. Complex background, camera angles and illumination conditions make the problem more difficult. Thus, this paper presents a fast and robust method for hand gesture recognition based on RGB video. First we detect the skin based on their color. Then we extract the contour and segment the hand region. Finally we recognize the gesture. The results of experiment demonstrate that the proposed method are efficient to recognize gesture with a higher accuracy than the state of the art.

1. Introduction

There has been great emphasis on Human-Computer-Interaction research to create easy-to-use interfaces by directly employing natural communication and manipulation skills of humans [1]. As an important part of the body, recognizing hand gesture is very important for Human-Computer-Interaction. In recent years, there has been a tremendous amount of research on hand gesture recognition [2]. While there are numerous researches focused on this topic, there are still several problems to be solved. The speed and accuracy are two main characteristics of the algorithm, thus a robust and fast method is needed to improve user experiences.

A contour based method for recognizing hand gesture using depth image data is shown in [3]. Using depth data, these method can distinguish the hand from background easily without getting confused by background color. However, depth data are not common and easily available while RGB data solves the problem. In [4], a hierarchical method of static hand gesture recognition that combines finger detection and histogram of oriented gradient (HOG) feature is proposed. An algorithm applied for locating fingertips in hand region extracted by Bayesian rule based skin color segmentation is proposed in [5]. In [6], the continuous gesture recognition problem is tackled with a two streams Recurrent Neural Networks (2S-RNN) for the RGB-D data input. These methods all get good effect, but they are not that efficient. This paper provides a more efficient way based on deep learning.

2. Proposed gesture recognition system

The proposed gesture recognition system (shown in figure 1) composed of three main parts. In the first part, we get RGB video of the user, then after threshold segmentation, the video become a binary video. Meanwhile, some preprocessing is needed, like expansion and corrosion. In the second part, we extract all the contours and find out the contour of hand based on several indexes. Finally, the gesture can be recognized using the pyramidal pooling module and attention mechanism.

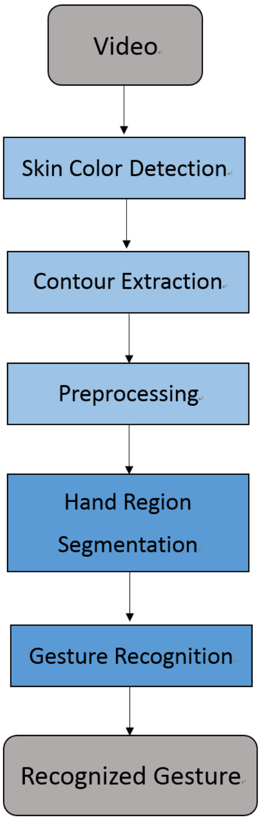

Figure 1. Gesture recognition system based on RGB video.

3. Skin Color Detection

In order to detect the skin from the video, we need to find out the character of the skin. According to [7], detecting skin-colored pixels, although seems a straightforward easy task, has proven to be quite a challenging task in images that are captured under complex unconstrained imaging conditions. So we developed a method based on the color feature for most human. The formula is shown as below.

$$R>85$$
$$R-B>10$$
$$R-G>10$$

Based on such criterion, we can efficiently segment the skin from the background, which can be considered to be human part. Then we can convert the image into binary image. The results are shown as figure 2.

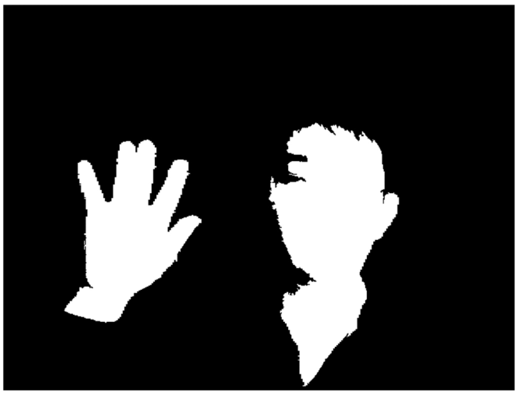

Figure 2. Binary image.

4. Preprocessing

Due to the nature of image, the hand region may have holes and cracks, which will definitely affect the accuracy of hand gesture. Usually the binary image will be noisy, so image preprocessing is necessary, which fills the holes. According to [8], there are two main methods for digital image restoration, texture-based method and non-textured-based method. In this paper, an exemplar image completion based on evolutionary algorithm is proposed. In the non-textured-based method, total variation method is a typical algorithm. An improved total variation algorithm is proposed in this paper. In the improved algorithm, the diffusion coefficients are defined according to the distance and direction between the damaged pixel and its neighborhood pixel. Although in other papers [9], a novel inpainting methods reach a good result. However, the methods are always too sophisticated and time-costing. Thus we employ some simple morphological operations like erosion and dilation to fulfill our request. In practice, we dilate two times and erode two times.

5. Contour Extraction and Hand Region Segmentation

After we remove the noise in the image, we need to extract contours. We consider each point cluster as a contour. Among these contours, there is only one contour which represents the hand region. Furthermore, hand region and face region are the largest two contours. Based on such fact, the problem of finding the hand from the contours becomes the problem of separating hands from faces. So we collect 100 samples of face region and hand region. Then we use VGGNet in [10] to classify them. VGGNet is a deep convolution neural network developed by the Visual Geometry Group and researchers at Google DeepMind. VGGNet explores the relationship between the depth and performance of convolutional neural networks. By stacking 3×3 small convolution cores and 2×2 maximum pool layer repeatedly, VGGNet has successfully constructed 16-19 layers (convolution layer and fully connected layer) deep convolution neural networks. In this paper, we use 16 layers VGGNet.

6. Gesture Recognition

In this paper, the pyramidal pooling module and attention mechanism are used to increase the receptive field and classify the details more efficiently. As can be seen from figure 3, the original input image passes through the convolution layer of 3 × 3 and the maximum pooling layer to get the size of 1/2 of the original image feature map. Then the four different scale spatial pyramids are pooled to get the size of 1/4, 1/8, 1/16, 1/32 feature map respectively. So the different scale features can be captured. Then, global average pooling is used to obtain the weights of global abstract features as channel dimensions at lower levels. Finally, the final probability score of each class is obtained by using fully connected layer and softmax. Comparing with stacked convolution and pooling structure, the proposed structure can not only get the feature maps of different receptive fields quickly, but also pool the high-resolution feature maps globally, and then reduce the channel dimension by 1 × 1 convolution, which can be used as the weight of adjacent low-resolution feature channels effectively. Guide the formation of abstract features. A large number of experiments on our gesture data show that this structure can accelerate the convergence of the network and improve the accuracy of recognition.

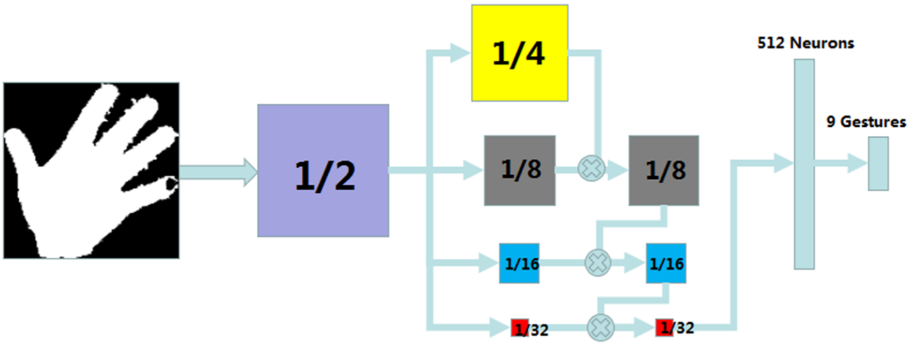

Figure 3. Training network structure.

7. Results

To validate the method proposed in this paper, we conducted two experiments. The first experiment is hand region segmentation. We use 100 hand region samples with different gesture and 100 face region samples. We take 70 samples as test data and 30 samples as validation data. The results are shown in figure 4 and figure 5. The accuracy of segmentation is 98.48%. To sum up, the experimental accuracy can meet actual needs.

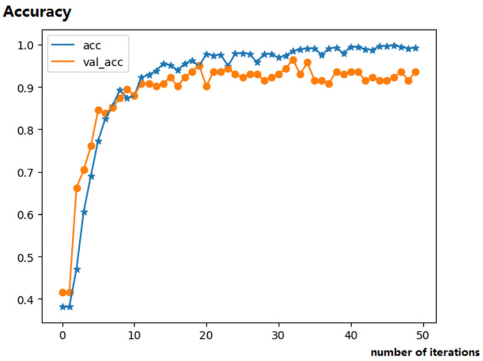

Figure 4. Hand region segmentation accuracy.

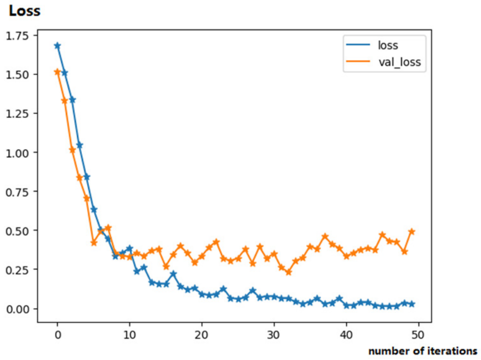

Figure 5. Hand region segmentation loss.

The second experiment is gesture recognition. To detect 9 common gestures (shown in figure 6), a total of 900 test samples form 5 people were tested. Each gesture have 70 test data and 30 validation data. The recognition results are shown in figure 7 and figure 8.

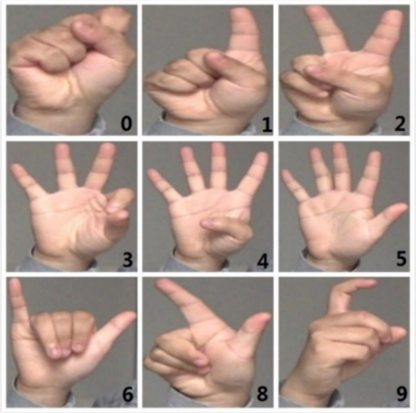

Figure 6. 9 Common gestures.

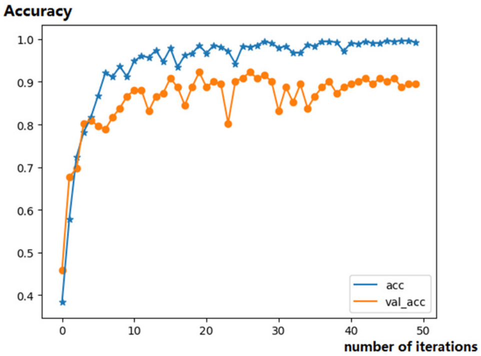

Figure 7. Gesture recognition accuracy.

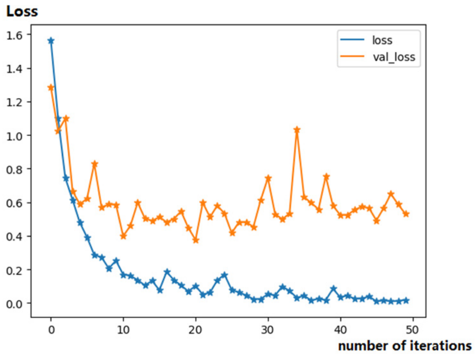

Figure 8. Gesture recognition loss.

As shown in these figures, we can notice that the method proposed in this paper has a high accuracy. The overall accuracy has reached 98.41%, which shows the effectiveness of the method. Furthermore, we also report the number of frames per second that can be processed. It can reach the speed of 10fps, which can achieve real-time gesture recognition. Comparing to other recent research, this method has higher accuracy. According to [11], in low dimensionality, the recognition rate of MAOMP algorithm can still be more than 85% and the recognition rate of SAMP, SP and SWOMP is maintained at 80–85%, while the ROMP and OMP algorithms are less than 80%. In [12], the paper comes up with a deep learning-based fast hand gesture recognition method using representative frames. The accuracy of the method is 95.18%. In [13], the paper uses appearance features and 3D Point Cloud to recognize gesture. It can recognize 9 gestures and the overall accuracy reaches 94.7%.These results show the accuracy and efficiency.

Table 1. Accuracy comparison of different methods

Methods	MAOMP	SAMP,SP, SWOMP	ROMP, OMP	Paper [14]	Paper [15]	Method in this paper
Accuracy	85%	80%~85%	80%	95.18%	94.7%	98.41%

8. Conclusion

Gesture plays an indispensable role in Human-Computer-Interaction. This paper comes up with an accurate and efficient method for gesture recognition. First, skin is detected using on rules based on experience and the picture is transformed into binary image. Then expansion and corrosion are adopted. After that, all the contours are extracted and the contour of hand are found. Finally, the gesture can be recognized using the pyramidal pooling module and attention mechanism.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61572372, 41671382), State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing Special Research Funding.

References

[1] Ren Z, Meng J, Yuan J. Depth camera based hand gesture recognition and its application in Human-Computer-Interaction[C]// Communications and Signal Processing. IEEE, 2011:1-5.

[2] Kang S K, Mi Y N, Rhee P K. Color Based Hand and Finger Detection Technology for User Interaction[C]// International Conference on Convergence and Hybrid Information Technology. IEEE, 2008:229-236.

[3] Le T N, Cong D T, Ba T N, et al. Contour Based Hand Gesture Recognition Using Depth Data[C]// The, International Conference on Signal Processing, Image Processing and Pattern Recognition. 2013:60-65.

[4] Liu S, Liu Y, Yu J, et al. Hierarchical static hand gesture recognition by combining finger detection and HOG features[J]. Journal of Image & Graphics, 2015.

[5] Bhuyan M K, Neog D R, Kar M K. Fingertip Detection for Hand Pose Recognition[J]. International Journal on Computer Science & Engineering, 2012, 4(3).

[6] Chai X, Liu Z, Yin F, et al. Two streams Recurrent Neural Networks for Large-Scale Continuous Gesture Recognition[C]// International Conference on Pattern Recognition. IEEE, 2017:31-36.

[7] Chen W, Wang K, Jiang H, et al. Skin color modeling for face detection and segmentation: a review and a new approach[J]. Multimedia Tools & Applications, 2016, 75(2):1-24.

[8] Li K, Wei Y, Yang Z, et al. Image inpainting algorithm based on TV model and evolutionary algorithm[J]. Soft Computing - A Fusion of Foundations, Methodologies and Applications, 2016, 20(3):885-893.

[9] Alexandru Telea. An Image Inpainting Technique Based on the Fast Marching Method[J]. Journal of Graphics Tools, 2004, 9(1):23-34.

[10] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, Proc. Int'l Conf. Learning Represent., 2015.

[11] Li B, Sun Y, Li G, et al. Gesture recognition based on modified adaptive orthogonal matching pursuit algorithm[J]. Cluster Computing, 2017(3):1-10.

[12] John V, Boyali A, Mita S, et al. Deep Learning-Based Fast Hand Gesture Recognition Using Representative Frames[C]// International Conference on Digital Image Computing: Techniques and Applications. IEEE, 2016.

[13] Chong, Y, Huang, J. and Pan, S. (2016) Hand Gesture Recognition Using Appearance Features Based on 3D Point Cloud. Journal of Software Engineering and Applications, 9, 103-111. doi:10.4236/jsea.2016.94009.