ATLANTIS PRESS
Advances in Computer Science Research, volume 71
4th International Conference on Machinery, Materials and Information Technology Applications (ICMMITA 2016)

Hand Gesture Recognition Using Deep Neural Network and Its Implementation in Augmented Reality

Huayong Yang, Xiaoli Lin

Department of Information Engineering, Wuhan College of Foreign Languages & Foreign Affairs, Wuhan 430083, China

Keywords:landmark detection, gesture recognition, object recognition, augmented reality.

Abstract. In this paper, we study the augmented reality and its application in intelligent human-computer interaction. The complex structure in the visual background is a major challenge for gesture segmentation. First, we use a robust skin color segmentation method to preprocess the input image. Second, visual texture features are analyzed and modeled for hand region recognition. Third, finger landmarks are annotated and configured by active shape model. Finally, the hand gestures are recognized and used for enhanced interaction. Experimental results show that the proposed gesture recognition system is robust against various background changes and illumination changes.

1. Introduction

Augmented Reality(AR) is a popular problem in computer vision. Gesture recognition is another important topic in computer vision[1,2,3]. It has many applications in Human-Computer Interaction (HCI). A robust localization is the key step for hand gesture modelling. The illumination changes and background changes make the localization problem more challenging.

Various algorithms have been studied in the field of landmark localization. Zhang et al.[4], proposed to use deep neural network to detect landmarks. The robustness was improved and a large amount of data was also required. Zhou et al.[5], used the convolutional network in a coarse-to-fine framework and it was robust against different backgrounds. The computational burden was relatively high. Sangineto [6] proposed to use the dense-SURF feature to locate landmarks.

In this paper we study the gesture recognition and its application in augmented reality. We propose to use skin detection and hand ROI (Region of Interest) detection for a robust initialization in landmark localization. Based on the robust landmark localization the gesture recognition is improved. The gesture recognition module is a part of our AR system. The system consists of camera, hand localizer, gesture recognition, scene generator, video generator and displayer.

2. Hand ROI Detection

The hand ROI is detected by two steps. First, the skin color is modelled and the pixels are labelled when the color is close to human skin. Second, the LBP (Local Binary Pattern) feature is extracted and the hand ROI is modelled by deep believe network (DBN). A sliding window is used for detection based on the DBN model.

The distance is measured by Euclid distance in the RGB color space and a threshold th_{skin}^{d} is used to decide whether the pixel belongs to human skin.

$$d = \sqrt{(b^{pixel} - b^{skin})^2 + (g^{pixel} - g^{skin})^2 + (r^{pixel} - r^{skin})^2} \quad (1)$$

A further skin classification is carried out for improved accuracy. The HSV(Hue, Saturation, Value) color space, YCbCr color space and TSL color space is used.

TSL color space is defined as following. The thresholds that used for the fine skin classification is achieved from a large testing dataset. The rules are manually adjusted for the acceptance of skin color as following:

Rule No.1: $Y > 60$ & $80 < Cb < 135$ & $130 < Cr < 175$ in YCbCr space.

Rule No.2: $0 < H < 410$ & $0.14 < S < 0.9$ in HSV space.

Rule No.3: $0.4 < T < 0.8$ & $0.02 < S < 0.5$ & $L > 60$ in TSL space.

The hand ROI detection results are shown in Fig.1. Under complex backgrounds, we can see from Fig.1 that the skin detection accuracy has dropped. Therefore the hand shape detection can only be used as a coarse initial step and more accurate localization algorithm is required.

(a)(b) (c) (d) (e)

Fig.1. Hand ROI Detectionin Complex background challenge: (a) Skin detection; (b) Hand ROI detection; (c) Original image with background; (d) Skin detection result; (e) Labelled hand region.

3. Landmark Detection

Both skin detection and Hough line detection are used to initial the Active Shape Model (ASM). High Order Markov Random Fieldsis used to configure the landmark detection results based on a pre-learnt shape model.

The shape model is represented as: $\hat{\boldsymbol{x}} = \overline{\boldsymbol{x}} + \Phi \boldsymbol{b}$. $\hat{\boldsymbol{x}}$is the shape vector generated from hand landmarks, $\overline{\boldsymbol{x}}$ is the mean shape vector, Φ contains the eigenvectors of covariance matrix from the distortion and $\boldsymbol{b}$ is a parameter for generating different shapes. As shown in Fig.2, the complex backgrounds and illumination changes make this task more challenging. Therefore we use the skin and hand ROI detection described in Sec.2 to get the initial position for ASM algorithm.

(a) (b)

Fig.2. Hand ROI Detection: (a) Landmark detection; (b) Illumination change.

4. Gesture Recognition

Support vector machine (SVM) is adopted for gesture recognition. The kernel function of SVM removes the rotation influence and separates different classes in high dimensional space.

In SVM the data set is represented as follows:

$$(\boldsymbol{f}_1, y_1), \ldots, (\boldsymbol{f}_n, y_n) \tag{2}$$

where$y$ is the class label, n is number of samples, and $\boldsymbol{f}$ is the feature vector.

A hyperplane can be written for the classification of the data points:

$$\boldsymbol{w}\boldsymbol{f} - b = 0 \tag{3}$$

where$\boldsymbol{w}$ is the normal vector to the hyperplane, and b determines the offset.

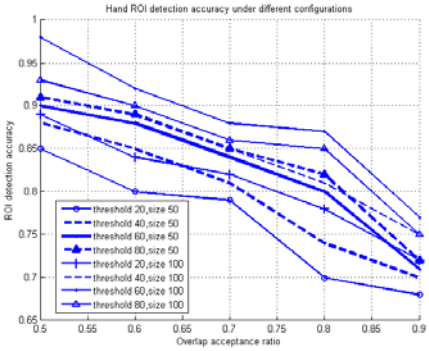
In SVM the optimization problem is to minimizing the following equation:

$$\left[\frac{1}{n}\sum_{i=1}^{n} m\,ax(0, 1 - y_i(\boldsymbol{w}\boldsymbol{f}_i + b))\right] + \lambda|\boldsymbol{w}|^2 \tag{4}$$

whereλ is a parameter determines the tradeoff between increasing the margin-size and ensuring that the data sample point lies on the correct side.

5. Experimental Result

We carry out three experiments to verify the effectiveness of the proposed algorithm. First, we test the localization accuracy of hand ROI (Region of Interest) detection. Second, we test the accuracy of localizing the tip of finger. Third, the recognition rate of different gesture types is tested.



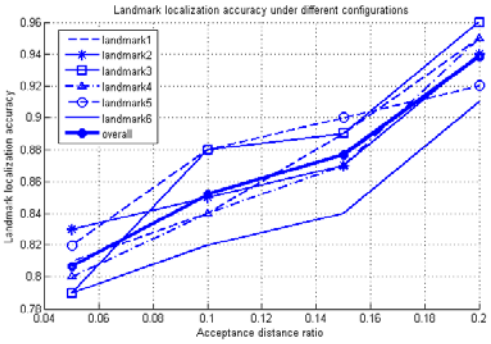


Fig.3. Hand ROI detection accuracy. **Fig.4. Landmark localization accuracy.**

As shown in Fig.3, the hand ROI detection rate is dependent on the threshold th_{skin}^{d} in the skin color detection stage and the training set size s^t in DBN model. The accuracy is measured by the overlap area of the hand ROI. A success detection of the ROI is defined by the ratio of overlap area. Generally above 80% overlap between the detected ROI and the ground-truth ROI is accepted in practical applications. From Fig.4, we can see that the success rate changes along the overlap acceptance ratio.

As show in Fig.4, the landmark localization accuracy is demonstrated. The Y-axis is the percentage of successfully located landmark and the X-axis is the acceptance distance ratio. The distance ratio is defined as $|P_l - bolP_g|_{L2}/(W+H)*2$, where P_l and P_g are the detected landmark location and the ground-truth landmark location, W and H are the width and height of hand ROI respectively.

We further compare the detection accuracy using the standard ASM algorithm without skin detection initialization and the proposed algorithm with skin detection, as shown in Tab.1.

Tab.1. Comparison between standard ASM and skin initialization.

Acceptance distance ratio	Standard ASM	Proposed algorithm	Improvement
0.05	80.67%	75.61%	5.06%
0.1	85.17%	83.24%	1.93%
0.15	87.67%	81.38%	6.29%
0.2	93.83%	89.21%	4.62%

As shown in Tab.2, the final gesture recognition accuracy is tested. Five different hand gestures are included. The training and testing ratio ranges from 10 to 5. We can see the recognition results are promising and the highest averaged rate reaches 92.0%. Compared with standard ASM, the improvement is show in the final column. This proves the effectiveness of our proposed landmark detection algorithm.

Tab.2. Final recognition accuracy on five hand gestures.

Ratio	Gesture I	Gesture II	Gesture III	Gesture IV	Gesture V	Average	Improvement
10	93.4%	95.6%	93.1%	89.3%	88.4%	92.0%	4.4%
9	92.1%	94.5%	92.3%	87.3%	86.4%	90.5%	5.4%
8	89.2%	92.3%	90.7%	86.1%	85.4%	88.7%	5.2%
7	88.3%	90.3%	88.3%	84.3%	82.9%	86.8%	6.3%
6	87.4%	87.3%	86.5%	82.4%	80.3%	84.8%	7.4%
5	87.1%	86.5%	86.1%	81.1%	79.9%	84.1%	4.9%

6.Summary

In this paper the robust skin detection and application in augmented reality is studied. Based on a good initial position the ASM landmark localization results are improved. The illumination changes and the background changes are discussed in this paper. DBN model is used for ROI modelling and SVM model is used for gesture recognition.Experimental results shownthat the proposed system has a promising application in augmented reality.

Acknowledgement

This work is supported by 2015 annual project of Hubei Province Education Science Planning (No. 2015GB343)

References

[1] Z. Ren, J. Yuan, J. Meng and Z. Zhang, Robust part-based hand gesture recognition using kinect sensor, *IEEE transactions on multimedia*, vol.15, no.5, pp.1110-1120, 2013.
[2] S. S. Rautaray and A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, *Artificial Intelligence Review*, vol.43, no.1, pp.1-54, 2015.
[3] Q. Pu, S. Gupta, S. Gollakota and S. Patel, Whole-home gesture recognition using wireless signals, *In Proc. of the 19th annual international conference on Mobile computing & networking*, pp.27-38, 2013.
[4] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, Facial landmark detection by deep multi-task learning, *Pro. In European Conference on Computer Vision*, pp.94-108, September, 2014.
[5] E. Zhou, H. Fan, Z. Cao, et al., Extensive facial landmark localization with coarse-to-fine convolutional network cascade, *Proc. of the IEEE International Conference on Computer Vision Workshops*, pp.386-391, 2013.
[6] E. Sangineto, Pose and expression independent facial landmark localization using dense-SURF and the Hausdorff distance, IEEE transactions on pattern analysis and machine intelligence, vol.35, no.3, pp.624-638, 2013.