Research on Fast Recognition Algorithm of Golf Swing

Qiao Tian, Jingmei Li(✉), Fangyuan Zheng, and Chao Lv

College of Computer Science and Technology, Harbin Engineering University, Harbin, China
lijingmei@hrbeu.edu.cn

Abstract. Due to the target recognition algorithm has high time complexity, a fast recognition algorithm of golf gesture based on video sequence is proposed. Firstly, the detector locates the salient region of image, then the gesture detector scans the fraction sequence generated by the video and the sequence is taken as feature data, finally, the linear support vector machine does real-time judgment of the data, thus completing the fast recognition. The experiments show that the recognition speed is over 30 fps and the accuracy of 97% can be achieved on iPhone5s and later version, proving the validity in practical application.

Keywords: Golf gesture recognition · Video sequence · Fraction sequence

1 Introduction

Gesture recognition in video sequence [1, 2] belongs to the research hotspot in computer vision. It is the basic function that auxiliary training equipment must have.

After fully studying the target recognition algorithm [3] based on machine learning, the thesis proposed a golf gesture recognition algorithm based on video sequences, and the experiments verify the algorithm effectiveness in practical application.

2 A Golf Gesture Recognition Algorithm in Video Sequence

The golf gesture recognition in video sequence adopts the machine learning method which is divided into training model and test phase, as shown in Fig. 1.

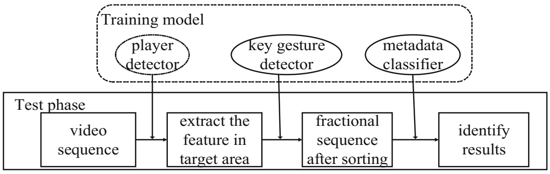

Fig. 1. Schematic diagram of video sequence

G. Sun and S. Liu (Eds.): ADHIP 2017, LNICST 219, pp. 505–509, 2018.
https://doi.org/10.1007/978-3-319-73317-3_59

2.1 Training Model

In training phase, the thesis learns three models: a player detector, several key gesture classifiers and a metadata classifier. These classifiers are shown below.

The player detector is trained by Dollar's fast pedestrian detection algorithm [4] in 2014, it is faster than other recognition algorithms such as HOG + SVM and DMP + SVM [5], because the recognition speed is 30 + fps in natural environment. The labeled player is selected as positive dataset, and the non-human region is as counter dataset. The trained detector acts as player detector, which locates the player in image to reduce the search range and improve the recognition speed, as shown in Fig. 2.

Fig. 2. The location graph of player with player detector

The golf swing action is defined as the combination of key actions, such as swing, back swing, down swing, batting and end swing. The self-created data set is selected to train on the key action classifier which is used to do real-time scoring of the images in the video sequence, the training key actions are selected as positive samples and the others as counter samples, as shown in Fig. 3.

Fig. 3. The key gesture classifier graph in training learn

The key action classifier group learned by training receives many fraction sequences while scanning input video sequence, but the sequence obtained by a single key action detector cannot judge the occurrence of action, therefore a metadata classifier is required, all the fraction sequences are judged simultaneously by a fixed time window L. The training dataset intercepts all fraction sequences through L and regroups them, selecting a fraction sequence that contains only one crest as positive samples, and the others as counter samples, as shown in Fig. 4.

The metadata classifier is learned by linear support vector machine (LSVM), its result is a fraction. This thesis sets the threshold for it, so the "positive" or "negative" judgment action can be simply output.

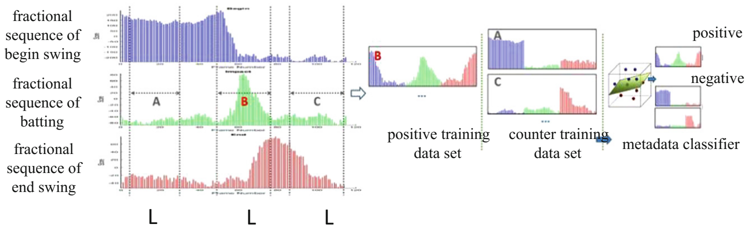

Fig. 4. Schematic diagram of training metadata classifier

2.2 Test Phase

In test phase, first the player detector locates the salience region of image while inputting the video sequence, and then scanning the video sequence to generate fraction sequence group with the key action classifier group which is trained in the previous phase, as shown in Fig. 5. Then, the metadata classifier is used to judge the fraction sequence obtained by the threshold sliding window, the length of L is got from most cases, sliding in a certain step on all fraction sequences to get the final judgment fraction sequence, as shown in Fig. 6.

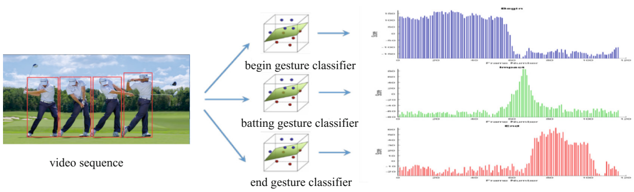

Fig. 5. The scanning video sequence graph by key gesture classifier

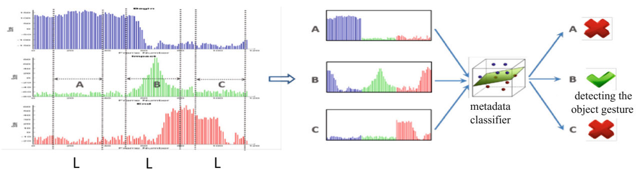

Fig. 6. The final classify result of metadata classifier

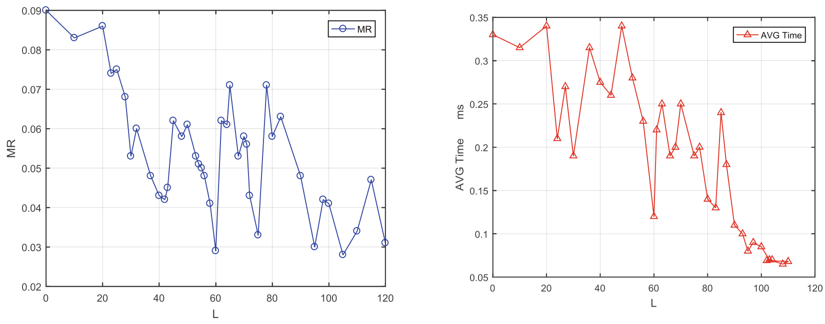

Fig. 7. The corresponding MR and average time with different L

3 Experimental Analysis

The algorithm operating environment: CPU i7, memory 8G, PC with Matlab2016b, and different iphones. The dataset uses UCF Sports and self-created dataset, all the resolution is normalized to 720×480. The experiments use 80% of samples as training data, 10% of samples as the validation set and the rest as test set.

In the aspect of golf gesture recognition based on video sequence, the length of sliding window is determined by missile rate (MR) and the average measuring time of each frame. The MR of algorithm with different window length L is shown in Fig. 7. The thesis takes L to 62, uses the precision and recall as performance measurement, and the cross validation method as an assessment method, the recognition frame rate on all iphones is 30 fps. Table 1 shows the experimental results in the motion dataset.

Table 1. The experiment results on UCF dataset

Data type on data set	Precision rate %	Recall rate %	The average time of every frame (ms)
Golf front swing	96.82	93.27	28.94
Golf side swing	97.37	94.87	28.74
Baseball side swing	96.88	83.78	29.12

4 Conclusion

How to reduce the time complexity of target recognition algorithm is an important issue in practical application. The thesis presents a fast golf gesture recognition algorithm based on video sequence. In different phase, player detector, key gesture classifiers and metadata classifier is used to deal with the image in video sequence. The experimental results show that the algorithm runs on iphone5s and later version, the recognition speed is more than 30 fps and the recognition accuracy is 97%, which proves the algorithm validity in practical application.

Acknowledgments. This work is supported by Research on Compiling Technology Based on FPGA Reconfigurable Hybrid System (No. 61003036).

References

1. Dollar, P., Tu, Z., Ponce, P., et al.: Integral channel features. In: BMVC (2009)
2. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR, vol. 6, no. 2, pp. 886–893 (2005)
4. Dollar, P., et al.: Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1532–1545 (2014)
5. Wohler, C., Anlauf, J.: An adaptable time-delay neural-networks algorithm for image sequence analysis. IEEE Trans. Neural Netw. **10**(6), 1531–1536 (1999)