ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 3065 – 3069

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

A Vision Based Method to Distinguish and Recognize Static and Dynamic Gesture

Zong-yuan Zhao[a], Wan-lin Gao[a]*, Miao-miao Zhu[a], Lina Yu[a]

[a]*China Arguiculture University,17[th] Qinghua East Road, Beijing100083,China*

Abstract

Vision based gesture recognizing acquired large improvements these years, as bonds of research achievements in labs become practical in factories. However, most of these findings are still restricted to either dynamic gesture or static posture, while the combination of them is always neglected but turns out to be common in practical. In this paper, we consider a vision based system that can interpret both dynamic and static gestures in real-time. Haar-like features and ada boost classifier are used to identify hands in images. Test results showed that when given suitable settings, the method of distinction runs successfully.

Keywords: gesture recognize, gesture distinction, ada boost classifier, haar-like feature

1. Introduction

As a kind of cutting aged technology, computer vision is wildly used not only in labs, but in the daily lives. Entertaining facilities are no longer limited in controlled by buttons and mouse, but seem to be more humane and amusing when users can handle them by gesture. Therefore, interpret users' gesture in real-time and complex background based on computer vision is a pivotal research area.

In this paper, a vision-based gesture recognizing system will be described, which can accurately distinguish and identify both dynamic gestures and static ones from sequences of images with complex background. Currently, we have limited our attention to single hand and we only use two dimensional models.

2. Backgrounds and related works

* Wan-lin Gao. Tel.:010-62738536.
E-mail address: gaowlin @cau.edu.cn.

doi:10.1016/j.proeng.2012.01.441

Gesture is a movement of part of our body, especially hands or head, to show what we mean or how we feel. It includes static gesture and dynamic gesture.

The origin of gesture recognizing is from 1983, when Grimes at AT&T got the patent of data glove and use it as sensor providing information of hands [1]. Vision based gesture recognizing is started in the 90s, at that time the labs in Fujitsu successfully identified 64 hand gesture [2]. In 1992, Fukunmoto, Mase and Suenaga developed a system called Finger-pointer, which can identify the points of fingers without gloves or signs [3]. Recently, more attentions are paid to bare hand recognizing, as Perry A. Stoll and Jun Ohya firstly used HMMs in the modeling and training of gestures and give out some method of recognizing [4].

For the complex background and real-time detection, Papegeorgiou, in 1998, came up with the concept of Haar Feature [5] which was followed by Haar-like feature, raised by Rainer Lienhart in 2002 [6]. In 1997, Freund and Schapire invited Ada boost, which is arithmetic of boost machine learning [7].

3. Dynamic and Static gesture recognizing

3.1. Static gesture recognizing

Static gesture recognizing aims at identify particular kinds of posture which remain still for a period of time in the videos. We use adaboost algorithm when training cascade classifiers, which can promote systems to be robust and real-time and can be used to identify static gestures.

3.1.1. Haar feature

Haar features are divided into three classes: edge features, line features and center-surround features.

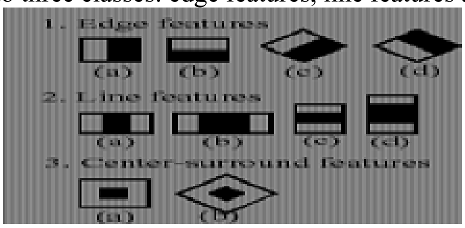

Fig. 1. Haar features

These features compose feature patterns, in which there are only rectangles in black and white. The feature of the patterns is defined as the deduction of the sum of pixel in white rectangles and black rectangles. The amount of haar-like features is depended on the size of the rectangles of the training sample images.

3.1.2. Cascade classifier training

Cascade adaboost classifier is a strong classifier which is made up of several cascade week classifiers. Week classifiers are the classifiers which are less accurate and single week classifier cannot classify objects accurately. But when they are combined to a cascade classifier, they can detect objects accurately, which is called a strong classifier.

The adaboost algorithm is a method to get a strong cascade classifier by training large amount of samples. The key of this iterative algorithm is to get a strong classifier by assembling week classifiers which are got from the same sample pool. The adaboost is realized by changing the distribution of data. It computes the weight of every sample according to whether it can be correctly classified and the precision of the last classification. The data with new weights will be delivered to the lower level to train the classifier, and the final cascade will be given after combining all these week classifier.

3.2. dynamic gesture recognizing

Dynamic gesture recognizing is mainly focusing on recording the position of hands in images and identifies the meaning of actions by analyzing them.

The first step is to identify hand automatically, that is to say, the system itself can get the position of the hand in images without manual intervention. We utilize the method used in static gesture recognizing to judge if any hands is in images. In our system, gestures cannot be judged at the time they are detected, since the dynamic gestures should be taken into consideration. Namely, we do not judge gesture only by one image but just record the position of hand because it may be one fragment in a series of images which contents a dynamic gesture such as a wave.

We record the position of hand until it cannot be detected or the number of frames which contain hand equals the set value. Then gesture can be identified by analyzing the recorded position which can judge whether it is a dynamic or static gesture and both of them can be recognized respectively.

4. Gesture distinguish

Gesture discrimination aims at distinguishing static and dynamic gestures automatically and recognizing them respectively. The workflow of our system is illustrated in the following picture. At first the system initializes itself as defining an empty queue to save hands' positions. After that, the main detect circle begins with capture an image from camera and does some preprocessing including converting forms, resizing and so on. If hand is detected in the image, then there are three conditions:

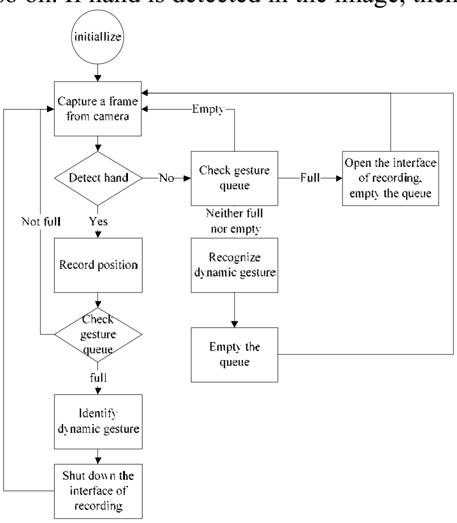

Fig. 2. The flowchart of the system.

1. The gesture queue is empty. In this situation no hand did any activity before this frame so no gesture shall be identified and should go straight to the beginning of the detect circle.

2. The gesture queue is neither empty nor full. In this situation hand must be just leave the camera detect area. That is to say, user just finished a dynamic gesture. The reason for this judgment is that hand

is detected in images before but get out of the detect area quickly. That means it is a typical dynamic gesture. Then positions that recorded before are utilized to identify the meaning of this gesture.

3. The gesture queue is full. In this situation a static gesture must had been finished for our system had detected hand in continuous frames for a long time, which is longer than any dynamic gestures' need. Thus, we had already identified this static gesture when it is finished and do not need any recognizing.

On the other hand, if hand is detected in image, position of it should be recorded in the gesture queue.

1. The gesture queue is full. In this situation, a static gesture is just finished for this frame is the last one in the gesture queue and dynamic gesture should be finished with less time. So we should identify the gesture using classifiers to get the meaning of it.

2. The gesture queue is not full. In this situation we could not judge whether it is a dynamic or static one, so no judgment should be made. We need to record the position and jump to the beginning of the detect circle.

5. Result

In this section, we made experiments about the method and theory above and results will be listed as follows.

5.1. Classifier training results

We have 487 sample images like the following images and 1000 negative background images. Positive samples are integrated to 20*20 pixels and BMP form. The region of hand takes more than 70% of every sample image and enjoys different and complex background and light.

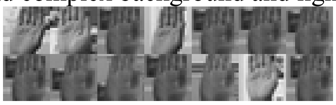

Fig. 3. Part of gesture samples

The number of cascade is set to 16 by experiments, as after training to the 16 stage the falsealarm reaches below $5*10^{-6}$. The order used in training is as follows:

-data samples -vec samples.vec -bg bg.dat -nstages 16 -nsplits 2 -minhitrate 0.999 -maxfalsealarm 0.5 -npos 487 -nneg 1000 -w 20 -h 20 –nonsym -mem 256 -mode ALL -minpos 200

After training we got a classifier which recognition rate is beyond 80%. However, when we detected hands in camera videos it seemed to enjoy a satisfied results and can avoid complex backgrounds and varied lights. The average time for recognizing is 0.06s.

5.2. Gesture distinction results

In this section we are just focusing on whether dynamic and static gestures can be accurately distinguished and the rate of identifying them is currently not taken into consideration.

Experiments suggest that people use no more than 1.5s to finish a dynamic gesture. During that period, our system can detect hand in 20 images, which is enough to recognize dynamic gesture.

Queue length is the length of the hand position queue. Dynamic rate means the percentage of recognized dynamic gestures. Dynamic error rate refers to the proportion of those dynamic gestures that are recognized as static ones. Our system used 0.7s in average to process one frame.

Table 1. Test results

Queue length	Dynamic rate	Dynamic error rate	Static rate	Static error rate
10	1/10	5/10	6/10	0/10
15	4/10	2/10	8/10	1/10
20	8/10	0/10	8/10	0/10
25	9/10	1/10	9/10	1/10
30	8/10	0/10	6/10	3/10

5.3. Discussion

We can see from the results that when given less queue length, dynamic recognition rate goes down and the error rate turns high. It is because hand moving too fast camera could not capture suitable images, while moving too slow the time to finish a dynamic gesture would be longer than the set time and be identified as static gesture. However, if given longer queue length, we got higher recognizing rate for dynamic gesture but lower for static. The reason lies on that people usually do not use that much time to finish a static gesture so some of them may be treated as dynamic ones. By analyzing and comparing the results, we set the queue length to 30 which can not only promise the recognizing rate but also be convenient and humanity.

6. Conclusion

In this paper we considered a vision based real-time system which can distinguish and recognize dynamic and static gestures automatically. We use haar-like feature and cascade ada boost classifiers to promise hand can be detected in complex background. The efficiency and accuracy of recognizing should be improved by refining classifier and training methodology and these are our work in the future.

References

[1] Ren Haibing, Zhu Yuanxin, Xu Guangyou, et al .Vision-based recognition of hand gestures: a survey. *Acta Electronica Sinica,* 2000, 28 (2): 118-121

[2] T. Takahashi and F. K. Shino. Hand gesture coding based on experiments using a hand gesture interface device. *SIGCHI Bulletin*，1991，23(2):67-73

[3] FUKUMOTO, M., MASE, K., and SUENAGA, Y.: ‘Real-time detection of pointing actions for a glove-free interface’, in *IAPR workshop on machine vision applications,* 1992, pp. 473-476

[4] Stoll, P.A. and J. Ohya. Applications of HMM modeling to recognizing human gestures in image sequences for a man-machine interface. in Robot and Human Communication, 1995. RO-MAN'95 TOKYO, Proceedings. *4th IEEE International Workshop* on. 1995.

[5] Papageorgiou, C.P., M. Oren, and T. Poggio. A general framework for object detection. in *Computer Vision*, 1998. Sixth International Conference on. 1998

[6] LIENHART R，MAYDT J．An extended set of haar-like features for rapid object detection．*IEEE ICIP2002*，2002，1:900-903

[7] Viola P.，&Jones，M..Robust real time object detection.*8th IEEE International Conference on Computer Vision Vancouver,* 2001

[8] Jianing Zhao , Wanlin Gao , Zili Liu , J: A Classification of Remote Sensing Image Based on Improved Compound Kernels of svm. *Computer and Computing Technologies in Agriculture,*.4Volume 317/2010,p15-20(2010)