Human Eye Recognition Tracking Based on Region Segmentation and Template Matching

Fei Li, Huadong Zheng*, Jia Huang, Jianbin Hu, Zhifang Gao

Department of Precision Mechanical Engineering, Shanghai University, Shanghai 200444, China

*Corresponding author: bluenote2008@shu.edu.cn

Abstract. As a very challenging field, computer vision has always attracted the in-depth exploration and research of many researchers because of its huge development potential. With the continuous improvement of computer performance and the popularity of electronic products, more and more researchers are committed to the important research direction in the field of computer vision - human eye detection. In this paper, the eye in the video sequence is taken as the research object, and the tracking and recognition algorithm of the dynamic object is taken as the main research purpose. The motion tracking of the human eye is deeply studied. Aiming at the existing problems of human eye tracking and recognition algorithms, this paper proposes a human eye tracking and recognition algorithm based on combination of region segmentation and template matching. Experiments show that this method greatly improves the accuracy and stability of human eye recognition.

1. Introduction

As an important feature of the human face, the eye plays an important role in face detection and recognition. The detection of the human eye mainly includes face detection and human eye detection. This is because in order to improve the detection speed, we use the process of detecting the human face first and then detecting the human eye on the basis of the face, which is consistent with the top-down. The visual recognition framework, because when we look at people with the naked eye, we often locate the face first and then recognize the position of the human eye. Since the pixels of the eye are often few, if one window is scanned, the detection speed is very slow and it is easy to misjudge. Adding an early face detection can avoid this problem.

For the face detection method based on the OpenCV database provided by Intel Corporation, the Adaboost face detection algorithm with high precision and fast features is recently proposed. The algorithm extracts the Harr-like feature of the image, which contains the frequency information of the image that can only be obtained by integer calculation, and then selects a series of important features from the extracted features. After classifying these important data, this series is important. The features can be combined into a cascade of classifiers that are very comprehensive and can detect a variety of faces in different lighting and different color spaces. For human eye detection, the OpenCV database has an Adaboost human eye detection algorithm for the human eye. The commonly used detection sequence is to detect the face first and then perform the human eye detection on the detected face.
In view of the above traditional human eye detection methods, the results have the disadvantages of unstable human eye detection and low efficiency. This paper proposes a traditional human eye detection method, which uses region segmentation and template matching in the process to greatly improve the detection results. Area segmentation refers to the elaboration of the area detected by the human eye, finding the smallest area where the human eye is located, and segmenting other external areas, so that the search range for the human eye is greatly reduced, and the search time is reduced. The human eye matching is to detect the human eye in the first few frames of the image, and then use it as a template, and then use the template search matching method instead of the human eye detection method, which reduces the calculation amount and improves the detection efficiency.

2. Algorithm and experimental result flow

2.1. Adaboost face detection algorithm
The face detection algorithm based on the rectangular feature proposed by Paul viola realizes the problem of rapid face detection. The idea of using the rectangular feature to describe the face feature algorithm is simple and effective. The integrated image is used to calculate the rectangular feature, which greatly improves the calculation speed of the rectangular feature and reduces the computational complexity.

In the classification, Paul Viola uses the AdaBoost algorithm to train the face classifier. Because the number of rectangular features is very large, the simple Adaboost classifier requires a large number of features to correctly classify the face, which affects the speed and the number of uses is huge. The feature training classifier wastes resources and time, so the training can be pre-selected with Adaboost, and Adaboost is used as a feature screening algorithm to find some features that are easy to distinguish faces from a large number of rectangular features. This part of the feature is further trained as a basic rectangular feature. Viola uses the Adaboost algorithm to train the classifier and proposes the idea of a cascade classifier. The Adaboost cascaded classifier algorithm can achieve the purpose of fast detection. The basic idea of the cascade classifier is that the first few layers train some simple classifiers. After these simple classifiers, more than half of the non-face windows can be eliminated. The burden of the later layers of complex classifiers, the more complex classifiers will only be further judged for those areas that are more like faces.

2.2. Adaboost cascade classifier algorithm
The Adaboost algorithm was proposed by Reund and Seapire in 1995. Adaboost is called Adaptive Boosting. The Adaboost algorithm does not require any prior knowledge about the performance of the weak learner. Plus it is as efficient as the original Boosting algorithm, so it can be very easily apply to actual problems. The Adaboost algorithm is composed of multiple weak classifiers to form a strong classifier. The learning rules of each weak classifier are very simple. After selecting a weak classifier, the weights are re-assigned to the samples. The samples of the last weak classifier are added with weights, and then the next weak class is selected. All weak classifiers are combined to form a strong classifier that is classified by threshold. The strong classifier thus achieved is better than any weak classifier.

The Adaboost algorithm can effectively select the most classifiers from many weak classifiers into strong classifiers, and experiments prove that this final strong classifier has a good detection rate. However, the number of classification algorithms based on rectangular features is huge, and more features are needed to achieve classification. In the classifier, the excessive number of weak classifications will make the generalization ability of the classifier weaker and affect the experimental results, so it is introduced. The concept of a cascaded classifier. The cascade classifier is composed of multiple Adaboost strong classifiers. Each layer classifier is a small Adaboost strong classifier. The requirement of the small classifier is that the classification requirements are lower in the first layer, along with the number of layers. The increase of the classification condition of the classifier is stronger, that is, the classification is closer to the accuracy. The simpler classifiers of the first few
layers can exclude most of the non-target windows, so that later complex classifiers can achieve lower false detection rates by performing operations on sub-windows that are more difficult to distinguish. Since the difficult-to-distinguish window requires a lot of features to be classified, the cascading method can be used to classify more detailed classifications step by step, so that the complex classifier only pays attention to detail classification, ignoring the characteristics of the rough classification, thus making the detection speed is relatively increased.

The detection process of the cascade classifier is similar to the decision tree. The first layer classifier sends the determined target result to the second layer classifier to continue the decision, and then the second layer classifier sends the decision result to the third layer. So on and so forth. The non-target results judged by any one classifier are directly excluded and no further judgment is made. So the strong classifiers of the current layer are faced with more difficult classification tasks than the previous one, because the samples of all the previous layers can be more difficult to distinguish from the normal samples. The structure of the cascade classifier is shown in Figure 2.6. It can be seen that only the child windows that are layered through all layers of the classifier are considered to be targets. Simple non-target windows are filtered out in the first few layers, and the fewer layers to be detected in the next few layers, which allows the latter classifier to focus on the discrimination of complex windows, reducing the burden on complex classifiers. Improve detection accuracy and speed.

Experimental results:  

**Fig. 1 Experimental results-face**

2.3. Human eye detection

Applying basic rectangular features for face detection, a good detection result is obtained. Similarly, since the eye also has many common features, the rectangular feature can also represent the characteristics of the eye. Since the eye has its own characteristics compared to the face, the features for face detection are not completely suitable for eye detection. Based on the preservation of several primitive rectangular features, several extended rectangular features and some new features have been added to suit the characteristics of the eye.
2.4. Regional segmentation

In the obtained face image, template matching is directly performed to find the eye. There is always a problem of computing speed. In a system with high real-time requirements, if the template matching method is directly used to find the target, the tracking efficiency of the algorithm is not very ideal. Therefore, before performing the pupil positioning of the human eye, we use the prior knowledge of the face geometry to coarsely locate the general position of the eye and reduce the moving area of the template matching method, thereby saving a certain computation time.

For the proportion structure of the human head, the ancients have the saying of "three courts and five eyes", which clarifies the basic proportion features of the facial structure. From the length of the hairline to the lower jaw, the length of the face is divided into three equal parts, which is the so-called three courts, where the court is from the hairline to the eyebrows; the second court is the eyebrow to the nose, and the third court is It is the area from the tip of the nose to the lower jaw; the five eyes refer to the width of the face: the boundary between the left and right hairs, the length between the two is exactly the length of five eyes, and the distance between the two inner corners is one eye. The length is one eye length on each side. Figure 3-10 is a schematic diagram of "three courts and five eyes".

Fig. 2 Experimental results-eye

Fig. 3 Schematic diagram of regional segmentation
Using the cascade classifier we first detect the face and get a rectangular area. If we search for the human eye in this rectangular area as usual, the above test results are unstable. In order to solve this problem, this paper proposes a region segmentation method, which divides the rectangular region and finds the minimum range of the eye's position. Searching for an eye within this range will reduce the amount of calculation and improve the stability of the test results.

The position of the eye is in the upper part of the rectangle, so we can intercept the upper part of the rectangle, as shown by the green line in the above picture. The first step is to divide the rectangle from the top and bottom into two. In the second step, we can round off the top 1/4 area of the upper part, because this area is generally the forehead and the eyes will not appear in this area. The third step divides the rectangular box where the eyes are located into two, and the left eye and the right eye are in the left and right boxes. At this point, we can subtract 1/4 of the area from the edge of the left and right frames, and finally in the rectangle where A and B are located. The eye detection area is reduced to the small rectangle where A and B are located, which greatly reduces the calculation amount and improves the detection quality.

![Fig. 4 Experimental results are unstable](image)

After the segmentation is performed, the speed of the detection result is increased, and there is no case where only one eye is recognized. However, there will still be instability, and occasional eye detection jumps will be lost.

2.5. Matching Tracking
In view of the above phenomenon of data hopping loss of eye detection, we use template matching to perform eye tracking. The object is detected by the method of human eye detection. If each frame is detected, it is very time consuming, and occasionally the phenomenon that the detection result jumps is lost. We turn the high calculation rate to a low calculation rate, and perform a test through the first few frames to use the detected human eye as a template. Then use template matching to track, that is, we first detect the eye, save it as a template, and then use the template matching method to perform human eye tracking in the last small rectangular area of the segmentation without using the human eye detection.

Opencv implements a part of the match Template () method for finding the best match between the template and the target image. The basic idea is to scan the template image one by one on the target image and match it by statistical basic methods, such as variance test, correlation test, etc. The method to find the best match.
The results show that the human eye recognition tracking effect is stable.

3. Conclusion
For the traditional cascaded classifier method, the detection results of face and human eye recognition are unstable and the disadvantages under the efficiency. This paper proposes a method based on region segmentation and template matching for human eye recognition, which effectively improves the detection. The efficiency and stability of the results.

References
[1] Horn B K, Schunck B G. Determine Optical Flow. Artificial Intelligence, 1981, 17(1-3):185-204.
[2] Zhu Kezhong, Research on a Fast Algorithm for Target Detection and Tracking, Technology information [J] 2007.7, 254-255.
[3] Forsyth, D.A. and Ponce, J. Computer vision: A Modern Approach [M] Prentice Hall Professional Technical Reference, 2002.
[4] Guo Baolong, Sun Wei Introduction to Digital Image Processing System Engineering. Xi'an University of Electronic Science and Technology Press, 2012.7, 67-69.
[5] Zhong Yu, jain,Anilk, Dubuisson-Jolly, Marie-Pierre. Object tracking using deformable templates. Proceedings of the IEEE international conference on Computer Vision. 1998.1, 440-445.