

Detection of the Eye Iris Using Regression Trees

Maxim S. Beloyvanov¹, Alexander V. Fisunov¹, Iakov S. Korovin² and Donat Ya. Ivanov^{2,*}

¹Don State Technical University, Shakhty, Russia

²Southern Federal University, Taganrog, Russia

*Corresponding author

Keywords: Eye tracking, Regression trees, Eye iris tracking.

Abstract. Recently, various systems of tracking the direction of a person's gaze have been of great interest. As a rule, these systems have a complex structure, and also use additional sources of illumination. This article discusses eye tracking methods, and proposes a method for tracking and determining the shape of the iris.

Introduction

Tracking the direction of a person's gaze is an important technology for creating various systems, such as interaction and non-contact control of digital equipment, tracking of human behavior, etc. There are devices [1] capable of performing this task, but they use additional sources of illumination, as well as IR illumination and specialized cameras operating in the IR range. The solution to this problem will simplify and cheapen such systems, as well as allow the use of ordinary webcams for personal computers and built-in cameras of mobile devices.

The Statement of the Problem

In [2], the captured from the webcam image is converted to grayscale to reduce the complexity of the calculations. The user's face is detected using Haar cascade [3], and the eye region is highlighted in the resulting region. The image from the eye area is cut out and undergoes preprocessing, after which it is transmitted to the input of the neural network. An example of preprocessing the image of the eye is shown in Figure 1.

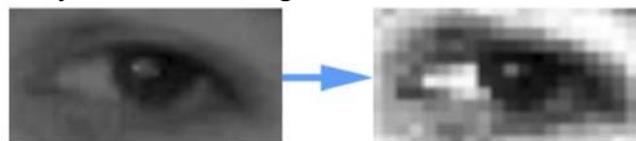


Figure 1. An example of image preprocessing.

Methods of eye detection, based on the full eyes image analysis, are taken as input for the image of the eyes, and then create functions for determining the direction of the eye. Lou et al. [4] proposed an adaptive linear regression method for sparse selection of training samples. The images of the eyes are divided into many sub-regions, and the total pixel intensity of each sub-region is used to summarize the feature vector for comparison with points on the screen. Zhang et al. [5] used multimodal convolutional neural networks to determine the direction of gaze. For training, large data is collected for training, obtained from 15 participants who have used a laptop for three months. Funese et al. [6] correlated the three-dimensional morphized model with the depth data obtained from the depth sensor, then the images of the eyes are cut off along the borders of the face to determine the look. The final result is determined by a combination of the frontal view and the position of the head. Sugano et al. [7] proposed a multi-chamber system in which the three-dimensional shape of the eye region is reconstructed and random forest is used to evaluate the gaze.

In contrast to methods based on features analysis, methods based on features evaluate the view by highlighting specific points of the face, such as the corners of the eyes, eyelids, centers of pupils or corneal highlights.

The Obtained Results

The purpose of this article is to find the way to track the position of the centers of the eyes, as well as determine the radius of the visible part of the iris. To achieve this goal, it was decided to apply the method based on an ensemble of regression trees proposed by Vahid Kazemi et. al. [8], for an accurate assessment of the position of personal landmarks in terms of computational efficiency. Each r_i regressor cascade predicts the update vector from the image and the current estimate - $S^{(t)}$ on each image I , which is added to the current estimate of the form $S^{(t)}$, to improve the estimate:

$$S^{(t+1)} = S^{(t)} + r_i(I, S^{(t)}) \quad (1)$$

Regressor r_i makes its predictions based on features such as pixel intensity values calculated from the image and indexed relative to the current shape estimate $S^{(t)}$. This introduces some form of geometric invariance into the process, and as you progress through the cascade, you can be more confident that the exact semantic location on the face is indexed. The output signal range, extended by the semantic emblem, is guaranteed to lie in the linear subspace of the training data, if the initial estimate $S^{(0)}$ belongs to this space. The initial shape can simply be selected as the average form of the training data, centered and scaled according to the output of the detector's bounding box. To train each r_i the gradient tree boosting algorithm with a sum of square error loss is used as described in, as described in [9]. Since the described method is essentially a form detector, it can be used to determine the shape of the iris and localize its center.

To train the model, it is necessary to mark up the set of images on which 7 points are annotated, six of which are marked along the contour of the iris and one in the center. The rectangle highlights the area that the model uses to find points. The training sample contains 1000 images of eyes previously transformed into images in grayscale, and a sample of 200 annotated images that are not involved in training is used to check the accuracy of work. Figure 2 shows an example of a tagged image from a training sample.



Figure 2. An example of eye image markup.

Since the training sample only serves to verify the accuracy of the work, the process of learning the model is governed by a set of parameters that affect the size, accuracy and speed of the generated model. The following parameters were chosen for the training.

The depth of the trees used in each cascade, and defining the capacity of the model, was chosen to be 4, which gave a compromise between the accuracy and size of the model.

The regularization parameter was chosen equal to 0.1, which will force the model to recognize patterns rather than fixed data, thereby eliminating the problem of retraining, and will allow training the model on a relatively small number of images.

The cascade depth is 10, which gives a balance of accuracy and size of the trained model. The size of the object pool is chosen to be 400 to ensure accuracy and good execution speed. Increasing this parameter will lead to improved accuracy, but the speed of work will begin to decline significantly. The number of separation elements selected at each node is 20. This parameter is responsible for selecting the best functions at each stage in the learning process. The parameter

affects the speed of training and the accuracy of the model; its further increase will result in an increase in training time. The number of randomly selected deformations applied to the training samples was chosen to be 40, which significantly increases the training time, however, it allows the use of small data sets.

Figure 3 shows the result of the work of the trained model, which was tested on images that were not involved in the learning process.



Figure 3. Iris detection result on test images set.

The trained model was tested on video from a webcam with a resolution of 1280x720 and a frequency of 30 frames per second. The eye areas were pre-detected using the HAAR classifier, the coordinates of which were used by the model to detect the iris shape. Figure 4 shows the result of the detector usage.



Figure 4. The result of the real time iris detection.

Conclusion

As a result of the work, an iris shape detector was obtained. The detector has high performance. On a system with an i5-5200u processor, the detection of points is performed in less than 2 ms for HD input image. This approach can be used on mobile devices and reduce the load on the computing device.

Acknowledgment

The reported study was funded by RFBR according to the research projects № 18-29-03225.

References

- [1] Eye Trackers for PC Gaming <https://gaming.tobii.com/products/> (2016)
- [2] Real-Time Eye Gaze Tracking With an Unmodified Commodity Webcam Employing a Neural Network

- [3] Rapid Object Detection using a Boosted Cascade of Simple Features. Viola, Paul and Jones, Michael. s.l.: IEEE, 2001. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. pp. I-511- I-518 vol.1
- [4] Bäck, David. Neural Network Gaze Tracking using Web Camera. Välkommen till institutionen för medicinsk teknik : Masters Thesis, 2005.
- [5] Non-Intrusive Gaze Tracking Using Artificial Neural Networks. Baluja, Shumeet and Pomerleau, Dean. 1994, Technical Report: CS-94-102.
- [6] Task-Dependent Evolution of Modularity in Neural Networks. Igel, Christian, Toussaint, Marc and Hüsken, Michael. 2002, Connection Science.
- [7] A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. Riedmiller, Martin and Braun, Heinrich. 1993. IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS.
- [8] One Millisecond Face Alignment with an Ensemble of Regression Trees
- [9] T.Hastie, R.Tibshirani, J.H.Friedman. The elements of statistical learning: data mining, inference, and prediction. New York: Springer-Verlag,2001. 2,3