An Accurate Eye Pupil Localization Approach Based on Adaptive Gradient Boosting Decision Tree

Dong Tian *, Guanghui He *, Jiaxiang Wu *, Hongtao Chen #, Yong Jiang #

*School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University
Dongchuan Road 800, Minhang District, Shanghai, 200240, China
#Electric Power Research Institute, State Grid Shanghai Electric Power Company
Shanghai, 200093, China

*guanghui.he@sjtu.edu.cn

***Abstract*—Eye pupil localization is an important part in computer vision applications such as face recognition, gaze estimation and so on. In this paper, we propose an improved method for precise and fast eye pupil localization. Based on gradient boosting decision tree(GBDT) algorithm, a more accurate localization is achieved by increasing the weight of the training samples with larger errors in a moderate rate. Furthermore, a pruning strategy is utilized to avoid overfitting and reduce the localization time without accuracy loss. Experimental results show that the improved method achieves an accuracy of 92.39% at a speed as fast as 1.7ms to locate in the range of eye pupil on BioID database. The proposed method outperforms most state-of-the-art methods in terms of localization accuracy and consumed time.**

***Index Terms*—computer vision, eye pupil localization, gradient boosting, adaptive, pruning method**

I. Introduction

The human eyes play an essential role in routine activities. Thus, eye pupil localization is widely used in human-computer interaction, gaze estimation and so on. Although valuable progresses have been made over the years, there still exists some challenges owing to facial expression, head pose, eye status, light conditions, variation in scale and ambient occlusion. To address these problems, numerous approaches have been proposed, which can be roughly divided into two categories, *i.e.*, feature-based methods and learning-based methods.

Feature-based methods use edges, eye corners and special points to detect eye centers. Curvature of isophotes [1] and geometric information [2], [3] are employed to detect eye pupil coordinate. Although feature-based methods are robust to shape and scale variations, the methods may fail to locate the eye pupil when extracted features are invalid in some circumstance, for example, when feature detection is confused by noise, or when eyes are almost close. Besides, these methods consume an amount of time to extract features.

More recent methods are implemented by machine learning algorithms. These methods often adopt Support Vector Machine and boosting algorithm to estimate the eye pupil location [4],[5]. Markus et al.[6] propose a framework based on an ensemble of randomized trees including GBDT. However, gradient boosting algorithm is likely to overfit the training set when the convergence rate is close to 1. The training time would increase significantly if the convergence rate descends. Besides, the methods may fail to estimate the eye pupil coordinates when requiring precise localization.

In this paper, we propose an improved approach based on gradient boosting decision tree model to locate eye pupil position accurately and efficiently. Inspired by Adaboost algorithm in the training process, the samples with larger errors will be emphasized for the next iteration in a moderate degree in case of overfitting. In addition, a pruning strategy is utilized to prevent the model from overfitting and reduce the time cost in the predicting process. Experimental results demonstrate that the proposed method outperforms most state-of-the-art methods when it comes to accuracy and time cost.

II. BASIC ALGORITHM

Gradient boosting algorithm is a machine learning technique for regression and classification problems. This model consists of additive regression trees by sequentially fitting a simple parameterized function to current residuals at each iteration.

Every single tree is constructed according to the difference of pixel intensity between two given points and a learned threshold. Binary tests can be utilized to split the original problem into two simpler ones recursively in each node of regression tree. A binary test[6] on image $\boldsymbol{I}$ is defined as

$$bintest(\boldsymbol{I}; I_1, I_2, t) = \begin{cases} 0 & \boldsymbol{I}(I_1) - \boldsymbol{I}(I_2) \le t \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

where $\boldsymbol{I}(I_i)$ indicates the pixel intensity at location I_i and t is the threshold. Location I_1 and I_2 is the normalized coordinates.

The training procedure follows the greedy algorithm at each node. Starting from the root node again, according to the level-order traversal, the clustering quality is computed for every

978-1-5090-5316-2/16/$31.00 © 2016 IEEE VCIP 2016, Nov. 27 – 30, 2016, Chengdu, China

combination of features and threshold at each non-leaf node.

$$Q_{node} = \sum_{\boldsymbol{p} \in S_l} \left\| \boldsymbol{p} - \frac{\sum_{\boldsymbol{q} \in S_l} \boldsymbol{q}}{\|S_l\|} \right\|^2 + \sum_{\boldsymbol{p} \in S_r} \left\| \boldsymbol{p} - \frac{\sum_{\boldsymbol{q} \in S_r} \boldsymbol{q}}{\|S_r\|} \right\|^2 \quad (2)$$

where the $i = 1, 2, \ldots, 2^d$ is one of 2^d classes, and the S_r and S_l are clusters that contain pupil coordinates of all eyes patches for which the outputs of binary test were 0 and 1, respectively.

In training process, the parameters of each binary tests in internal nodes are optimized by maximizing the clustering quality. In this case, the optimization problem can be transformed to the problem of minimizing the node quality Q_{node}.

The combination of features and threshold that achieves the best clustering quality is selected as the binary test parameter, and labeled at this node. When reaching the terminating condition, the output at this leaf node is formulated by averaging all the remaining samples annotated prediction result.

Every regression tree trained from a vast amount of training samples is able to predict the eye pupil coordinate given an image of the face region. Given that each of these predictors is relatively weak, thus gradient boosting algorithm could be used to combine these weak predictors together to be a strong predictor through iterations.

III. PROPOSED ALGORITHM

In this section, an adaptive gradient boosting decision tree is presented to improve the accuracy of localization. Moreover, a pruning strategy is utilized to avoid overfitting and reduce consumed time.

A. Adaptive GBDT

The gradient boosting aims to minimize the loss function $L(\boldsymbol{y}, F(\boldsymbol{x}))$ through iterations, where $F(\boldsymbol{x})$ denotes the learned decision tree model and $\boldsymbol{y}$ is the coordinate of the training set. In every iteration, the loss function descends towards its negative gradient direction, so that the loss function can be minimized quickly. When the square error, defined as $L(\boldsymbol{y}, F(\boldsymbol{x})) = \frac{1}{2}(\boldsymbol{y} - F(\boldsymbol{x}))^2$, is used in loss function, the gradient $\widetilde{\boldsymbol{y}}$ is the residue itself. In another word, the decision trees fit the residue of all previous stages in every iteration.

$$\begin{aligned} \widetilde{\boldsymbol{y}} &= \frac{\partial L(\boldsymbol{y}, F(\boldsymbol{x}))}{\partial F(\boldsymbol{x})} = \frac{1}{2} \frac{\partial (\boldsymbol{y} - F(\boldsymbol{x}))^2}{\partial F(\boldsymbol{x})} \\ &= \boldsymbol{y} - F(\boldsymbol{x}) \end{aligned} \quad (3)$$

Once the descent direction is settled, we should determine the proper descent stride ρ_m. ρ_m is selected by solving the following one-dimensional optimization problem.

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} \|\widetilde{\boldsymbol{y}}_i - \rho h(\boldsymbol{x}_i)\|^2 \quad (4)$$

where $h(\boldsymbol{x}_i)$ indicates that the fitting result of the residue.

Inspired by Adaboost algorithm, during the process of training weak classifiers, the weight of wrong samples will be increased for the next iteration owing to the wrong classification. Meanwhile, the weight of right samples will relatively decrease. So normalized mean squared error is considered as the weight of training samples.

$$w_i = \frac{\|y_i - F_{m-1}(x_i)\|^2}{\sum_{i=1}^{N} \|y_i - F_{m-1}(x_i)\|^2} \quad (5)$$

where $F_{m-1}(x_i)$ denotes the result of $m-1$ iteration corresponding to the i_{th} component.

Through the iterations, a more accurate classifier is modeled according to current weight distribution. The weighted GBDT method will help adjust the error of weak classifier adaptively and improve the accuracy and robustness.

In addition, learning rate ν is introduced to control the convergent rate of iteration in the algorithm. It has been verified that small learning rates yields dramatic improvements in accuracy. However, it comes at the expense of increasing computational time both during training and predicting because of more iterations.

Algorithm1 summarizes the algorithm flow of the training process.

Algorithm 1 Adaptive Gradient Boosting Decision Tree

Initialization: $F_0(\boldsymbol{x}) = \arg\min_{F(\boldsymbol{x})} \sum_{i=1}^{N} L(\boldsymbol{y}_i, F(\boldsymbol{x}_i))$

// Iterate M times
for $m = 1 \to M$ **do**
  //Determine the weight of training sample
  $w_i = \|\boldsymbol{y}_i - F_{m-1}(\boldsymbol{x}_i)\|^2 / (\sum_{i=1}^{N} \|\boldsymbol{y}_i - F_{m-1}(\boldsymbol{x}_i)\|^2)$
  //Compute the residual
  $\widetilde{\boldsymbol{y}}_i = \boldsymbol{y}_i - F_{m-1}(\boldsymbol{x}_i),\ i\ from\ 1\ to\ N$
  //Fit weighted residual
  $\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} w_i \|\widetilde{\boldsymbol{y}}_i - \rho h(\boldsymbol{x}_i)\|^2$
  //Update model for next iteration
  $F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \nu \rho_m h(\boldsymbol{x})$
end for
return $F_M(\boldsymbol{x})$

B. Pruning Strategy

Model for predicting the eye pupil is learned by fitting the previous training set. However, our goal is to build up a model capable of predicting test set. In the trained model, the eye pupil coordinate will be determined after 25 iterations. Yet some samples in test set may achieve the least error rate in less than 25 iterations. Thus, during the iterations, we achieve the regularization by introducing the pruning strategy that terminates the iterative process once the normalized error is less than threshold .

There exists similar pruning strategy in single tree locating process. When predicting the eye center of a given image through one single tree, the algorithm starts at the root node of the decision tree, at each non-leaf node, the process moves to the left or right child node according to the answer to the pre-trained binary test. Meanwhile, the normalized error is computed to decide whether to continue the process. The eye

Algorithm 2 Localization Based on Adaptive GBDT

Input: $TestImage : \boldsymbol{I}$
Output: $\boldsymbol{y} \in \Re^2$
for $layer = 1 \rightarrow depth$ **do**
  $node = 2 \cdot node + bintest(\boldsymbol{I}; I_1, I_2, t)$
  if $err <= T_{err}$ **then**
    $break$;
  end if
end for
return $\boldsymbol{y}_{node}$

pupil coordinate y_{node} will be returned once the normalized error is less than the pruning threshold T_{err} or the process reaches the leaf node. Algorithm2 summarizes the algorithm flow of locating process through a single tree.

Moreover, the localization time will be reduced because the process may stop within 25 iterations. The extent of pruning is controlled by tunable threshold.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we describe the implementation detail and the evaluation metric. Then experimental results are presented to demonstrate the performance of the improved eye pupil localization model in comparison to state-of-the-art methods in terms of accuracy and consumed time.

A. Implementation and Evaluation Metric

LFW[11] public data set is used to train the model. The LFW data set contains 13233 frontal faces.

Face images are obtained using face detector in OpenCV and eye regions are estimated by simple anthropometric relations. Adaptive GBDT method with multi-level eye detection schemes is applied to locate the eye pupil. What's more, random perturbation, which samples multiple rectangular patches around the eye, is added to improve the robustness of localization. During the training process, all face images are resized according to the bounding box width of the stage to ensure the same scale of face images. We employ an estimation chain consisting of 5 stages with 25 trees of depth equal to 11.

The normalized error, indicating the error obtained by the worst eye estimation, is adopted as the evaluation metric for the estimated eye pupil location. This metric is proposed by Jesorsky et al.[12] and is defined as

$$err = \frac{max\,(d_l, d_r)}{\|C_l - C_r\|} \tag{6}$$

where d_l and d_r are the euclidean distance between the found left and right eye centers and the ones in the ground truth, and $\|C_l - C_r\|$ is the euclidean distance between the eyes in the ground truth. In this metric, an error of err = 0.25 corresponds to the distance between two eye centers, err = 0.10 corresponds to the range of the iris, and err = 0.05 corresponds to the range of the pupil. When err $\leq$ 0.05, the localization can be considered to be accurate.

The BioID database(http://www.bioid.com) is used for test. It consists of 1521 near frontal face grayscale images with significant variation in illumination, scale and pose. In the database some eyes are closed, turned away from the camera, partly hidden by strong highlights on the glasses. Due to these conditions, the BioID database is considered difficult and realistic.

B. Effectiveness of Adaptive GBDT and Pruning Strategy

To verify the effectiveness of the improvements, the basic GBDT and adaptive GBDT methods have been trained with the same training set. We test all models on BioID database by comparing the accuracy and consumed time.

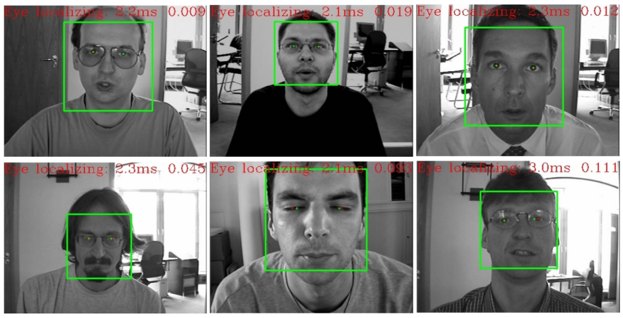

Fig. 1: Some examples of pupil localization on BioID database

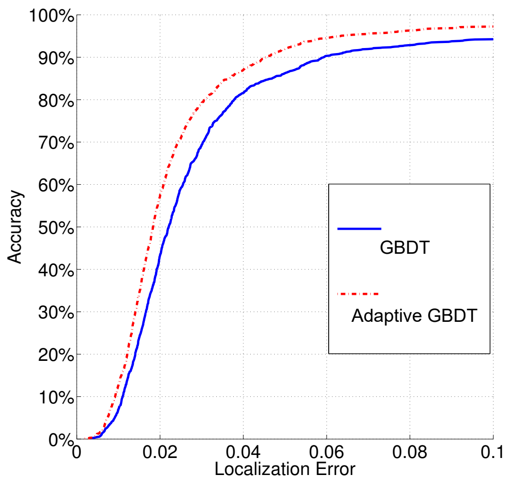

Fig. 2: GBDT vs. Adaptive GBDT($\nu = 0.2$)

Fig.1 shows some examples obtained on different subjects of BioID database by adaptive GBDT method, where green crosses are the found landmarks and red crosses are the annotated landmarks. The normalized error is in the top right corner. The first row provides exact localization samples normally. When presented with closed eyes and strong highlight on the glasses like the second row, the proposed method can still locate the eye center with larger normalized error.

Fig.2 compares the performance of the basic and adaptive GBDT method in localization accuracy. We can observe that

TABLE I: COMPARISON OF NORMALIZED ERROR SCORES AND CONSUMED TIME ON BIOID DATABASE

Method	$e \leq 0.025$	$e \leq 0.05$	$e \leq 0.1$	$e \leq 0.15$	$e \leq 0.2$	Consumed Time(ms)
Adaptive GBDT	**70.06**	**91.54**	**97.07**	**99.10**	**99.89**	**2.0**
AGBDT(T_{err}=0.025)	**71.17**	**93.93**	**98.22**	**99.36**	**99.90**	**2.0**
AGBDT(T_{err}=0.05)	**-**	**92.39**	**99.13**	**99.81**	**99.97**	**1.7**
Zhang et al.,2016[2]	*52.00	85.66	93.68	*95.00	*97.00	-
Soelistio et al.,2015[3]	*40.00	80.75	95.15	97.78	98.88	-
Valenti et al.,2014[7]	*35.00	71.30	95.80	98.60	99.30	-
Yi et al.,2011[8]	*54.00	81.50	99.00	*99.50	*99.80	-
Valenti et al.,2008[1]	*48.00	84.10	90.85	*92.00	96.50	33.0
Zhang et al.,2014[5]	*38.00	81.00	93.30	-	*98.00	11.3
Markus et al.,2014[6]	*61.00	89.90	97.10	98.00	98.90	**0.24**
Zhou et al.,2015[9]	*50.00	**93.80**	**99.80**	***99.90**	*99.90	2.0
Araujo et al.,2015[10]	-	88.30	92.70	94.50	*96.30	83.4

[1] * = the value estimated from author's graphs

adaptive method dramatically improves the accuracy in all ranges. This result demonstrates that the adaptive approach provides significant benefits.

The first three lines of Table I present the difference between the adaptive GBDT method and the pruning ones when learning rate equals to 0.4. Obviously, the pruning strategy contributes to the accuracy increase in all ranges. On the other hand, the consumed time decreases about 15% with the threshold increase from 0.025 to 0.05. Experimental results indicate that the pruning method can be applied to improve the model's generalization ability and reduce the localization time meanwhile.

C. Comparison on BioID Database

Table I summarizes the accuracy and consumed time of the proposed method and the state-of-the-art methods which use the same database and metric. It can be observed that the proposed method significantly achieves much better performance than most of the other methods especially when err $\leq$ 0.025 and 0.05 in terms of accuracy. The results verify that the method proves to be an accurate localization approach.

We measure the average time to locate the eye pupil using the proposed method with standard parameters in a PC with an Intel Core i7-3.6GHz and 8G RAM. The adaptive approach with a pruning strategy consumes only 1.7ms in eye pupil localization, less than most of the other methods. It can be concluded that the proposed approach can reduce consumed time without accuracy loss.

V. CONCLUSION

In this paper, we propose an improved method for eye pupil localization based on adaptive gradient boosting decision tree. By increasing the weight of wrong samples and utilizing the pruning strategy, the proposed method achieves the accuracy of 92.39% at a speed of 1.7ms on BioID database. Experimental results demonstrate that the proposed method outperforms most state-of-the-art methods in terms of accuracy and consumed time.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China under Grant No. 61674103 and in part by Science and Technology Project of State Grid Corporation of China.

REFERENCES

[1] R. Valenti and T. Gevers, "Accurate eye center location and tracking using isophote curvature," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[2] W. Zhang, M. L. Smith, L. N. Smith, and A. Farooq, "Eye center localization and gaze gesture recognition for human–computer interaction," *JOSA A*, vol. 33, no. 3, pp. 314–325, 2016.

[3] Y. E. Soelistio, E. Postma, and A. Maes, "Circle-based eye center localization (cecl)," in *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*. IEEE, 2015, pp. 349–352.

[4] W. Zhang, H. Chen, P. Yao, B. Li, and Z. Zhuang, "Precise eye localization with adaboost and fast radial symmetry," in *International Conference on Computational and Information Science*. Springer, 2006, pp. 1068–1077.

[5] C. Zhang, X. Sun, J. Hu, and W. Deng, "Precise eye localization by fast local linear svm," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1–6.

[6] N. Markus, M. Frljak, I. S. Pandzic, J. Ahlberg, and R. Forchheimer, "Eye pupil localization with an ensemble of randomized trees," *Pattern Recognition*, vol. 47, no. 2, pp. 578 – 587, 2014.

[7] R. Valenti, N. Sebe, and T. Gevers, "Combining head pose and eye location information for gaze estimation," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 802–815, 2012.

[8] D. Yi, Z. Lei, and S. Li, "A robust eye localization method for low quality face images," in *Biometrics (IJCB), 2011 International Joint Conference on*, 2011, pp. 1–6.

[9] M. Zhou, X. Wang, H. Wang, J. Heo, and D. Nam, "Precise eye localization with improved sdm," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4466–4470.

[10] G. M. Araujo, F. M. L. Ribeiro, E. A. B. Silva, and S. K. Goldenstein, "Fast eye localization without a face model using inner product detectors," in *IEEE International Conference on Image Processing*, 2014, pp. 1366–1370.

[11] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2007.

[12] O. JESORSKY, "Robust face detection using the hausdorff distance," *AVBPA, 2001*, 2001.