

Research Article

Robust Eye Localization by Combining Classification and Regression Methods

Pak Il Nam,¹ Ri Song Jin,¹ and Peter Peer²

¹ *Institute of Mathematics, State Academy of Sciences, Democratic People's Republic of Korea*

² *Faculty of Computer and Information Science, University of Ljubljana, Slovenia*

Correspondence should be addressed to Pak Il Nam; pakilnam@yahoo.com

Received 24 November 2013; Accepted 4 March 2014; Published 30 March 2014

Academic Editors: Y. Dimakopoulos and Z. Huang

Copyright © 2014 Pak Il Nam et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Eye localization is an important part in face recognition system, because its precision closely affects the performance of the system. In this paper we analyze the limitations of classification and regression methods and propose a robust and accurate eye localization method combining these two methods. The classification method in eye localization is robust, but its precision is not so high, while the regression method is sensitive to the initial position, but in case the initial position is near to the eye position, it can converge to the eye position accurately. Experiments on BioID and LFW databases show that the proposed method gives very good results on both low and high quality images.

1. Introduction

Because face images should be normalized based on the coordinates of eyes in most face recognition systems, eye localization is an important part in face recognition systems. Its precision closely affects the performance of face recognition [1, 2]. Eye localization methods considering geometric properties of eyes such as edges, shape, and probabilistic characteristics are high in precision in normal conditions, but they are sensitive to illumination, pose, expression, and glasses [3–6].

State-of-the-art methods in eye localization are based on boosting classification, regression, boosting and cascade, boosting and SVM, and other variants [1, 2, 7–11]. In particular, the method in [1] is very effective, guaranteeing high precision even in unconstrained environment. It integrates the following three characteristics:

- (i) probabilistic cascade,
- (ii) two-level localization framework,
- (iii) extended local binary pattern (ELBP).

In eye localization, the boundary between the positive and negative samples is ambiguous, especially in low quality images. Thus, positive samples with low quality are easily

rejected by the thresholds in the cascade and fail to contribute to the final result. In [1] the authors introduced a quality adaptive cascade that works in a probabilistic framework (P cascade). In the P cascade framework all image patches have a chance to contribute to the final result and their contributions are determined by their corresponding probability. In this way P cascade can adapt to face images of arbitrary quality. Furthermore, they constructed two-level localization framework with a coarse-to-fine localization for the system to be robust and accurate. Figure 1 shows the size and geometry of the eye training samples for two-level stacked classifiers.

In order to enhance the classification ability, they introduced ELBP as improvement to LBP and multiblock LBP (MB-LBP) to improve the system's precision. Figure 2 shows the ELBP used in [1].

In this paper we propose an eye localization method with two-level localization framework, which is both robust and accurate even in unconstrained environment. In the coarse level we use the classification method similar to [1], primarily upgrading it with pyramid structure and postprocessing. Moreover, in the fine level we apply shape regression method similar to [12], where we improve the robustness primarily by using the coarse level information, normalization based on two-eye centers, and shape initialization.

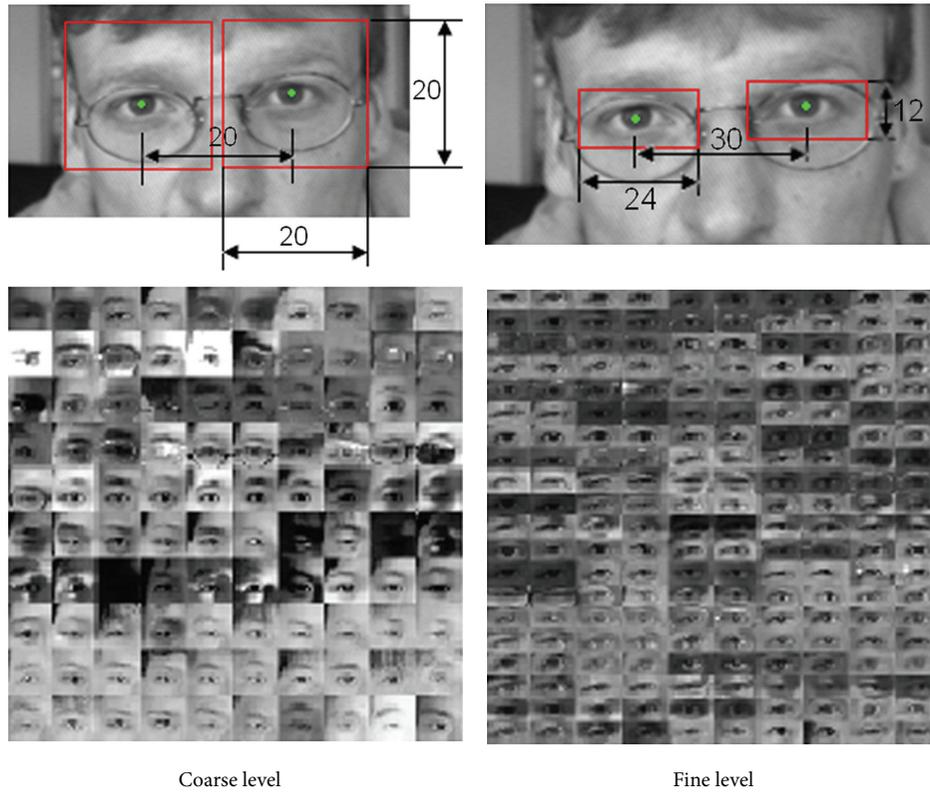


FIGURE 1: The size and geometry of training samples for two-level stacked classifiers.

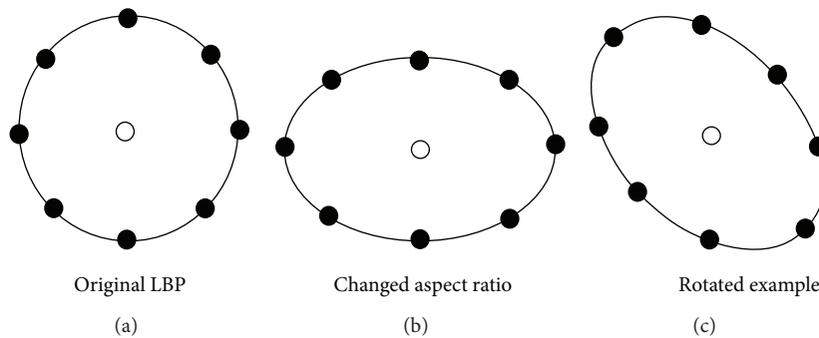


FIGURE 2: LBP and two ELBP operators.

This paper is organized as follows. Section 2 introduces proposed method through a definition of an eye center, analysis of the limitations of classification and regression method, and discussion of the two-level approach to eye localization, namely, the coarse level using classifier and fine level using regression. Section 3 conducts experiments on high and low quality databases to illustrate the superiorities of the proposed method. Section 4 concludes the paper.

2. Proposed Eye Localization Method

2.1. Definition of the Eye Center. In general, eye center is considered as the center of the pupil. But according to the eye gaze direction the center of pupil is offset to some extent. Furthermore, for closed eyes the center of pupil is not seen.

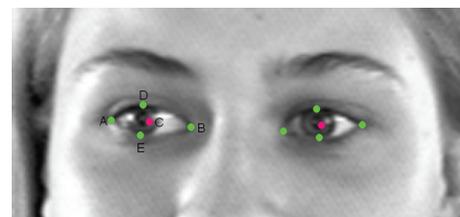


FIGURE 3: Definition of the eye center.

In order to estimate the eye center more accurately, we introduce a definition of the eye center. In Figure 3, $A = (A_x, A_y)$ and $B = (B_x, B_y)$ are left and right end points of the eye and $D = (D_x, D_y)$ and $E = (E_x, E_y)$ are upper and lower



FIGURE 4: Example of the eye localization failure by regression method.

points of the pupil, respectively. Then eye center $C = (C_x, C_y)$ is defined as

$$\begin{aligned} C_x &= \frac{(A_x + B_x)}{2}, \\ C_y &= \frac{(D_y + E_y)}{2}. \end{aligned} \quad (1)$$

This definition approximately coincides with the center of pupil in most cases and is able to give a robust eye center position even in case when pupil is offset much from the eye center. The definition can also give the state of two eyes, namely, closeness and openness. If the center of pupil is needed, one can estimate it using the points D and E easily.

The problem however is finding the method that can estimate the points A , B , D , and E . We are going to estimate these points by the combination of classification and regression method. In general, the classification method is robust but not so accurate. The classification method, which determines whether input data are positive or negative, is unable to express correctly how far the data are from the eye center. If the data, which are very near to the eye center, are used as negative examples, the training error increases, and low quality eye examples may be misrecognized as noneye examples. Unlike this, regression methods train the distance from current position to the eye center and consequently have better possibility for precision improvement. A drawback of regression method is its sensitivity to the initial position, namely, in case that initial position is far from the eye center, and there exist some patterns similar to the eye near to the initial position, and it may not converge to the eye center (Figure 4). In Figure 4 green triangles represent an initial shape and red squares represent the final shape obtained by regression method [12]. But if an initial shape is near enough to two-eye centers, the result is better.

In order to solve this sensitivity problem, we propose a robust and accurate eye localization method by the combination of classification and regression methods. We first estimate initial positions of two eyes by the coarse localization and then find around initial positions more accurate eye centers by the regression method.

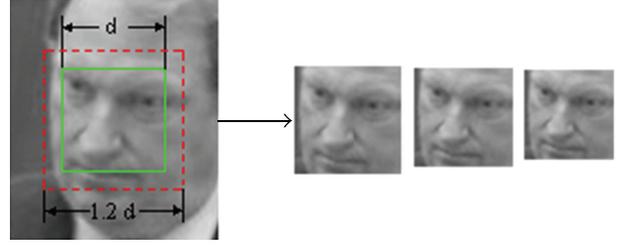


FIGURE 5: Extension of face detection region and pyramid construction.

2.2. Coarse Eye Localization. Because the main problem in coarse eye localization is robustness rather than precision under many factors, such as eyebrow, hair, and glasses, we perform the coarse eye localization by primarily paying attention to robustness.

2.2.1. Feature Selection and Classifier Construction. We use the same size and geometry of the eye training samples as in the coarse level of [1] (Figure 1). Furthermore, we use ELBP as the eye detection feature (Figure 2). Compared to LBP, ELBP has two radiuses and an angle, which makes the shape of ELBP a rotated ellipse. Given a 20×20 image, the dimension of ELBP feature set is very high when using 4 orientations. Considering the characteristic of ELBP, we use gentle boost algorithm to select efficient features and construct classifier. We also use P cascade with rejection stage and probability.

2.2.2. Eye Detection by Image Pyramid and Postprocessing. In general, eye detection is performed in the face region after face is detected. In literature on eye detection authors usually do not construct image pyramid because the ratio between the size of face region and the size of corresponding eye region usually falls in certain range. But this is not always satisfied and for some faces these ratios may exceed the expected range considerably. Moreover, if face region does not contain two eyes, eye detection fails. As shown in Figure 1, training samples we use are extracted under the condition that the distance between two eyes is constant. Thus, image pyramid structure is certainly needed for eye detection that is robust to scale variation of face detection regions.

In order to perform such robust eye localization, we first extend the face detection region in all directions, so that it contains two eyes completely, and normalize face region to 3 image pyramids: 60×60 , 55×55 , and 50×50 , and then eye detector scans each region. Figure 5 shows the extension of face detection region and pyramid construction.

Then we calculate the maximum rejection stage corresponding to P cascade from 3 image pyramids and find eye candidate positions with the maximum value. As shown in Figure 6, most eye candidate positions (each white pixel) are usually around eye center, but some candidates might be around either the frame of glasses, eyebrow, or similar.

Considering that most candidates are around eye center and their classifier probabilities are usually high, we first perform hierarchical clustering [13] on them. Then we determine the cluster having the largest number of candidates as

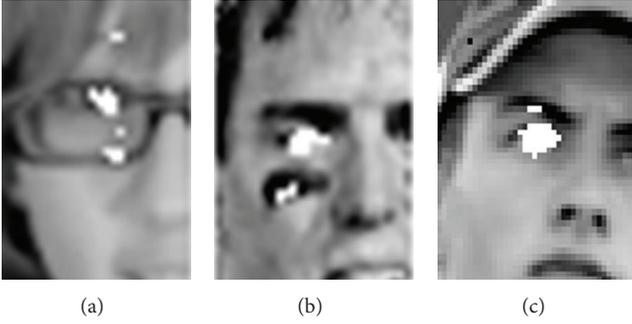


FIGURE 6: Detected eye candidate positions.

the best cluster. If two clusters have the same number of candidates, we select the best cluster by comparing their mean classifier probabilities. Furthermore, we select m candidates with the highest classifier probability in the best cluster and determine the best eye candidate position by weighted arithmetic average:

$$e^* = \frac{\sum_{e_i \in \{e_{\text{rank-}m}\}} p_i e_i}{\sum_{e_i \in \{e_{\text{rank-}m}\}} p_i}, \quad (2)$$

where e_i is a candidate in face image, p_i is the classifier probability of e_i , and $e_{\text{rank-}m}$ is the first m candidates with the highest classifier probabilities in the best cluster ($m = 5$ in our case). With this method, even in case that some non-eyes are recognized as eyes with high probabilities, we can accurately estimate eye centers.

2.3. Fine Eye Localization by Explicit Shape Regression. After eye candidate position is determined, fine eye localization for higher precision is performed. As described in Section 2.1, we convert eye localization problem into the estimation of left and right end points of an eye and upper and lower points of a pupil. Moreover, regarding robustness, we are convinced that considering two eyes at the same time is better than considering each eye individually.

On the basis of these considerations, we apply the explicit shape regression used in face landmark detection to refine eye localization [12].

2.3.1. Explicit Shape Regression. Given a facial image I and an initial face shape S^0 , each regressor computes a shape increment δS from image features and then updates the face shape in a cascaded manner:

$$S^t = S^{t-1} + R^t(I, S^{t-1}), \quad t = 1 \cdots T, \quad (3)$$

where the t -th weak regressor R^t updates the previous shape S^{t-1} to the new shape S^t .

The regressor R^t depends on both image I and previous estimated shape S^{t-1} . Given N training examples $\{(I_i, \widehat{S}_i)\}_1^N$ the regressor is obtained by explicitly minimizing the sum of alignment errors:

$$R^t = \underset{R}{\operatorname{argmin}} \sum_{i=1}^N \left\| \widehat{S}_i - (S_i^{t-1} + R(I_i, S_i^{t-1})) \right\|, \quad (4)$$

where S_i^{t-1} is the estimated shape in previous stage.

The regressors $(R^1, \dots, R^t, \dots, R^T)$ are sequentially learnt until the training error no longer decreases. In [12], each weak regressor R^t is learnt by a second level boosted regression, that is, $R^t = (r^1, \dots, r^k, \dots, r^K)$.

Shape indexed image features used in [12] are indexed only relative to S^{t-1} and no longer change when those r 's are obtained. Very simple and efficient Fern is used as each primitive regressor r^i . A Fern is a composition of F features and thresholds that divide the feature space and all training samples into 2^F bins. Each bin b is associated with a regression output δS_b that minimizes the alignment error of training samples Ω_b falling into the bin:

$$\delta S_b = \frac{1}{1 + \beta / |\Omega_b|} \frac{\sum_{i \in \Omega_b} (\widehat{S}_i - S_i)}{|\Omega_b|}, \quad (5)$$

where β is a free shrinkage parameter. When the bin has sufficient training samples, β has little effect; otherwise, it adaptively reduces the estimation. Efficient feature selection and Fern construction are performed by using fast correlation computation. Methods that generate feature set select efficient features and construct Ferns are the same as in [12]; thus, in continuation we discuss our contributions considering characteristics of fine eye localization.

2.3.2. Normalization Based on Two-Eye Centers. In order to apply explicit shape regression to eye localization, we randomly generate P pixels and calculate P^2 pixel difference features. Each pixel is indexed relative to the currently estimated shape S^{t-1} rather than the original image coordinates. A similarity transform to normalize the current shape to a mean shape is needed. Given the current shape S_i^{t-1} , two-eye centers $L = (L_x, L_y)$ and $R = (R_x, R_y)$ are easily calculated by (1). Then rotation and scale transformation based on two-eye centers can normalize the current shape to a mean shape. Denoting the distance between two eyes in mean shape by dist , rotation angle θ and scale constant α are determined as follows:

$$\theta = -\arctan \frac{(R_y - L_y)}{(R_x - L_x)}, \quad (6)$$

$$\alpha = \frac{\text{dist}}{\|L - R\|}.$$

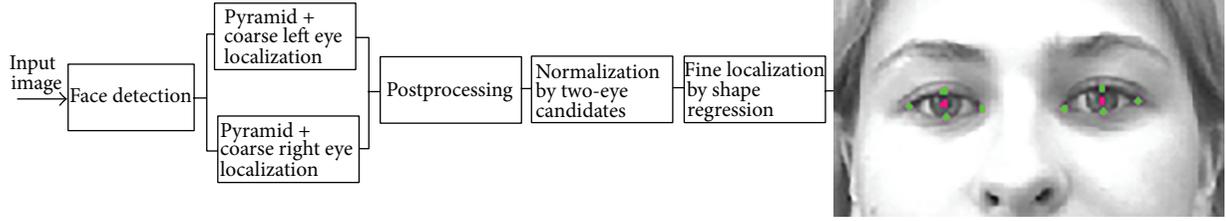


FIGURE 7: Eye localization flowchart.

Then the transformation that normalizes the current shape to a mean shape is

$$T(x, y, c, \alpha, \theta) = \alpha R_\theta \begin{pmatrix} x - c_x \\ y - c_y \end{pmatrix}, \quad (7)$$

where $R_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ and $c = (c_x, c_y)$ is a rotation center.

T^{-1} , the inverse of T , is then

$$T^{-1}(u, v, c, \alpha, \theta) = \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \frac{1}{\alpha} R_{-\theta} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (8)$$

2.3.3. Shape Initialization. As described in Section 2.1, regression method is somewhat sensitive to initial point. In other words, farther away from object position the initial points are, the smaller the convergence tendency to the points is. If we set initial shape using only face detection information, because of instability of face detection information in some cases, its shape (especially scale) may be quite different from the real shape.

To solve this problem, we use coarse eye localization information described in Section 2.2. But because coarse eye localization information is also not completely accurate, we observe the following. According to the experimental result in [1], even for LFW and VS (Video Surveillance) databases, which have low quality images, the success rate at normalized error $\text{err} < 0.25$ (9) is more than 99 percent (this result is from the cumulative curves for eye localization in [1]). Surely, for images that are not low quality the result is nearly 100 percent. Consequently, in order to model coarse eye localization information statistically, we add zero-mean Gaussian noise with constant variance to real eye centers (ground truth) so that their (noisy two eyes) normalized errors are smaller than 0.25. Based on noisy two-eye positions and (7), (8), we normalize the face regions to constant size and then these normalized images and corresponding real eye positions are used to explicit shape regression training.

After training, by using the trained regression function and the coarse positions of two eyes, we find left and right end points of an eye and upper and lower points of a pupil and determine two-eye centers using (1). Because of using the coarse positions of two eyes, the scales of the eyes are determined very robust and their initial positions are also near to the real positions. Consequently we overcome the limitation of the regression method.

Figure 7 shows the flowchart of proposed eye localization system. After performing face detection, the left and right eye

detectors scan tree level pyramid images, get eye candidate points, and determine the best candidate points for two eyes by using the proposed postprocessing. Then we normalize the input image based on the candidates of two eyes, find the left and right end points of an eye and upper and lower points of a pupil by regression method, and output the best positions of two-eye centers.

3. Experimental Results

To train an eye detector, we construct the training set which contains 12,000 face images from various databases including ColorFERET, MUCT, PICS, and CVL Face DB. These training images are considered as high quality images. The test set is divided into two categories: high quality and low quality. BioID database is used for high quality evaluation, and LFW database is used for low quality evaluation. BioID database consists of 1,521 images, while LFW database consists of 13,233 images.

The normalized error is used to evaluate the error between the localized eye positions and the ground truth:

$$\text{err} = \frac{\max(\|l_g - l\|, \|r_g - r\|)}{\|l_g - r_g\|}, \quad (9)$$

where l_g and r_g are the ground truth positions of the left and right eyes and l and r are the eyes positions localized by an algorithm, respectively. Considering that similarity drops in face recognition if normalized eye localization error is more than 5%, we use 5% and 10% normalized error to evaluate eye localization performance.

3.1. Coarse Eye Localization. Because left eye and right eye are almost symmetric, we may construct eye detector only for left eye and therefore we can use 24,000 eye training samples for coarse eye localization training.

Coarse eye localization training is similar as in [1], and it differs in our additional steps using pyramid structure and postprocessing. To show the effectiveness of the proposed pyramid structure and postprocessing, we trained coarse level method described in [1] and compared the performance with the proposed method. Tables 1 and 2 show the performance of both methods on BioID and LFW databases. As shown, the proposed method effectively improves the performance.

TABLE 1: Comparison of coarse eye localization methods on BioID.

	err < 0.05	err < 0.1
Coarse level from [1]	61.4%	96.14%
Proposed method	66.9%	96.66%

TABLE 2: Comparison of coarse eye localization methods on LFW.

	err < 0.05	err < 0.1
Coarse level from [1]	40.5%	85.1%
Proposed method	44%	89%

TABLE 3: Comparison of eye localization methods on BioID.

	err < 0.05	err < 0.1
[1]	70.5%	97.9%
Proposed method	92.6%	99.26%

TABLE 4: Comparison of eye localization methods on LFW.

	err < 0.05	err < 0.1
[1]	51.7%	87.9%
Proposed method	70.68%	95.40%

3.2. Fine Eye Localization. Training images used in shape regression training are the same as in coarse eye localization training, with addition of Gaussian noise to real eye positions of each image. The images are normalized as described in Section 2.3. Each training sample consists of a training image, an initial shape, and a ground truth shape. To achieve better generalization ability, we augment the training data by randomly sampling multiple (20 in our implementation) shapes from other annotated images as initial shapes of each training image. This is very effective in terms of robustness against large posevariation and rough initial shapes during the testing. We run the regressor several times (5 times in our implementation) and take the median result as the final estimation. We determined the parameters by cross-validation and set them to $F = 5$, $\beta = 1000$, $T = 9$, $K = 500$, $P = 500$, and $\text{dist} = 30$, as they represent good tradeoff between computational cost and precision.

Because the reported results of method [1] are the highest in literature, we compare our results with them. Tables 3 and 4 show the results: our method gets better results for both high and low quality images. Particularly for LFW database our method gives much better results and they show that our method works very well even for low quality images in unconstrained environment.

Furthermore, the numbers prove that results are substantially improved by the combination of classification and regression methods.

4. Conclusion

In this paper we proposed an eye localization method of high precision by the combination of classification and regression

methods. The proposed method is based on facts that classification method is robust but less accurate, while regression method is less robust but very accurate because it has more information about object position. Proposed method is both robust and accurate and gives the highest precision in terms of normalized error. We believe that the concept of the proposed method can be widely used not only in eye localization problem, but in many other object localization problems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] D. Yi, Z. Lei, and S. Z. Li, "A robust eye localization method for low quality face images," in *Proceedings of the International Joint Conference on Biometrics (IJCB '11)*, pp. 1–6, Washington, DC, USA, October 2011.
- [2] P. Wang, M. B. Green, J. Qiang, and J. Wayman, "Automatic eye detection and its validation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition—Workshops (CVPR '05)*, vol. 3, pp. 164–170, San Diego, Calif, USA, June 2005.
- [3] Z.-H. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recognition*, vol. 37, no. 5, pp. 1049–1056, 2004.
- [4] Q. Chen, K. Kotani, F. Lee, and T. Ohmi, "A robust eye detection approach based on edge related information," *International Journal of Computer Science and Network Security*, vol. 9, no. 9, pp. 22–27, 2009.
- [5] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas, "A novel eye-detection algorithm utilizing edge-related geometrical information," in *Proceedings of the European Signal Processing Conference*, Florence, Italy, September 2006.
- [6] Q. Chen, K. Kotani, F. Lee, and T. Ohmi, "An accurate eye detection method using elliptical separability filter and combined features," *International Journal of Computer Science and Network Security*, vol. 9, no. 8, pp. 65–72, 2009.
- [7] Y. Ma, X. Ding, Z. Wang, and N. Wang, "Robust precise eye location under probabilistic framework," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 339–344, Seoul, Korea, May 2004.
- [8] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gao, "2D cascaded AdaBoost for eye localization," in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 2, pp. 1216–1219, Hong Kong, August 2006.
- [9] X. Tan, F. Song, Z.-H. Zhou, and S. Chen, "Enhanced pictorial structures for precise eye localization under uncontrolled conditions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1621–1628, Miami, Fla, USA, June 2009.
- [10] X. Tang, Z. Ou, T. Su, H. Sun, and P. Zhao, "Robust precise eye location by adaboost and svm techniques," in *Advances in Neural Networks*, vol. 3497 of *Lecture Notes in Computer Science*, pp. 93–98, Springer, Berlin, Germany, 2005.
- [11] M. Everingham and A. Zisserman, "Regression and classification approaches to eye localization in face images," in *Proceedings of the 7th International Conference on Automatic Face and*

Gesture Recognition (FGR '06), pp. 441–446, Southampton, Uk, April 2006.

- [12] X. Cao, Y. Wei, F. Wen, and J. Sun, “Face alignment by explicit shape regression,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2887–2894, Providence, RI, USA, June 2012.
- [13] T. Hastie, *The Elements of Statistical Learning*, Springer, Berlin, Germany, 2008.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

