

EYE CENTER LOCALIZATION AND DETECTION USING RADIAL MAPPING

Karan Ahuja^G, Ruchika Banerjee^D, Seema Nagar^I, Kuntal Dey^I, Ferdous Barbhuiya^G

^GIIT Guwahati, ^DIIT Delhi, ^IIBM Research India

ABSTRACT

We propose a geometrical method, applied over eye-specific features, to improve the accuracy of the art of eye-center localization. Our solution is built upon: (a) checking radially constrained gradient vectors, (b) adding weightage to iris specific features and (c) considering bi-directional image gradients to eliminate errors due to reflection on pupil. Our system outperforms the state of the art methods, when compared collectively across multiple benchmark databases, such as BioID and FERET. Our process is lightweight, robust and significantly fast: achieving 50-60 fps for eye center localization, using a single threaded approach on a 2.4 GHz CPU with no GPU. This makes it practicable for real-life applications.

Index Terms— eye, eye center, localization, geometry based approach, radial mapping

1. INTRODUCTION

Several techniques, both intrusive and non-intrusive, exist for eye center localization. Special equipments, such as active infrared illumination, electrooculography (EOG), scleral search coils and head-mounted devices, can be used only under constrained setting, with different degrees of intrusiveness. Hence these are not viable in real-life applications.

Image-based approaches that use non-specialized equipment under non-intrusive settings, are affordable and usable in real-life settings. These would allow for free head movements, and detect the eye center in spite of different angles of position with respect to the camera capturing the eye image. Image-based approaches avoid special hardware, but call for development of challenging image-processing algorithms.

Algorithms for eye center localization tend to follow one of the following approaches [1]. (i) Shape-based approaches approximate eyes as a simple elliptical shape, or a more complex shape with a template, such as a deformable eye model comprising of two parabolas representing eyelids, a circle representing the iris *etc* [2,3]. (ii) Feature-based approaches use the typical human eye characteristics to identify a set of distinctive features around the eyes, such as eye corners, iris, pupil, eyelids, cornea reflections *etc* [4,5]. (iii) Appearance-based methods detect eyes based upon the photometric appearance characterized by color distribution or filter responses of the eye and surroundings [6,7]. (iv) Hybrid models combine

multiple other model types, attempting to overcome the shortcomings faced by each of these model types [3,8]. Some other works exist, exploring eye detection under active infrared illumination, symmetry operators and blink and motion.

We propose a geometry-based approach, based upon a specific set of eye features involving the iris and pupil. We create a four-stage process. First, we **preprocess**, performing histogram equalization for image intensity normalization to improve image contrasts. Second, we **perform adaptive skin color thresholding**, based on differential analysis of the image intensity to remove the homogeneous regions of the skin and limit our region of interest. Third, we extract **isocentric and radial curvature based patterns of the iris**, restricting our points of consideration within the expected eyeball radius values and considering the annulus design of the iris, improving both speed and accuracy. Fourth, we consider the **gradients directed inwards**, towards the eye pupil. This significantly reduces errors in cases where the pupil is brighter than the iris, such as light reflecting on the pupil.

We test with two popular databases: BioID [9] and FERET [10], and also test on the lesser-explored Talking Face Video database. Note that, Ge et al. [11] (95%) and Zhou et al. [12] (93.8%) claim higher accuracies on BioID compared to ours (92.06%). However, they do not test across multiple datasets, such as FERET or other databases. The former reports merely cross-validation results within BioID, with no separate testing dataset, making the study rather incomplete. The latter tests only on a part of the BioID database, showing results only on 1,251 images out of the available 1,521, surprisingly ignoring the remaining images. Further, it needs prior annotation for performing inference, which is impractical in real life.

Kroon et al. [13] provide reliable and accurate results across datasets. Our method also works well across datasets, across image sizes and resolutions, without any additional adaptation, and is accurate, outperforming Kroon [13] in most test conditions. We achieve a speed of 16 ms for detecting eye center pairs. Our algorithm, being accurate, robust, fast, lightweight and scale-invariant, can be used as a building block for applications that rely on eye center localization.

2. MODEL CONSIDERATIONS

Basic Definitions: Let c be a possible center and g_i the gradient vector at position p_i . Figure 1 shows examples of inward

and outward gradients. Then, at the eye center, the normalized displacement vector d_i will have the same orientation with the gradient vector g_i . We define d_i and g_i as:

$$d_i = (p_i - c) / |p_i - c|$$

$$g_i = (\delta I(x_i, y_i) / \delta x_i, \delta I(x_i, y_i) / \delta y_i)^T$$

For each pixel p_i , we compute the square of dot product of the displacement vector d_i (between the potential center c and the point p_i), and the image gradient g_i (at point p_i). The position of c , where most of the image gradients intersect, is the probable eye center. [14] [15]

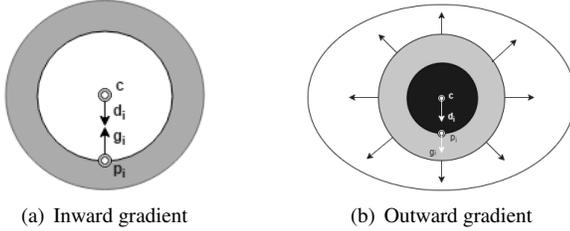


Fig. 1. Inward and outward gradients

2.1. Preprocessing for Normalization

We first detect the face using the well-accepted Viola and Jones face detector [16]. Based on the position and size of the detected face, we extract rough eye regions of interest. We use histogram equalization on the low-resolution images (images with face size less than 100×100), to spread the contrast distribution of the data throughout the spectrum equally by using a Cumulative Distribution Function (CDF) [17]. This increases the visibility of low resolution images while retaining the sharpness of the higher resolution ones, and therefore increases the accuracy of our system.

2.2. Adaptive Skin-Color Thresholding

Based on the skin color of the presented image, we carry out an adaptive thresholding process, to remove the homogeneous regions of the skin. This eliminates the points lying on the skin rather than inside the eye, limiting our region of interest, and hence (i) reduces the potential for errors and (ii) reduces the number of points to process, improving execution speed.

To perform adaptive skin-color threshold by differential analysis of image intensity, we compute the mean and standard deviation of the matrix comprising of the magnitude of the gradient vectors. Gradient magnitudes signify the differences between the pixel intensities. Two adjacent pixels will have similar magnitudes unless they have a sharp contrast, such as the darker pupil and iris on the lighter sclera or skin. We set a dynamic threshold as $mean + (standardDeviation * stdDevFactor)$, where $stdDevFactor$ represents a correction constant. We discard all gradient vectors below this threshold. The process is inherently adaptive, as the threshold value varies for each given

image, and is determined by the gradient vectors specific to that image. For experiments, we use $stdDevFactor = 14$ as the optimal value (observing over a range from 0.1 to 35).

2.3. Radius Based Cutoff: Prior Knowledge of Iris

The radius of the iris is close to an anatomical constant (around 7 mm) [18], and the radius of the eyeball ranges 12-13 mm according to anthropometric data [19]. We use this premise to implement radius based cutoff strategy. We only check the points within the eyeball, as the points outside cannot contain the eye center.

We note that, presence of light sources or strong reflections may inhibit the pupil to appear as the darkest point in the eye images, and may even be the brightest in some cases. However, the iris is always darker than the sclera, and surrounds the pupil. Hence, we assign a high weight to the points surrounding the eye center and within the iris radius based cutoff region, leveraging the annulus nature of the iris. The weighted matrix is calculated by applying a Gaussian Filter to the inverted preprocessed image of the eye ROI. This also reduces errors due to dark zones, as eyebrows, eyelashes and hair.

2.4. Gradients of the Inward Direction

In case of reflections or the light shining off the optic nerve, the iris will be darker than the pupil. To further tackle such cases and extract the isocentric pattern of the eye Figure 1, we take the gradient vectors that are inwards and opposite to the direction of the displacement vector (directed towards the pupil of the eye). After applying the weighted Gaussian blur that smoothens and darkens the high intensity parts of the image, the iris will be further darker compared to the pupil. Therefore, in such cases, considering the gradients g_i in the inward direction will point towards the probable center.

Thus, we have, a set of pixel positions after adaptive thresholding P , a calculated eye center c^* , probable eye centers c , normalized displacement vector d_i , normalized gradient vector g_i , total number of pixel positions in the image N , radius of the eyeball R , radius of the iris r , weight given to center based on darkness w_c , weight assigned to points within iris cut off based on darkness $f(w_p)$, and weight assigned to points based on distance from c as $f(d)$. The optimal center c^* of a circular object in an image with pixel positions p_i , $i \in \{1, \dots, N\}$, is calculated as:

$$c^* = \operatorname{argmax}_c \frac{1}{N} \sum_{i=1}^N \frac{w_c + f(w_p)}{f(d)} \left(d_i^T g_i \right)^2 \quad (1)$$

Our overall methodology is captured in Algorithm 1.

3. EVALUATION

For experiments, we use a hardware configuration of Inter Pentium CPU 2020M @ 2.40 GHz and 4 GB RAM.

Algorithm 1 OUR ALGORITHM

```
1:  $output(p \in \{1, \dots, N\}) = 0$ 
2: for  $i \in P$  do
3:   for  $c \in \{1, \dots, N\}$  do
4:     if  $|c - p_i| > R$  then
5:       continue
6:     end if
7:      $d_i = (p_i - c) / |(p_i - c)|$ 
8:      $g_i = (\delta I(x_i, y_i) / \delta x_i, \delta I(x_i, y_i) / \delta y_i)^T$ 
9:      $g_i = g_i / |g_i|$ 
10:    if  $|c - p_i| \leq r$  then
11:       $f(w_p) = w_{p_i}$ 
12:       $f(d) = 0.8$ 
13:    else
14:       $f(w_p) = 0$ 
15:       $f(d) = 1.0$ 
16:    end if
17:     $output(c) += ((w_c + f(w_p)) / f(d)) * (d_i^T g_i)^2$ 
18:  end for
19: end for
20:  $optimalCenter = \operatorname{argmax}_c((1/N) * (output))$ 
21: Output:  $optimalCenter$  as  $c^*$ 
```

3.1. Dataset Description

We run experiments on BioID [9], FERET [10] and Talking Face Video [20] databases. BioID is a collection of 1,521 gray-scale images, with the left and right eye centers annotated, across 23 different individuals, with varying locations, illumination, camera angles, scale and head pose. The color FERET database contains a total of 11,338 facial images collected by photographing 994 subjects at various angles. Out of this, 2,409 images are present in the frontal face (fa) and alternate frontal face (fb) partitions of the database, and thereby have a frontal view of the eyes present in the images. We test on this subset. We also test on the Talking Face Video [20], comprising of 5,000 image sequences of a person, designed to model the behavior of the face in natural conversations.

Normalized Error: We use the convention of measuring the *normalized error*, which has been the widely accepted form of error reporting in the literature of eye center localization. The measure of normalized error was introduced by Jerosky et. al. [23]. It estimates the error obtained by the worst of both eye estimations, measured as:

$$e \leq \frac{1}{d} \max(e_l, e_r) \quad (2)$$

Here e_l and e_r are the Euclidean distances between the estimated and the actual left and right eye centers. The distance between the actual eye centers is denoted as d . $e \leq 0.25 \approx$ is the distance between eye center and eye corners, $e \leq 0.10 \approx$ the iris diameter and $e \leq 0.05 \approx$ the pupil diameter.

Method	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.25$	AR
Database: BioID				
Valenti [3]	86.09%(5)	91.67%(6)	97.87%(5)	5.3
Fabian [14]	82.5%(6)	93.4%(5)	98.0%(4)	5
Yi [21]	86.5%(4)	99.1%(2)	99.6%(3)	3
Kroon [13]	92.3%(2)	97.9%(4)	99.9%(2)	2.6
Our	92.06%(3)	97.96%(3)	100%(1)	2.3
Zhou [12]	93.8%(1)	99.8%(1)	99.9%(2)	1.3
Database: FERET				
Valenti [3]	74.38%(1)	96.27%(3)	99.17%(3)	2.3
Kroon [13]	65.7%(2)	97.6%(2)	99.6%(2)	2
Our	64.57%(3)	98.25%(1)	99.87%(1)	1.6
Database: Talking Face Video				
Pang [22]	NA	96.2(2)%	NA	2
Our	94.78%(1)	99%(1)	99.42%(1)	1

Table 1. Performance of different eye center localization methods on various databases. AR \rightarrow Average Rank.

Database	Resolution	$e \leq 0.10$	$e \leq 0.25$
FERET	512X768	98.25%	99.87%
FERET	256X484	98.34%	99.95%
FERET	128X192	96%	99.5%
BioID	384X286	97.96%	100%
Talking Face Video	720X576	99%	99.42%

Table 2. Performance of our method across image resolutions

3.2. Characteristics of Our System

We provide a comparative performance measurement of the existing methods and our method on Table 1. The characteristics of our system are the following.

Accuracy: We test on the well-accepted BioID and FERET databases, as well as the lesser-explored Talking Face Video database. The works by Ge et al. [11] and Zhou et al. [12] claim higher accuracies compared to ours, on the BioID database. However, [11] (accuracy 95% for $e \leq 0.05$) perform only cross-validation and no test on unseen data, and is thereby prone to over-fitting. They also test on no other database or on real-life, thereby raising questions on model portability and variance with image scale. And [12] (accuracy 93.8% for $e \leq 0.05$) learn on an external dataset of 12,344 near-frontal images, but test only on a part of BioID, on 1,251 images out of 1,521, surprisingly ignoring the remaining. They further depend upon prior annotation of eye corners. Neither of the two works validate on any other benchmark database. This raises questions of practical portability and reliability of these studies.

Note that, as also pointed by [3], the annotation in the color FERET database is sometimes unreliable. Figure 3 shows some examples. Clearly, the ground truth labels of the pupil are often inaccurate, and are marked within the iris. Hence, our method yields (and other approaches also yield) high accuracies on FERET at $e \leq 0.10$ and $e \leq 0.25$, but under-

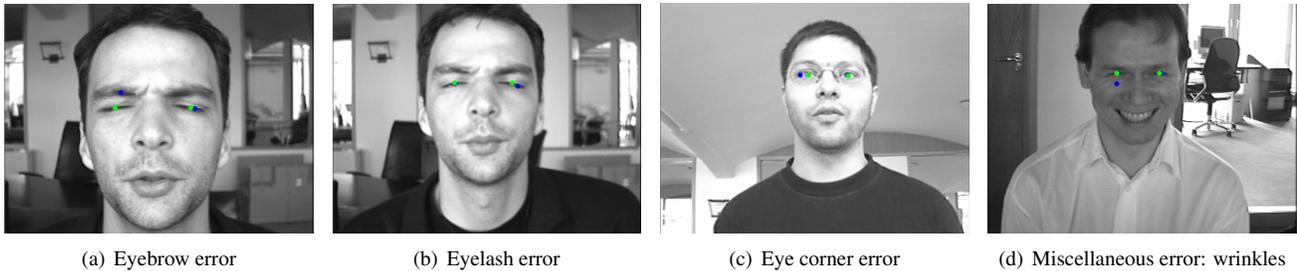


Fig. 2. Errors committed by our system on BioID. True eye center positions are in green and positions we estimate are in blue.

performs (64.57%) for $e \leq 0.05$ (Table 1). It is visibly obvious from Figure 3 that our method gives more accurate estimates of the eye center compared to the (often erroneous) ground truth labels, for $e \leq 0.05$.

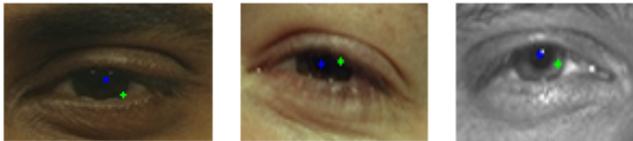


Fig. 3. FERET database. Green dots: Ground truth. Blue Dots: Estimated values from our method.

As shown on Table 1, our average rank is the best for FERET and Talking Face Video databases, and the second-best for the BioID database, ignoring the mere cross-validation based work by Ge et al. [11]. However, since the work by Zhou et al. [12] provides results only on a part of the BioID database, our system is practically the best for BioID also. Thus, for all practical purposes, across image scales, and across databases, our system is the most accurate one, among the ones can be used under practical settings.

Execution speed and stability: Our system is stable, and fast enough to use in real-life applications such as gaze tracking. Our methodology was able to scale to anywhere between 50-60 images (frames) per second using a single threaded approach, without GPUs. Systems using shift clustering and classifiers are not as stable and effective, when required to process a large number of frames per second. This gives our system a significant advantage over such other systems.

Simplicity of computation: Since our system does not use classifiers or clustering, and only uses geometrical features, as well as because of the cutoff and filtering, the practical load of computation is significantly lower on our method, compared to the other systems having comparable accuracy.

Scale invariance: Our system is scale invariant. Experiments show that our method performs well across different image resolutions over different databases (Table 2). The execution speed, accuracy, scale invariance, stability and low computing resource loads, makes our method usable in practical applications, including resource-constrained mobiles.

3.3. Error Analysis

We perform a detailed error analysis on all the databases, but present only for BioID here for space constraints.

Errors due to eyebrow and eyelid (eyelash) detection: The eyebrows and eyelids are detected due to the fact that they are predominantly dark. Hence when a weightage is applied to darker regions, their contribution increases and hence they are detected. The radius-based cutoff technique plays a major role in keeping the errors due to false pupil detections on the eyebrows to a negligible 0.067% for $e \leq 0.10$.

Errors due to eye corner detection: Eye corners have a structure somewhat symmetrical to the pupil. In low resolution blurred images, the junction between the eyelashes creates an almost circular dark structure. This is in contrast with the brighter sclera, hence receives a higher contribution.

Miscellaneous errors: A few other errors occur from stray factors such as wrinkles, spectacles and occlusions.

Table 3 shows the errors made by our system for $e \leq 0.10$. Figure 2 provides a view of different error types for BioID.

Error Type	Error Count	Error Percentage
Eyebrows	1	3.34%
Eyelashes	9	30%
Eye corners	14	46.66%
Miscellaneous	6	20%

Table 3. Error classification of our system for $e \leq 0.10$, Error percentage = $(|Error\ Count| / \sum |Error\ Count|) \times 100$

4. CONCLUSION

We attempted to detect and localize the eye center using a geometric approach built upon eye-specific features. We used a multistage radial mapping process as the core of our methodology, involving image histogram equalization, adaptive skin color thresholding, iris radius based cutoff and by specifically considering the directly inward gradients. Our method is robust and scale-invariant: it works across different datasets with different face image sizes and resolutions without any additional adaptation. We outperformed the state of the art when collectively compared across multiple benchmark databases. We process 50-60 frames per second on a single thread. This makes our system usable in real-life applications, such as gaze tracking and mobile based applications.

5. REFERENCES

- [1] Dan Witzner Hansen and Qiang Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 3, pp. 478–500, 2010.
- [2] Antonio Pérez, ML Cordoba, A Garcia, Rafael Méndez, ML Munoz, José Luis Pedraza, and F Sanchez, "A precise eye-gaze detection and tracking system," *UNION Agency-Science Press*, 2003.
- [3] Roberto Valenti and Theo Gevers, "Accurate eye center location through invariant isocentric patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1785–1798, 2012.
- [4] S Sirohey, Azriel Rosenfeld, and Zoran Duric, "A method of detecting and tracking irises and eyelids in video," *Pattern Recognition*, vol. 35, no. 6, pp. 1389–1401, 2002.
- [5] TD Orazio, Marco Leo, Grazia Cicirelli, and Arcangelo Distanto, "An algorithm for real time eye detection in face images," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. IEEE, 2004, vol. 3, pp. 278–281.
- [6] Peng Wang, Matthew B Green, Qiang Ji, and James Wayman, "Automatic eye detection and its validation," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 164–164.
- [7] Dan Witzner Hansen and John Paulin Hansen, "Robustifying eye interaction," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*. IEEE, 2006, pp. 152–152.
- [8] Fei Yang, Junzhou Huang, Peng Yang, and Dimitris Metaxas, "Eye localization through multiscale sparse dictionaries," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 514–518.
- [9] "Bioid database," <http://www.bioid.com/About/BioID-Face-Database>.
- [10] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss, "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [11] Shiming Ge, Rui Yang, Hui Wen, Shuixian Chen, and Limin Sun, "Eye localization based on correlation filter bank," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1–5.
- [12] Mingcai Zhou, Xiyang Wang, Haitao Wang, Jingu Heo, and Dongkyung Nam, "Precise eye localization with improved sdm," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4466–4470.
- [13] Bart Kroon, Sander Maas, Sabri Boughorbel, and Alan Hanjalic, "Eye localization in low and standard definition content with application to face matching," *Computer Vision and Image Understanding*, vol. 113, no. 8, pp. 921–933, 2009.
- [14] Fabian Timm and Erhardt Barth, "Accurate eye centre localisation by means of gradients.," in *VISAPP*, 2011, pp. 125–130.
- [15] Ravi Kothari and Jason L Mitchell, "Detection of eye locations in unconstrained visual images," in *Image Processing, 1996. Proceedings., International Conference on*. IEEE, 1996, vol. 3, pp. 519–522.
- [16] Paul Viola and Michael J Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [17] Mansour Asadifard and Jamshid Shanbezadeh, "Automatic adaptive center of pupil detection using face detection and cdf analysis," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2010, vol. 1, p. 3.
- [18] Rhys Newman, Yoshio Matsumoto, Sebastien Rougeaux, and Alexander Zelinsky, "Real-time stereo tracking for head pose and gaze estimation," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 122–128.
- [19] Kyung-Nam Kim and RS Ramakrishna, "Vision-based eye-gaze tracking for human computer interface," in *Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on*. IEEE, 1999, vol. 2, pp. 324–329.
- [20] "Talking face video," http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html.
- [21] Dong Yi, Zhen Lei, and Stan Z Li, "A robust eye localization method for low quality face images," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–6.
- [22] Zhiyong Pang, Chuansheng Wei, Dongdong Teng, Dihu Chen, and Hongzhou Tan, "Robust eye center localization through face alignment and invariant isocentric patterns," *PLoS one*, vol. 10, no. 10, pp. e0139098, 2015.
- [23] Oliver Jesorsky, Klaus J Kirchberg, and Robert W Frischholz, "Robust face detection using the hausdorff distance," in *Audio- and video-based biometric person authentication*. Springer, 2001, pp. 90–95.