

Circle-based Eye Center Localization (CECL)

Yustinus Eko Soelistio^{1,2}, Eric Postma², Alfons Maes²
 Information System Department¹, Tilburg center for Cognition and Communication²
 Universitas Multimedia Nusantara¹, Tilburg University²
 Tangerang, Indonesia¹, Tilburg, The Netherlands²

Abstract

The ability to automatically detect eye center locations in video images allows for estimating gaze direction. This, in turn, facilitates the study of human-computer interaction and behavioral analyses of social interactions. We propose an improved eye center localization method based on the Hough transform, called Circle-based Eye Center Localization (CECL) that is simple, robust, and achieves accuracy at a par with typically more complex state-of-the-art methods. The CECL method relies on color and shape cues that distinguish the iris from other facial structures. The circle enclosing the iris is localized by means of the Hough transform and the center of the iris is determined using the intensity level within the detected circle. The accuracy of the CECL method is demonstrated through a comparison with 15 state-of-the-art eye center localization methods against five error thresholds, as reported in the literature. The CECL method achieved an accuracy of 80.8% to 99.4% and ranked first for 2 of the 5 thresholds. It is concluded that the CECL method offers an attractive alternative to existing methods for automatic eye center localization.

1. Introduction

The automatic localization of the centers of eyes is an active area of research [1-9, 13-23]. In theory, finding the centers of the only circular structures in facial images does not seem very hard. However, in practice eye center localization is a tough challenge due to the noisy nature of natural visual images (occlusion by eyelids, hair or glasses) and inappropriate illumination conditions.

There are two types of methods to localize the eye centers. The first type relies on specialized devices, such as infrared cameras [1], the second type makes use of standard video cameras, such as webcams [2-9, 13-23]. Although the use of specialized devices can give very accurate results, their application is often limited to unnatural situations. Hence, there is a need for eye center localization methods applicable to standard video cameras in visible light.

Two state-of-the-art examples of video-based eye-center localization methods are due to Timm and Barth (2011) [3] and Valenti and Gevers (2008) [5]. Their methods can achieve high accuracy, which means that the error, i.e., the distance between the estimated and true eye center, is small. Both methods use color characteristics of the pupil and the iris to determine the approximate eye center location. However both methods reported that they suffered from interference from visual contours such as occlusion by highlight and objects located outside the iris-sclera region (such as eyebrows,

glasses, or hair). Similar types of interference are reported in the eye-gaze detection method proposed by Smith, Yin, Feiner, and Nayar (2013) [2].

The sensitivity to interference from adjacent visual contours may be overcome by adopting a model that relies on a shape cue for the irises. Circularity provides a unique cue to the irises, since no other visual structure in the face is circular. Several methods have been proposed using this idea, by combining edge detection [7, 8, 9] with the Hough transform (HT) [7-12]. Soltany, Zadeh, and Poureza (2011; henceforth referred to as SZP) [8] proposed a two-step eye center localization method. In the first step, the horizontal and vertical intensity histograms of the image region containing the eye are used to determine the region of interest (approximate eye region). A dark image region (iris) reveals itself as minima in the two histograms. Using the Euclidean locations of the minima, the SZP method determines the location of the region of interest (ROI). In the second step, within the ROI the eye center is located by searching for a circle using the HT. The SZP method works well on clearly visible images of the eyes, but its first stage fails on images containing dark artifacts Figure 1 illustrates two failures of the SZP method.

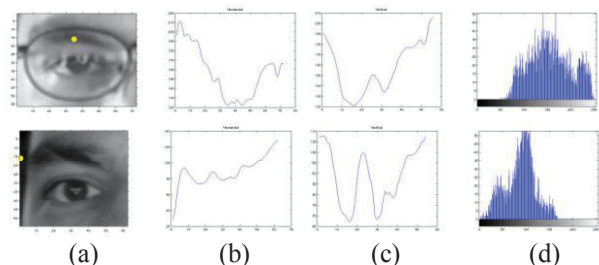


Figure 1. Illustration of failure of the SZP method. (a) Two images of the eye region containing dark artifacts (glasses and hair). (b, c) Horizontal and vertical histograms. (d) Image histograms. The small dots in (a) show the eye center estimates generated by the second stage of the SZP method.

This paper presents a new eye center localization method called the Circular-based Eye Center Localization (CECL) method that improves upon the SZP method by not suffering from dark artifacts.

The outline of the remainder of this paper is as follows. In Section II and III, we present the CECL method and its comparative. Sections IV and V contain the general discussion and conclusions.

2. The CECL method

The CECL method consists of three stages, (1) eye region detection, (2) eye-region pre-processing, and (3) eye

center localization. In the first stage, the approximate eye locations are estimated by first localizing the face region and then localizing the eye region within the face region. In the second, pre-processing stage, potential disturbing visual elements (e.g., eyebrows, hair, or wrinkles) are removed. Finally, in the third stage, circular objects are detected by means of the HT. In some cases where HT fails to detect any circular object, CECL uses the center of the eye region as approximate eye center.

1.1. Eye localization

Eye localization proceeds in four steps. First, the image is smoothed with a Gaussian filter (size of $5 \times 5px$, standard deviation = 5% of image width). Second the image region containing the face is detected. Third, the face region is cropped to 60% of its original height and the cropped image is cut into two equal halves along the vertical direction. Fourth, for each half, the eyes are detected. For both the face and the eye detection steps, the Viola-Jones algorithm is used (Matlab's and mexopencv's implementations).

Whenever during the fourth step the eyes cannot be localized, the CECL method exhaustively searches for the most probable eye regions.

1.2. Pre-processing

The main purpose of the pre-processing stage is to remove unwanted artifacts that may hamper the HT-based circle detection. Pre-processing proceeds in four steps (figure 2): (b) cropping, (c) contrast enhancement, (d) static binarization, and (e) morphological closing.

In the cropping step the top part of the image containing the eyebrow is removed. The size of the top part is defined by the parameter T_e .

In the second step the contrast of the cropped image is enhanced by means of histogram equalization.

The third step maps the equalized grayscale image onto a binary image by setting all pixels with a value $>T_b$ to 1 and to 0, otherwise. The optimal value of the threshold parameter T_b is determined by means of machine learning (5-folds cross validation). Ideally, the resulting binary image will show one or more full dark objects surrounded by a white background. However, some white spot might appear in the middle of these dark objects.

In the fourth step, morphological closing is used to remove these white spots.

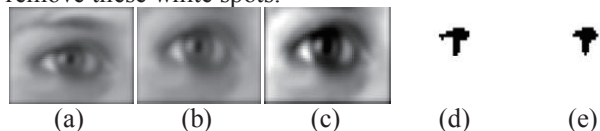


Figure 2. Illustration of the results of the four pre-processing steps. (a) Original eye image. (b-e) The four steps (see text).

1.3. Eye center localization

In the third stage of the CECL method, consists of two steps: detecting the iris and finding its center. The circular HT [7-12] is used for detecting the circle enclosing the iris from the binary image. Given the resulting circular region, the minimum intensity value within the circular region is determined from the corresponding gray scale image.

Figure 3 provides an illustration of CECL's operation.

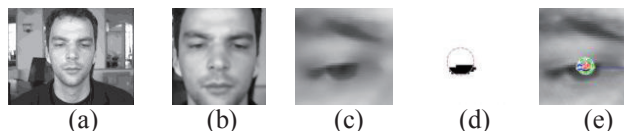


Figure 3. Illustration of the CECL method's operation. (a) Original image. (b) Result of face detection. (c) Result of eye detection. (d) Result of circle detection in binary image. (e) Eye center localization in the original image.

3. Performance evaluation

CECL's performance is evaluated on the BioID database [13], a challenging set of facial pictures taken under realistic conditions. The database contains 1521 grayscale images ($384 \times 286 pixels$) from 23 different subjects in different locations and under various illumination conditions. Some of the subjects wear glasses and some subjects have curly hair obstructing their eyes. They have different head pose and eye states. Detection performance is defined as the ratio of Euclidean distance between the prediction error (distance between the true and the prediction position of the irises) and the true distance of both eyes.

We evaluate CECL in terms of accuracy by adopting the standard measure of normalized error (e) [13]:

$$e_{worst} = \frac{\max(d_l, d_r)}{\|C_l - C_r\|} \quad (1)$$

where C_l and C_r are the Euclidean positions of the true eye center locations of the left and right eyes, and d_l and d_r are the Euclidean distances between the predicted and real eye center locations, for the left and right eye respectively.

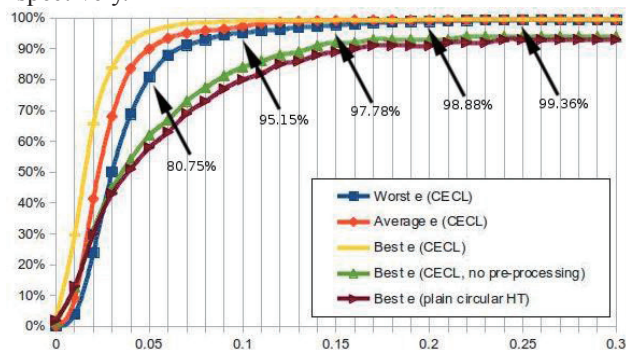


Figure 4. Performance results of CECL for three different error measures. The vertical axis represents the detection performance (%) and the horizontal axis represents the error threshold e .

Timm and Barth [3] associated different error values as follows: $e = 0.25$ corresponds roughly to the distance between the eye center and the eye corner, $e = 0.10$ the diameter of the iris, and $e = 0.05$ the diameter of the pupil. Predictions are considered good when $e \leq 0.05$.

We compare the CECL accuracy test result with the results from fifteen other methods. We also provide two additional normalized errors, the upper bound error, $e_{best} = \frac{\min(d_l, d_r)}{\|C_l - C_r\|}$, and the average error, $e_{average} = \frac{d_l + d_r}{2 \times \|C_l - C_r\|}$ [3, 5] to allow for comparison with other me-

thods.

We implemented our method using Matlab R2012a and mexopencv on an Intel i5 2.5GHz multicore PC. CECL was evaluated using 5-fold cross validation.

The top three curves in figure 4 show the detection performances of CECL as a function of threshold e , for e_{best} (top curve), $e_{average}$ (middle curve) and e_{worst} (lowest of the three curves). CECL can reach an accuracy of 80.75% ($e \leq 0.05$), 95.15% ($e \leq 0.1$), and 97.78% ($e \leq 0.15$). The accuracy is relatively constant when $e > 0.15$. The bottom two curves in figure 4 show the performance of CECL without pre-processing (upper curve), and the result from a plain circular HT (lower curve). Apparently, the combination of pre-processing and the circular HT contributes considerably to the per-

formance.

To compare CECL’s performances with those of state-of-the-art methods, we adopt table 1 from Timm and Barth [3] extended with the evaluations reported in table 1 in Leo, Cazzato, De Marco, and Distante [23]. Adaptive methods are listed separately from non-adaptive methods to facilitate the comparison. Table 1 shows that CECL is outperforming all other methods when $e \leq 0.15$ and $e \leq 0.20$, and ranks second in all others except in $e \leq 0.05$ where it ranks third.

In terms of processing speed, the CECL method is somewhat slower than competitive methods. Figure 5 shows examples of successful (a-d) and unsuccessful (e-h) eye center localizations by the CECL method.

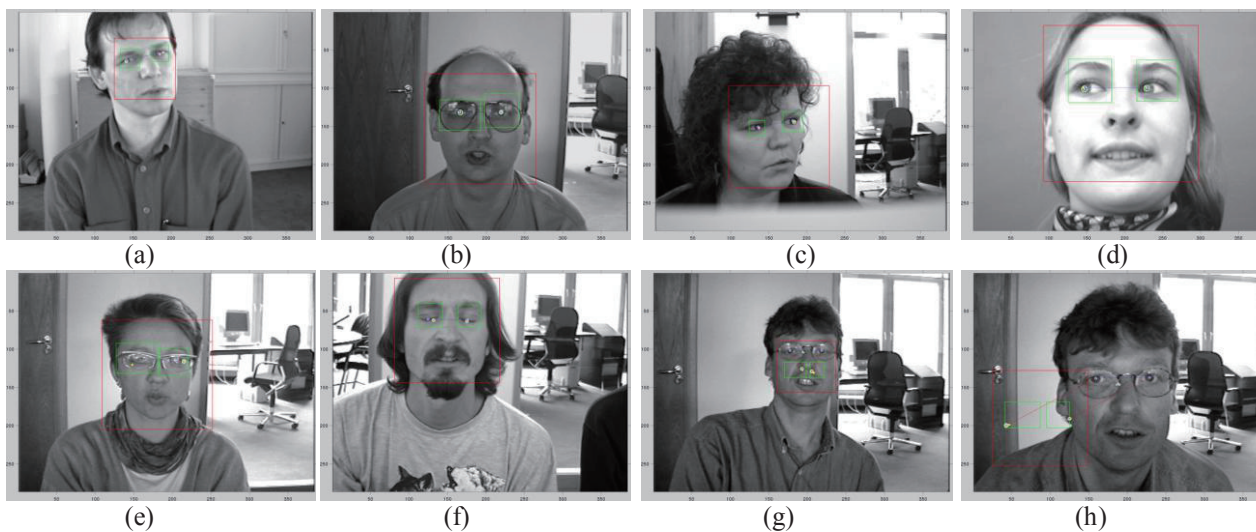


Figure 5. Samples of CECL eye center localization.

4. Discussion

The CECL method performs on a par with state-of-the-art eye center localization methods. The use of the HT to find circular objects appears to overcome the main limitation of image gradient methods [3], i.e., the interference of artifacts in the eye region (eyebrow, eyelid, and hair). CECL can increase the accuracy simply by limiting the ROI to the circular area determined using the HT. We ran a few tests on high-resolution images (3840×2400 pixels) and found that CECL achieves good predictions ($e \leq 0.03$). The superior performance of CECL is mainly due to its effective combination of pre-processing, which removes most of the possible obstructions and accentuates the circular characteristic of the iris, and the HT.

Nevertheless, CECL’s performance is restricted by its reliance on the presence of a visible circular feature of the iris. Specific illumination conditions can render the circular structure of the iris invisible, which gives rise to errors in CECL’s adaptive binarization stage. The erroneous localizations in Figure 5(e-h) have various sources: strong reflection from glasses (e), thick eyelids during eye closure (f), incorrect eye localization (g), and incorrect face localization (h). These examples illustrate the two main weaknesses of the CECL method: inaccurate face/eye detection and lack of “no detection” output. The inaccurate face and eye detections follow from the limitations of the employed Viola-Jones implementations.

Future versions of CECL will be based on improved detectors. The lack of a “no-detection” output follows from CECL’s implicit assumption that the iris is always visible. As a consequence, in cases where the iris is totally occluded, as in Figure 5(f), CECL will generate a false positive.

5. Conclusion

We proposed the CECL method, a simple and accurate eye center localization method using a HT-based circle detection. The CECL method performs at a level comparable with state-of-the-art methods, is less sensitive to interference from artifacts than gradient methods, is relatively simpler than other HT-based eye center localization methods, and operates successfully on low resolution images. By combining the shape and intensity cues, CECL can surpass other methods that rely on only either one cue.

Future research will address the two main weaknesses of the CECL method in an attempt to further boost eye-center localization performance.

We conclude by stating that the CECL method provides a viable approach to eye-center localization that deserves further study to achieve top level performance.

Acknowledgments: The authors would like to thank Bart Joosten for his valuable comments and suggestions.

Table 1. Comparison of CECL with other methods. (•) methods that do not use learning model. (S) method consider to be thecurrent state-of-the-art. **Bold**value indicate best accuracy. (*) indicate second best accuracy.

Method	$e \leq 0.05$	$e \leq 0.10$	$e \leq 0.15$	$e \leq 0.20$	$e \leq 0.25$	Note
Asadifard and Shanbezadeh (2010)	47.0%	86.0%	89.0%	93.0%	96.0%	•
Asteriadis et al. (2005)	44.0%	81.7%	92.6%	96.0%	97.4%	•
Zhou and Geng (2004)	-	-	-	-	94.8%	•
Timm and Barth (2011)	82.5%*	93.4%	95.2%	96.4%	98.0%	•, S
Leo et al. (2014)	80.7%	87.3%	88.8%	90.9%	-	•
Kroon et al. (2008)	65.0%	87.0%	-	-	98.8%	
Valenti and Gevers (2008)	84.1%	90.9%	93.8%	97.0%	98.5%	MIC+SIFT+kNN, S
Turkan et al. (2007)	18.6%	73.7%	94.2%	98.7%*	99.6%	
Campadelli et al. (2006)	62.0%	85.2%	87.6%	91.6%	96.1%	
Niu et al. (2006)	75.0%	93.0%	95.8%	96.4%	97.0%	
Chen et al., (2006)	-	89.7%	-	-	95.7%	
Hamouz et al. (2005)	58.6%	75.0%	80.8%	87.6%	91.0%	
Christinacce et al. (2004)	57.0%	96.0%	96.5%*	97.0%	97.1%	
Behnke (2002)	37.0%	86.0%	95.0%	97.5%	98.0%	
Jesorsky et al. (2001)	38.0%	78.8%	84.7%	87.2%	91.8%	
CECL	80.8%	95.2%*	97.8%	98.9%	99.4%*	

References

- [1] K. Aai, and R. Mardiyanto: "Comparative Study on Blink Detection and Gaze Estimation Methods for HCI, in Particular, Gabor Filter Utilized Blink Detection Method," *Eighth International Conference on Information Technology: New Generation*, pp. 441-446, 2011.
- [2] B. A. Smith, Q. Yin, S. K. Feiner, and S. K. Nayar: "Gaze Locking: Passive Eye Contact Detection for Human-Object Interaction," *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 271-280, 2013.
- [3] F. Timm, and E. Barth: "Accurate Eye Center Localization by Means of Gradients," *VISAPP*, pp. 125-130, 2011.
- [4] M. Asadifard, and J. Shanbezadeh: "Automatic adaptive center of pupil detection using face detection and cdf analysis," *Proceedings of the IMECS*, pp. 130-133, 2010.
- [5] R. Valenti, and T. Gevers: "Accurate Eye Center Location and Tracking Using Isophote Curvature," *Computer Vision and Pattern Recognition*, pp. 1-8, 2008.
- [6] S. Asteriadis, N. Nikolaidis, A. Hajdu, and I. Pitas: "An Eye Detection Algorithm Using Pixel to Edge Information," *Int. Symp. on Control, Commun. and Sign. Proc.*, 2006.
- [7] D. E. Benn, M. S. Nixon, and J. N. Carter: "Robust eye centre extraction using the Hough transform," *Audio-and Video-based Biometric Person Authentication*, pp. 1-9, 1997.
- [8] M. Soltany, S. T. Zadeh, and H. R. Pourreza: "Fast and accurate pupil positioning algorithm using circular Hough transform and gray projection," *International Conference on Computer Communication and Management*, 2011.
- [9] Y. Ito, W. Ohshima, T. Wakabayashi, and F. Kimura: "Detection of eyes by circular Hough transform and histogram of gradient," *Pattern Recognition, 21st International Conference*, pp. 1795-1798, 2012.
- [10] H. K. Yuen, J. Princen, J. Illingworth, and J. Kittler: "Comparative study of Hough transform methods for circle finding," *Image and Vision Computing*, pp. 71-77, 1990.
- [11] J. Illingworth, and J. Kittler, "A survey of the Hough transform," *Computer vision, graphics, and image processing*, 1988, 87-116.
- [12] R. O. Duda, and P.E. Hart: "Use of Hough transformation to detect lines and curve in pictures," *Communications of the ACM*, pp. 11-15, 1972.
- [13] O. Jesorsky, K.J. Kircheberg, and R. W. Frischholz: "Robust Face Detection Using the Hausdorff Distance," *Audio-and video-based biometric person authentication*, pp. 90-95, 2001.
- [14] B. Kroon, A. Hanjalic, and S. Maas: "Eye localization for face matching: is it always useful and under what conditions?," *Proceedings of the 2008 CIVR*, pp. 379-388, 2008.
- [15] M. Turkan, M. Pardas, and A. E. Cetin: "Human eye localization using edge projections," *Proceedings of the 2nd VISAPP*, pp. 410-415, 2007.
- [16] P. Campadelli, R. Lanzarotti, and G. Lopori: "Precise eye localization through a general-to-specific model definition," *Proceedings of the 17th BMVC*, pp. 187-196, 2006.
- [17] Z. Niu, S. Shan, S. Yan, X. Chen, and W. Gaw: "2d cascaded adaboost for eye localization," *Proceedings of the 18th IEEE ICPR*, pp. 1216-1219, 2006.
- [18] D. Chen, X. Tang, Z. Ou, and N. Xi: "A hierarchical floatboost and mlp classifier for mobile phone embedded eye location system," *Proceedings of the 3rd ISNN*, pp. 20-25, 2006.
- [19] M. Hamouz, J. Kittler, J. Kamarainen, P. Paalanen, H. Kalviainen, and J. Matas: "Feature-based affine-invariant localization of faces," *IEEE Transactions on PAMI*, 27(9):1490.
- [20] Z. Zhou, and X. Geng: "Projection functions for eye detection," *Pattern Recognition*, 37(5):1049-1056, 2004.
- [21] D. Cristinacce, T. Cootes, and I. Scoot: "A multistage approach to facial feature detection," *Proceedings of the 15th BMVC*, pp. 277-286, 2004.
- [22] S. Behnke, "Learning face localization using hierarchical recurrent networks," *Proceedings of the ICANN*, pp. 135-135, 2002.
- [23] M. Leo, D. Cazzato, T. De Marco, and C. Distanto, "Unsupervised eye pupil localization through differential geometry and local self-similarity matching," *PLoS one*, 9(8), e102829, 2014.