A Robust Eye-Corner Detection Method for Real-World Data

Gil Santos and Hugo Proença

Abstract—Corner detection has motivated a great deal of research and is particularly important in a variety of tasks related to computer vision, acting as a basis for further stages. In particular, the detection of eye-corners in facial images is important in applications in biometric systems and assisted-driving systems. We empirically evaluated the state-of-the-art of eye-corner detection proposals and found that they achieve satisfactory results only when dealing with high-quality data. Hence, in this paper, we describe an eye-corner detection method that emphasizes robustness, i.e., its ability to deal with degraded data, and applicability to real-world conditions. Our experiments show that the proposed method outperforms others in both noise-free and degraded data (blurred and rotated images and images with significant variations in scale), which is a major achievement.

I. INTRODUCTION

A corner is defined by the intersection of at least two edges. For decades, it was believed that most primitives of the human visual system were based on the detection of such points, which have well-defined positions. Corner detection is known to have particular relevance in computer vision, as it is often used as a starting point for other image recognition processes. Hence, various corner detection strategies have been emphasized in previous investigations of image segmentation, tracking, recognition and motion detection systems.

In this paper, we are particularly interested in the detection of both the temporal and nasal eye-corners in facial images. Eye-corners constitute relevant points of interest, and the ability to accurately pinpoint them is of great value in areas, such as biometrics, and applications, such as driving assistance systems. In biometrics, an emerging type of recognition is called periocular, based on human recognition by using data collected from around the eyes. The periocular region is particularly useful when the quality of data reduces the efficacy of other recognition strategies, such as with uncooperative subjects, when using visible light imagery or when acquiring data from moving subjects at a distance (e.g., [12], [10], [9], [15]).

Among all of the points of interest that can be extracted from the periocular region, we highlight eye-corners – the intersections between the upper and lower eyelids – because the position of eye corners does not vary with different facial expressions, levels of eye closure, gaze, eyelashes or makeup. After reviewing the state-of-the-art research on eye-corner detection, we concluded that published methods lack robustness and were developed to operate successfully only with high-quality data. We empirically determined that the performance of these approaches tends to significantly deteriorate with real-world data of significantly higher heterogeneity. Hence, this work proposes an eye-corner detection method suitable for imperfect environments, such as uneven lighting conditions and rotated or blurred data, with substantial differences in scale and levels of eye closure. Our method uses a periocular image as input, segments the iris and the sclera and defines a region of interest from which candidate points are extracted. Then, multiple features are linearly combined in an objective function whose optimization determines the pair of points that constitute the nasal and temporal eye-corners.

A. Related Works

Several approaches for the detection of eye-corners can be found in the literature. Harris and Stephens [6] proposed a general purpose corner detection method, which is often used in the specific case in which eye-corners with satisfactory results with high-quality data are available. Zheng et al. [17] estimated an initial region of interest from integral projections and located eye-corners according to a bank of Gabor-based filters, convolved at five different scales and orientations, from which averaged outputs yielded the final detection kernel. A more in-depth description of this strategy can be found in [18]. Khoorsavi and Safabakhsh [7] localized eye-corners in gray data, starting from the center of the iris and selecting two points on its scleric boundary at symmetric angles. Next, they found points on the eyelids according to local differences in brightness and used four masks to define motion direction. Xu et al. [16] used the approach of Harris and Stephens to select candidate points and then parsed them, combining semantic features using logistic regression. However, this method relies on image edges, which are difficult to obtain in unconstrained acquisition environments. Haiying and Gouping [5] proposed the weighting of Harris’s response function with the variance projection function, achieving a more robust system for frontal images with no significant lighting variations or rotation. The variance projection function itself was proposed for similar purposes by Feng and Yuen [4]. More recently, Erdogmus and Dugelay [3] proposed a method that achieves good results on frontal images but also heavily relies on edge detection, and eye-corners result from the interception of polynomial functions fitted to these edges.

The remainder of this paper is organized as follows: in Section II we describe our methods in detail; Section III
presents our experiments and discusses the obtained results, with an emphasis on the robustness factor. Finally, Section IV presents our conclusions.

II. PROPOSED METHOD

A. Iris Segmentation and the Definition of the Region of Interest

As illustrated in figure 1, our method uses a periocular image as the input, and the first step is to obtain the corresponding noise-free iris binary segmentation mask. This mask discriminates between the noise-free regions of the iris and all of the remaining data and was obtained as described by Tan et al. [14]. This method has been shown to be effective with real-world data. In addition, this iris segmentation algorithm was selected because it outperformed in the NICE.I contest 1. The segmented iris data are represented by the black regions of figure 1(b) and contain holes that correspond to the pupil and occluded iris regions. These holes were removed by zeroing out all of the regions that were unreachable when filling out the background from the edges of the image, as described in [13].

Next, we defined a region of interest (ROI) from which subsequent processing would be completed. This region is illustrated in figure 2 and was obtained by cropping the input image and the segmentation mask, avoiding unnecessary regions, such as the eyebrow and the skin underneath the eye. With an input image of dimensions $M \times N_0$, this yields regions of dimensions $M \times N_1$, according to horizontal projection techniques. This procedure ensures that the ROI is composed of the extreme coordinates of pixels belonging to the iris ($P_u$ and $P_l$ of figure 2):

$$
y_u = \max(y_p)$$
$$
y_l = \min(y_p)
$$

where $y_p$ are the row coordinates of all of the pixels that belong to the iris.

B. Sclera Segmentation

The localization of regions that correspond to the sclera inside the ROI is critical to our method, as both eye corners should be adjacent to the sclera. In addition, pixels belonging to the human sclera have particularly low levels of saturation, which is illustrated by figure 3. The left image gives the saturation channel of the HSV colorspace (figure 3(a)), and the right image shows the result of the convolution with a unidimensional horizontal median filter [8] for eyelash attenuation, followed by data quantization and histogram equalization (figure 3(b)). This example illustrates that the sclera became more homogenous and had evidently lower intensities, enabling their classification using empirically adjusted thresholds.

C. Eye Contour Approximation

Once the iris and sclera were segmented, the next stage involved approximating the contours of the eyelids. This was performed in two steps: 1) a morphological dilation of the iris segmentation mask with a horizontal structuring element, which horizontally expands the iris regions, and 2) a point-by-point multiplication between the dilated and the enhanced data illustrated in figure 3(b), as described by Caselles [2]. We obtained an image similar to that illustrated in figure 4(b) and whose boundary constitutes a close approximation of the contours of the eyelids.

1NICE.I: Noisy Iris Challenge Evaluation - Part I http://nicel.di.ubi.pt
D. The Generation of Eye-Corner Candidates

This stage involved the generation of a set of candidate points for the positions of the eye-corners, which was performed by using the approach of Harris and Stephens [6]. However, because of the high probability of producing too many false positives, this detector was exclusively applied inside the nasal (Rn) and temporal (Rt) regions, cropped from the extremes of the major axis of the sclera mask, as illustrated in figure 5.

Fig. 5. An approximation of the eyelid contour (white snake) and the regions from which corner candidates are extracted (represented by white rectangles).

E. Feature Set

This stage involved finding the appropriate features to discriminate between the set of corner candidates. We also wanted to ensure that such a feature set would be robust in response to differences in translation, rotation, scale, affine-transformation and blurred data. In all subsequent descriptions, we consider \( \{c_i\}_{i=1}^n, c_i = (x_i, y_i) \) to be the set of eye-corner candidates.

a) Harris Pixel Weight \( H(P_i) \): Because all candidates were generated according to the Harris and Stephens method, it is straightforward to include the corresponding score in the proposed feature set. This score is given by

\[
H = |M| - k \tr(M)^2 \tag{2}
\]

where \(|.|\) denotes the determinant, \(\tr(.)\) is the trace of a matrix and \(M\) is the Hessian matrix obtained from a blurred version of the original data:

\[
M(x,y) = \begin{bmatrix}
G_{xx}(x,y) & G_{xy}(x,y) \\
G_{yx}(x,y) & G_{yy}(x,y)
\end{bmatrix} = I(x,y) \otimes h(x,y), \text{with } h(x,y) = \frac{1}{\pi} \exp \left( \frac{x^2 + y^2}{2} \right) \text{ and } \otimes \text{ denotes convolution.}
\]

b) Internal Angles: Let \( B = \{b_i\}_{i=1}^N, b_i = (x_i, y_i) \) be the set of pixels belonging to the eyelid boundary obtained as described in section II-C. An ellipse fitted to \( B \) points is parameterized as follows:

\[
E = (x_e, y_e) + Q(\gamma) \begin{bmatrix} A \cos(\sigma) \\ B \sin(\sigma) \end{bmatrix} \tag{3}
\]

where \((x_e, y_e)\) is the central point of the ellipse, \(Q(\gamma)\) is a rotation matrix and \(A\) and \(B\) are the lengths of the major and minor axes, respectively. Two sets of pixels located along the opposite directions of the ellipse’s minor axis are given by

\[
b_i = (x_e - \cos \left( \gamma - \frac{\pi}{2} \right) \cdot B, y_e - \sin \left( \gamma - \frac{\pi}{2} \right) \cdot B) \tag{4a}
\]

For every candidate point \( c_i \), two vectors \( \overrightarrow{u} = c_i - b_u \) and \( \overrightarrow{v} = c_i - b_v \) were obtained, and their internal angle \( \theta(c_i, E) \) is given by

\[
\theta_1(c_i, E) = \arccos \left( \frac{\langle u, v \rangle}{||u|| \cdot ||v||} \right) \tag{5}
\]

where \( \langle u, v \rangle \) is the dot product between \( u \) and \( v \), and \( || \cdot || \) denotes the norm of a vector.

Let \( m_1 \) be the slope of the ellipse’s major axis and \( m_2 \) be the slope of the line connecting \((x_e, y_e)\) and the candidate point \( c_i \):

\[
m_2 = \frac{y_e - y_i}{x_e - x_i} \tag{6}
\]

Their internal angle measures the agreement between the directions of the ellipse’s major axis and the straight line that passes through the candidate point and the center of the ellipse:

\[
\alpha_2(c_i, E) = \arctan \left( \frac{m_2 - m_1}{1 + m_1 \cdot m_2} \right) \tag{7}
\]

Finally, because we are interested in pairs of eye corners, we found it useful to obtain a feature that relates any two candidates as a pair rather than scoring them independently. Let \( c_{i1} \) and \( c_{i2} \) be two corner candidates, one from the temporal and the other from the nasal region, and let \( l_{12} \) be the line that passes through both points. If the plausibility of both candidates is high, the direction of \( l_{12} \) should be similar to that of the major axis of the previously defined ellipse \( E \). Thus, according to (7), we obtained the internal angle between these vectors (\( \alpha_3(c_{i1}, c_{i2}, E) \)).

c) Positions in ROIs: A complementary feature measures the relative position of each candidate in the ROIs, i.e., the proportion of pixels inside the ROI that are above each candidate. This feature is given by

\[
p(c_i, R) = \frac{\sum_{i=1}^{N_1} \sum_{j=1}^{M_1} I_{R}(i,j) \in R}{\sum_{i=1}^{N_1} \sum_{j=1}^{M_1} I_{R}(i,j) \in R} \tag{8}
\]

where \( I_{R} \) is an indicator function.

d) Relative Distances: This type of feature considers the distance between each candidate point \( c_i \) and the center of the ellipse:

\[
d_1(c_i, E) = \sqrt{\left( x_i - x_e \right)^2 + \left( y_i - y_e \right)^2} / A \tag{9}
\]

where \((x_e, y_e)\) denotes the coordinates of the center of the ellipse and \(A\) the length of the ellipse’s major axis to compensate for the imbalance between acquisition distance and eye size.

Let \( \overrightarrow{v}_{tan} \) be a vector with the same direction of the major axis of the ellipse and \( p_1 = (x_1, y_1) \) and \( p_2 = (x_2, y_2) \) be the antipodal points of the ellipse. Let \( p_{tan} = (x_{tan}, y_{tan}) \) be a point tangential to the ellipse that belongs to a line that passes through \( c_i \).
\[ x_{\tan} = x_1 + u(x_2 - x_1) \]
\[ y_{\tan} = y_1 + u(y_2 - y_1) \]

where \( u \) is given by:
\[
u = \frac{(x_c - x_1)(x_2 - x_1) + (y_c - y_1)(y_2 - y_1)}{||p_2 - p_1||^2}
\]

The Euclidean distance between \( p_{\tan} \) and each candidate \( c_i (d_2(c_i, p_{\tan})) \) was also added to the feature set.

d) The Intersection of Interpolating Polynomials: The nasal and temporal eye corners can be regarded as the intersections between the upper and lower eyelids. Because of this, we parameterized two lines, each corresponding to one eyelid. The intersections \( t \) of both polynomials are illustrated by figure 6 and provide a rough estimate of the upper and lower eyelids, respectively. Thus, the Euclidean distance between each candidate and the interception point of the corresponding ROI \( (d_2(c_i, t)) \) also acts as a measure of goodness for that candidate.

\[ \text{Fig. 6. Interpolating second (upper eyelid) and third (lower eyelid) degree polynomials. The interception points of both polynomials constitute an accurate approximation of the eye corners.} \]

F. Objective Function

According to the description given in section II-E, the proposed feature set is composed of seven features: \( F = \{b(c_i), \beta_1(c_i, E), \beta_2(c_i, E), p(c_i, R), d_1(c_i, E), d_2(c_i, p_{\tan}), \text{and} d_3(c_i, t)\} \), which should be fused to produce the final score. With two sets of corner candidates (nasal and temporal), the final score for every pair of nasal \( c_n \) and temporal \( c_t \) candidates is given by the weighted sum of these features:
\[
\Gamma(c_t, c_n) = \sum_{i=1}^{7} \beta_i f_i + \sum_{j=8}^{14} \beta_j f_{j-7}
\]

where \( \{\beta_1, \ldots, \beta_{14}\} \) are regularization terms adjusted to maximize performance in a training set. This optimization procedure was carried out by linear regression, and these terms were adjusted to minimize the mean squared error between the predicted values and the ground-truth data using the Akaike criterion [1]:
\[
J(c_t, c_n) = (\Gamma(c_t, c_n) - g(c_t, c_n))^2
\]

Regularization coefficients were estimated on a sub-set of frontal images, resampled in a ten-fold cross-validation.

III. EXPERIMENTS

A. Datasets

The performance of the proposed method was assessed on right-eye images of the UBIRIS.v2 database [11]. The images have dimensions of 400 \( \times \) 300 pixels and were acquired from moving subjects in visible wavelengths at different distances and under varying lighting conditions. Additionally, the quality of the images was degraded by different factors, such as blur, motion, rotation and gaze. To check the reduction in the performance of the proposed method with respect to each factor, five dataset configurations were used and are illustrated in figure 7:

- **Frontal** – includes 300 images with the subjects’ gazes aligned toward the camera;
- **Deviated Gaze** – 200 images in which the subjects’ heads were deviated;
- **Blur** – images with an artificially made 50-pixel-length motion blur in the \( \pi/4 \) direction;
- **Clockwise rotation (CR)** – images artificially rotated by \( \pi/8 \) clockwise;
- **Counter-clockwise rotation (CCR)** – the same as the previous but with a counter-clockwise rotation.

For the Blur, CR and CCR experiments, the images selected from the UBIRIS.v2 database were not enough, and variations were artificially made by image processing software, starting from the **frontal** subset. For all images, the data were ground-truthed manually by different experts in order to reduce subjectivity.

\[ \text{Fig. 7. Sample images from the different datasets.} \]

B. Results

Based on the analysis of previously published research, the type of data used in this research and the results reported by the authors, we compared the performance of our method to the strategies employed by Haiying and Guoping [5] and Erdogmus and Dugelay [3]. The methods we compare ourselves to were implemented on the scope of this work and, although designed for different databases, were the ones best fitting our purposes. In addition, because we found that one of the proposed features (the intersection of polynomials) constitutes a strong estimator even when used alone, we also included this feature in our comparisons (Polyfit I.). All of the error values provided in this section correspond to the Euclidean distance between the estimated location of the eye-corners and the true location obtained by a manual annotation of all the images in our datasets.
methods that are generally more efficient in detecting the nasal eye-corner, with the exception of the Erdogmus and Dugelay strategy. Again, the proposed method outperformed the previous methods.

C. Analysis of Bias

To analyze the errors that are predominant in the outputs of each method, for each case, we obtained a vector $\vec{v} = (m, \theta)$, where $m$ is the Euclidean distance between the estimated $(x_e, y_e)$ and true $(x_t, y_t)$ corner position, and $\theta$ is the arctangent of $(x_e - x_t, y_e - y_t)$. The relative

Figure 8 gives the results obtained for frontal images, which is the data subset that in appearance most closely resembles the type of data the other methods are concerned with. Figure 8(a) provides the global detection rates, and figures 8(b) and 8(c) specify the results obtained for the temporal and nasal eye-corners. The horizontal axes denote the error values, and the vertical axes illustrate the proportion of images with such error values. From the analysis, it is evident that the proposed approach clearly outperformed previously reported strategies in the frontal images. When the analysis was performed separately for the nasal and temporal corners and for the temporal region, the polynomial interpolation interception was more accurate than the Erdogmus and Dugelay method, and in most cases, it showed performance similar to the proposed method. Regarding the nasal corners, we observed that all three methods behave similarly for small error values, whereas our proposal is notably better for moderate and large error values (larger than 25 pixels).

For the sake of clarity, figure 9 compares the boxplots of the error values observed for the proposed method and the methods used for comparison in the temporal (black bars) and nasal (gray bars) corners. The median of the observed performance range (horizontal solid lines) and the first and third quartile values of the observations (top and bottom of the box marks) are shown. The upper and lower whiskers are denoted by the horizontal lines outside of each box, and the outliers are denoted by dot points. This plots highlights the

Fig. 9. The distances between the predicted corners and the true locations on frontal images. Black and gray represent the temporal and nasal regions, respectively.

Fig. 10. The relative frequencies of the observed deviations between the predicted and true positions of eye-corners. The left and right images represent the temporal and nasal corners, respectively.
frequency of these values is illustrated in figure 10, where the horizontal axis denotes the angle, and the vertical axis denotes magnitude. Deviations from the proposed method and from the polynomial interpolation interceptions are homogeneously distributed in all directions, slightly skewed toward the $[0, \frac{\pi}{2}]$ interval. Considering that our datasets are composed exclusively of right-eye images, the estimates tend to be biased northeast of the true eye-corners. On the nasal region, the prediction tends to be closer to the center of the eye than the true location. This fact is especially evident for the estimates using the Haiying and Guoping method. With the Erdogmus and Dugeley approach, temporal deviations were observed more rarely, with a slight predominance to the right of the true corner. Whereas the other methods seem to have a clear bias toward the center of the face in the nasal region, deviations were spread in all directions with the Erdogmus and Dugeley approach. This atypical behavior shown by the Erdogmus and Dugeley method in both regions probably results from the fact that, as this method is heavily dependent on edge detection, it is also considerably affected by data degradation. Notably, such distributions of deviations are in concordance with the observed correlation values, where a higher similarity between the proposed method, the interception of the polynomials and the Haying and Guoping methods was observed.

D. Robustness to Variations in the Data

Robustness is a key requirement for the proposed method, and we aimed to assess the decrease in performance when the quality of the data was degraded by different factors. In this analysis, we decided to exclusively compare the results obtained by the proposed method with those obtained using the Haiying and Guoping approach, as the latter is considered a state-of-the-art approach, and its performance was closest to ours. Figure 11 summarizes the obtained error values in the dataset, where the images were substantially degraded as a result of the corresponding factor. The black boxplots denote the results of our method and the gray bars those determined by the Haiying and Guoping method. The analysis demonstrates the higher stability of the performance of our method across the different datasets, as the average error values are steady and remain under 50 pixels. The performance of the Haiying and Guoping method, in contrast, notably diminished when handling rotated iris data, simultaneously increasing its variance.

Figure 12 highlights these decreases in performance and provides the detection rates with respect to the error value (in pixels). Here, the higher slope of our method’s performance plots for small errors is especially evident, which may indicate that large errors in the estimates are quite unlikely, as opposed to the values observed for the other strategy.

![Figure 12. Detection rate as a function of the distance for all image variations.](image)

1) Blur: Acquiring sharp data in less controlled acquisition environments is an important issue, as slight movements of subjects often correspond to severely blurred data, a result of small depth-of-focus ranges. Thus, the ability to handle blurred data is a desirable property of any robust corner detection method. Our method only slightly decreased in performance, whereas Haiying and Guoping’s performed better in some circumstances (distances from 55 to 130 pixel present an higher detection rate), with blurred data than with the focused images. The minor degradation in performance of our proposal occurred during the stage that defines the ROIs, as illustrated in figure 13; the edges become less prominent in blurred data, the region growing process stops at different iterations and consequently, the candidate search areas are also different. This, coupled with the fact that the blur also degrades the performance of the method used for the extraction of the candidates, led to a worse outcome in our proposed method.

![Figure 13. Extraction of candidate points in frontal image and in the corresponding blurred version.](image)

2) Deviated Gaze: Gaze is another important factor in less controlled acquisition environments, as it is expected that most of the time, a subject’s head and eyes will not be aligned with the camera. In this case, our method behaves robustly, which was regarded as extremely positive and may indicate good performance with this type of data. There was a typical case in which our method performed better than the others: when the images had a visible background or notable facial
elements (e.g., the nasal bone). Figure 14 illustrates such cases and highlights the robustness of the proposed method for deviations in gaze.

![Image](image_url)

Fig. 14. An illustration of the results typically obtained in gaze-deviated images. White squares and black circles represent the outputs of our method and Haiying and Guoping’s method, respectively.

3) Rotation: Rotation is another case of special interest, and significant rotations in data are expected as a result of different types of movements in an uncontrolled acquisition scene. Again, our method showed a much more robust behavior than the approach of Haiying and Guoping, which had a significantly diminished performance. We believe that this was the result of the vertical and horizontal variance projection functions that produce different results in rotated data and, consequently, bias further processing. This is highlighted by figure 15(a), in which a visible predominant bias in the opposite direction of the rotation can be seen. This is in opposition to our method, as illustrated in figure 15(b), in which a different behavior for each corner was observed: in the nasal corner, vectors counteract the direction, but angle changes are minimal. For the temporal corner, the prediction tends to follow the rotation with a larger angle variation.

![Image](image_url)

Fig. 15. The relative frequencies of the deviations in clockwise rotated data. The images on the left and right images are of the temporal and nasal regions, respectively.

IV. CONCLUSIONS

Several researchers are working on eye-corner detection, and the performances of different proposed methods have been found to significantly diminish in response to degraded data acquired under less controlled conditions. These shortcomings led us to propose a new method for the detection of eye-corners in periocular images that simulate real-world data. We compared the results obtained by our proposal to other state-of-the-art methods and concluded that our method consistently outperformed these methods, both when operating with noise-free and with degraded data (rotated, blurred, affine-transformed and with significant differences in scale). Finally, these improvements were obtained without significant increases in the computational demands of the task, which is a significant asset, considering the real-time demands that eye-corner detection techniques typically impose.

REFERENCES