A Fully Automatic Approach for the Accurate Localization of the Pupils

Marco Leo, Dario Cazzato, Tommaso De Marco, and Cosimo Distante

National Research Council of Italy - Institute of Optics
via della Libertà, 3 73010 Arnesano (Lecce)
marco.leo@cnr.it

Abstract. This paper presents a new method to automatically locate pupils in images (even with low-resolution) containing human faces. In particular pupils are localized by a two steps procedure: at first self-similarity information is extracted by considering the appearance variability of local regions and then they are combined with an estimator of circular shapes based on a modified version of the Circular Hough Transform. Experimental evidence of the effectiveness of the method was achieved on challenging databases containing facial images acquired under different lighting conditions and with different scales and poses.

Keywords: Self-similarity, Saliency, Circularity Analysis, Pupil Localization.

1 Introduction

As one of the most salient features of the human face, eyes and their movements play an important role in expressing a person's desires, needs, cognitive processes, emotional states, and interpersonal relations. For this reason the definition of a robust and non-intrusive eye detection and tracking system is crucial for a large number of applications: advanced interfaces, control of the level of human attention, biometrics, gaze estimation for example for marketing purposes, etc. A detailed review of recent techniques for eye detection and tracking can be found in [1], where it is clear that the most promising solutions use invasive devices (*active eye localization and tracking*). In particular some of them are already available on the market and require the user to be equipped with a head mounted device [2], while others obtain accurate eye location through corneal reflection under active infrared (IR) illumination [3]. Passive eye detection and tracking systems are only recently introduced and they attempt to obtain information about eye location just starting from images acquired from one or more cameras. Most popular approaches in this area use complex shape models of the eye: they work only if the important elements of the eye are visible and then zoomed, or high resolution views are required [4]. Other approaches explore the characteristics of the human eye to identify a set of distinctive features around the eyes and/or to characterize the eye and its surroundings by the color distribution or filter responses. The method proposed by Asteriadis et al. [5] assigns a vector to every pixel in the edge map of the eye area, which points to the closest edge pixel. The length and the slope information of these vectors are consequently used to detect and localize the eyes by matching them with a training

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 503–512, 2013.

set. Timm and al. [6] proposed an approach for accurate and robust eye center localization by using image gradients. They derived an objective function whose maximum corresponds to the location where most gradient vectors intersect and thus to the eye's center. A post-processing step is introduced to reduce wrong detection on structures such as hair, eyebrows, or glasses. In [7] the center of (semi)circular patterns is inferred by using isophotes. In a more recent paper by the same authors, additional enhancements are proposed (using mean shift for density estimation and machine learning for classification) to overcome problems that arise in certain lighting conditions and occlusions from the eyelids [8]. A filter inspired by the Fisher Linear Discriminant classifier is instead proposed in [9] to localize the eyes. A sophisticated training of the left and right eye filters is required. In [10] a cascaded AdaBoost framework is proposed. Two cascade classifiers in two directions are used: the first one is a cascade designed by bootstrapping the positive samples, and the second one, as the component classifiers of the first one, is cascaded by bootstrapping the negative samples. A method for precise eye localization that uses two Support Vector Machines trained on properly selected Haar wavelet coefficients is presented in [11]. In [12] an Active Appearance Model (or AAM) is used to model edge and corner features in order to localize eye regions.

Unfortunately, the analysis of the state of the art reveals that most of the methods uses a supervised training phase for modeling the appearance of the eye or, alternatively, introduces ad-hoc reasoning to filter missing or incorrect detection. For this reason, although leading to excellent performance in specific contexts, they can not be directly used in different contexts (especially in the real world ones) without some adjustments of the models previously learned.

This paper explores the possibility to introduce a pupil detection approach that does not require any training phase (or post filtering strategy). It detects the pupil in low-resolution images by combining self-similarity and circularity information: in other words the total variability of local regions are characterized by saliency maps that are then related with gradient based features specifying point-wise circularity. This way the proposed approach is more suitable to operate in real contexts where it is not generally possible to ensure uniform boundary conditions. Experimental evidence of the effectiveness of the method was achieved on benchmark data sets containing facial images acquired under different lighting conditions and with different scales and poses.

2 Overview of the Proposed approach

Figure 1 schematically shows the main steps of the proposed approach. Each input image is initially analyzed by using the boosted cascade face detector proposed by Viola and Jones [13]. The rough positions of the left and right eye regions are then estimated using anthropometric relations. In fact, pupils are always contained within two regions starting from 20×30 percent (left eye) and 60×30 percent (right eye) of the detected face region, with dimensions of 25×20 percent of the latter [14]. The innovative procedure based on the combination of self-similarity and circularity information is finally applied to the cropped patches in order to accurately find the pupil location.

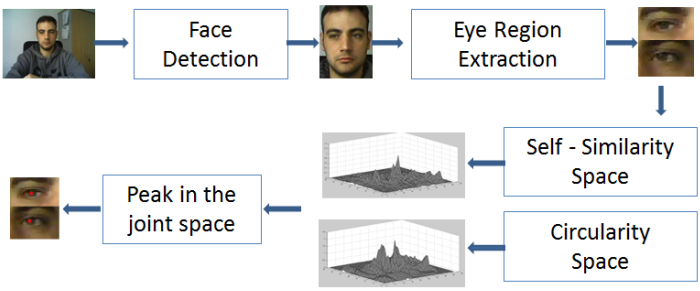


Fig. 1. A quick overview of the proposed multistep approach for pupil localization

2.1 Self-similarity Space Computation

The key idea is to initially find regions with high self-similarity, i.e. regions that retain their peculiar characteristics even under geometric transformations (such as rotations or reflections), changes of scale, viewpoint or lighting conditions and possibly also in the presence of noise. Self-similarity score can be effectively computed as a normalized correlation coefficient between the intensity values of a local region and the intensity values of the same geometrically transformed local region [15]. A local region is self-similar if a linear relationship exists:

$$I(T(x)) = a + bI(x) \qquad \forall x \in P \tag{1}$$

where P is a circular region of radius r and x is a point located in P. $I(x)$ denotes the intensity value of the image I at location x, and T represents a geometric transformation defined on P. For the purposes of the paper, T is limited to a reflection and a rotation. Both reflection and rotation, preserve distances, angles, sizes, and shapes. To better clarify the notions of reflection and rotation into the specific context under consideration, point locations can be represented in polar coordinates, hence $x = (r, \phi)$. Every reflection is associated to a mirror line going through the center of P and having orientation denoted by $\vartheta \in [0; 2\pi]$. Having said that, a reflection is defined as the geometric transformation that maps the location (r, ϕ) to location $(r, 2\vartheta - \phi)$.

Similarly every rotation has a centre and an angle. Let the centre of the rotation be the centre of P and let the rotation angle α be one of the angles $\frac{2\pi}{n}$, where n is an integer. A rotation maps the location (r, ϕ) to location $(r, \phi + \alpha)$.

Given these preliminary concepts, from the operational point of view, the cornerstone of this first phase is the search of the points that are closest to satisfy the condition in equation 1 considering that on real data, it can hardly be fulfilled for all points of P. This way, highlighted points should correspond to the pixels of the eye which has both (almost) radial and rotational symmetry. In particular the strength of the linear relationship in equation 1 can be measured by the normalized correlation coefficient:

$$ncc(P, T) = \frac{\sum_i (I(x_i) - \bar{I})(I(T(x_i)) - \bar{I})}{\sqrt{(\sum_i (I(x_i) - \bar{I}))^2 (I(T(x_i)) - \bar{I})^2}} \tag{2}$$

Here i counts all points of P and $\bar{I}$ represents the average intensity value of points of P.

At a given location, the normalized correlation coefficients in equation 2 can be computed for different mirror line orientations or different angles of rotation. All give information of region self-similarity.

In this paper the average normalized correlation coefficient computed over all orientations of the mirror line (*radial similarity map* S) at a given location is used as a measure of region self-similarity[1].

Let the sampling intervals for θ be $\Delta\theta = \frac{2\pi}{N}$ the similarity measure is then computed as

$$S(P) = \frac{1}{N}\sum_{i=0}^{N-1} ncc(P, T_{\theta_i}) \tag{3}$$

To overcome the problems related to the processing near the borders, for the calculation of the self-similarity scores, a wider area (10 pixels in each direction) is considered and then the portion of the self-similarity space relative to the original size of the eye patches is extracted. At the end a $(m \times n \times M)$ data structure S_r is available where $m \times n$ is the size of the input image and M is the number of sampled radii r (i.e. the number of considered scales). Local maxima and minima are then obtained by comparing each pixel value to its eight neighbors in the current similarity map and nine neighbors in the scale level above and below. A point is selected only if it has an extreme value compared to all its neighbors. The self-similarity map (of size $m \times n$) at the scale corresponding to the region (among the selected ones) with the highest self similarity score is the outcome of this first phase.

2.2 Circularity Measurements

The second phase starts with the estimation of the spatial distribution of circular shapes and this is done by a modified version of the Circular Hough Transform (CHT). A circle detection operator, that is applied over all the image pixels, produces a maximal value when a circle with a radius in the range $[Rmin, Rmax]$ is detected. It is defined as follows:

$$u(x,y) = \frac{\int\int_{D(x,y)} \boldsymbol{e}(\alpha,\beta)\cdot \boldsymbol{O}(\alpha - x, \beta - y) d\alpha d\beta}{2\pi(R_{max} - R_{mim})} \tag{4}$$

The domain D(x,y) is defined as:

$$D(x,y) = \{(\alpha,\beta) \in \Re^2 | R_{min}^2 \leq (\alpha - x)^2 + (\beta - y)^2 \leq R_{max}^2\} \tag{5}$$

where $\boldsymbol{e}$ is the normalized gradient vector:

$$\boldsymbol{e}(x,y) = [\frac{E_x(x,y)}{|E|}, \frac{E_y(x,y)}{|E|}]^T \tag{6}$$

[1] The self-similarity coefficients computed when T is a reflection are equal to those computed when T is a rotation. This has been mathematically proven in [15]

and $\boldsymbol{O}$ is the kernel vector

$$\boldsymbol{O}(x,y) = [\frac{\cos(\tan^{-1}(y/x))}{\sqrt{x^2+y^2}}, \frac{\sin(\tan^{-1}(y/x))}{\sqrt{x^2+y^2}}]^T \quad (7)$$

The use of the normalized gradient vector in the equation (4) is necessary in order to have an operator whose results are independent from the intensity of the gradient in each point.

2.3 Pupil Localization

The final step of the proposed approach integrates the selected self-similarity map and the circular shape distribution space. Both data structures are normalized in the range [0,1] and then point-wise added. The peak in the resulting data structure is the point that is finally selected as the pupil center.

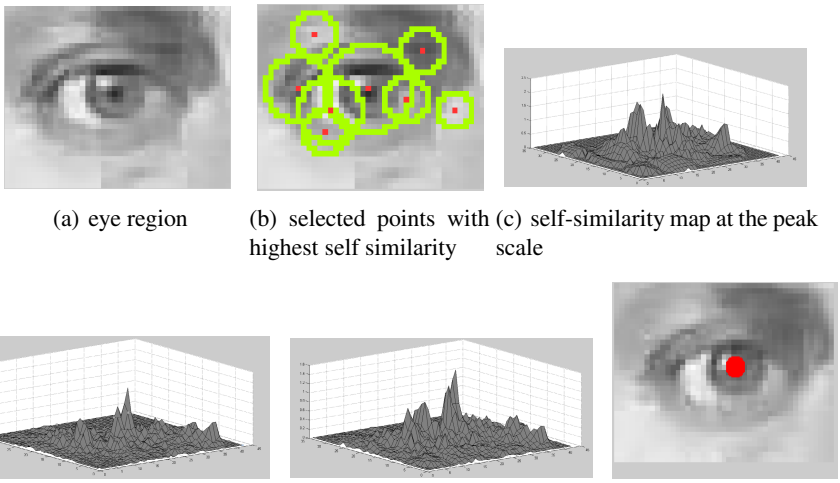
Figure 2 shows an example of how the proposed procedure works; in particular figure 2(a) shows the cropped region of the eye whereas figure 2(b) shows the points with highest self similarity values across all the scales. The resulting self-similarity map at the peak scale is then reported in figure 2(c) and figure 2(d) reports the circular shape distribution space built by using the modified version of the Circular Hough Transform. Finally figure 2(e) shows the joint space obtained by appropriately combining the two data structures. Notice how this joined space is the most suited to easily locate the pupil that is highlighted in figure 2(f) by a red circle.

From section 2 it is quite straightforward to derive that the proposed approach is invariant to rotation, illumination and scale changes. Moreover it works without making the assumption that the image sequences contain only face images.

3 Experimental Results

Experimental evidence of the effectiveness of the method was achieved on challenging benchmark data sets. The MATLAB implementation of the boosted cascade face detector proposed by Viola and Jones [13] with default parameters is used in our experiments, discarding false negatives from the test set.

In the first experimental phase the BioID database [16] is used for testing and in particular the accuracy of the approach in the localization of the pupils was evaluated. The BioID database consists of 1.521 gray-scale images of 23 different subjects and has been taken in different locations and at different times of the day under uncontrolled lighting conditions. Besides non-uniform changes in illumination, the positions of the subjects change both in scale and pose. Furthermore in several examples of the database the subjects are wearing glasses. In some instances the eyes are partially closed, turned away from the camera, or completely hidden by strong highlights on the glasses. Due to these conditions, the BioID database is considered a difficult and realistic database. The size of each image is 384×288 pixels. A ground truth of the left and right eye centers is provided with the database. The *normalized error*, indicating the error obtained by the worse eye estimation, is adopted as the accuracy measure for the found eye locations.



(a) eye region (b) selected points with highest self similarity (c) self-similarity map at the peak scale

(d) Hough space (e) Joint space (f) pupil location

Fig. 2. The outcomes of the proposed approach

This measure is defined in [17] as $e = \frac{max(d_{left}, d_{right})}{w}$ where d_{left} and d_{right} are the euclidean distances between the found left and right eye centers and the ones in the ground truth and w is the euclidean distance between the eyes in the ground truth. In this measure, $e \leq 0.25$ (a quarter of the interocular distance) roughly corresponds to the distance between the eye center and the eye corners, $e \leq 0.1$ corresponds to the range of the iris, and $e \leq 0.05$ to the range of the pupil.

In figure 3 the performances of the proposed approach on the BioID database are reported (blue line) and compared with those obtained using self-similarity (i.e. the point with the maximum value in the saliency map) or circularity (i.e. the point with the maximum value in the accumulation space) information. The graph shows that the combination of the feature related to appearance (that is to say the self-similarity) and to oriented edge location as feasible centers of circumferences (i.e. Modified Circular Hough Transform) allows to increase the localization performance. These results are very encouraging especially when correlated with state-of-the-art methods in the literature. To this end in figure 4 the comparison with some of the most accurate techniques in the literature which use the same dataset and the same performance metric is shown. Looking at the graph it can be seen that only the methods proposed in [8] and [6] provide slightly better results both for $e \leq 0.1$ and $e \leq 0.05$ measures. In general the proposed approach outperforms most of the related methods, even if it does not make use of supervised training or post processing adjustments.

In figure 5 two images of the BioID database processed by the proposed approach are shown. In the one on the left pupils are correctly detected (normalized error 0.0267), whereas in the one on the right some highlights on the glasses mislead the algorithms that miss the detection of the pupil of the left eye (normalized error 0.1158).

To systematically evaluate the robustness of the proposed pupil locator to lighting and pose changes, one subset of the Extended Yale Face Database B [18] is used.

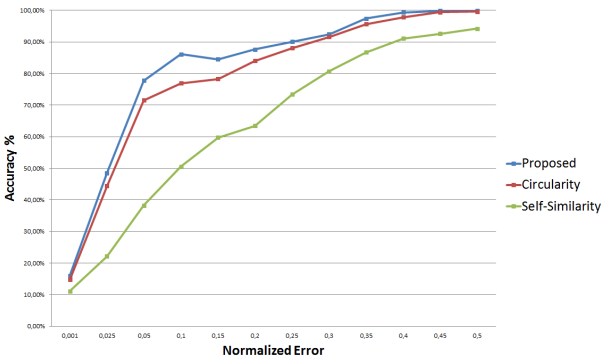


Fig. 3. Results obtained on the BioID database

Fig. 4. Comparison with state-of-the-art methods in the literature on the BioID database

The full database contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions. The size of each image is 640×480 pixels. In particular the system was tested on the 585 images of the subset $\#39B$. The performance in accuracy of the proposed approach on this challenging dataset are $61,66\%$ ($e \leq 0.05$) and $73,16\%$ ($e \leq 0.01$). By analyzing the results, it is possible to note that the system is able to deal with light source directions varying from $\pm 35°$ azimuth and from $\pm 40°$ elevation with respect to the camera axis. The results obtained under these conditions

Fig. 5. Two images of the BioID database processed by the proposed approach (first row) and the the corresponding details on the eye regions (second row). In the image on the left both pupils are correctly detected, whereas in the one on the right some highlights on the glasses mislead the algorithms that miss the detection of the pupil in the left eye.

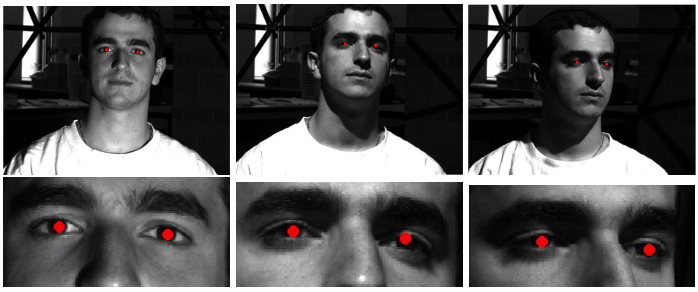


Fig. 6. Some images of the Extended YALE database B in which the approach correctly detects the pupils

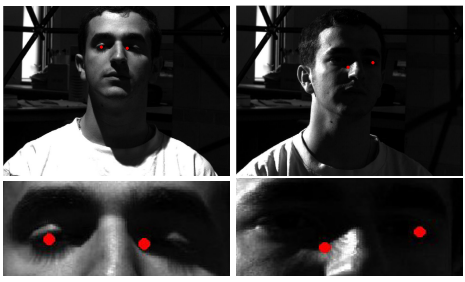


Fig. 7. Some images of the Extended YALE database B in which the detection of the pupils is either less accurate or completely fails

are $77,95\%$ ($e \leq 0.05$) and $84,66\%$($e \leq 0.01$). For higher angles, the method is often successful for the less illuminated eye and sporadically for the most illuminated one: if the eye is uniformly illuminated, the pupil is correctly located, even for low-intensity images. In figure 6 some images of the Extended YALE database B in which the approach correctly detects the pupils even under different lighting conditions and pose changing are shown. In figure 7 some images in which the detection of the pupils is either less accurate or completely fails are instead shown.

4 Conclusions and Future Works

A new method to automatically locate pupils in low-resolution images containing human faces is proposed in this paper. In particular pupils are localized by a two steps procedure: at first self-similarity information is extracted by considering the appearance variability of local regions and then they are combined with an estimator of circular shapes based on a modified version of the Circular Hough Transform. The proposed approach does not require any training phase or decision rules embedding some a priori knowledge about the operating environment. Experimental evidence of the effectiveness of the method was achieved on challenging benchmark data sets. The results obtained are comparable (sometimes outperform) with those obtained by the approaches proposed in literature (making use of training phase and machine learning strategies).

With regard to the computational load, the calculation of the similarity space has a complexity $O(kM^2)$ where k is the number of pixels in the image and M represents the maximal considered scale. The circle detection has instead $O(kn)$ complexity where k is again the number of pixels in the image and n is the dimensionality of the operator used in the convolution implemented by equation 4. However considering that the calculation of the two spaces is embarrassingly parallel (no effort is required to separate the problem into a number of parallel tasks) it is possible to approximate the computational load to the maximum of the two terms above. This therefore leads to a complexity comparable to that of the state of the art methods, however, offering better performance of detection and although not requiring training or other specific post-processing steps that limit their ability to work under various operating conditions.

Future work will address the improvement of the construction of the area of circularity through techniques derived from differential geometry in order to make the system even more accurate.

References

1. Hansen, D.W., Qiang, J.: In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(3), 478–500 (2010)
2. SMI, SensoMotoric Instruments, `http://smivision.com/en/gaze-and-eye-tracking-systems/products/iview-x-hed.html` (retrieved on December 2012)
3. Zhu, Z., Ji, Q.: Robust Real-Time Eye Detection and Tracking under Variable Lighting Conditions and Various face Orientations. Computer Vision and Image Understanding 98(1), 124–154 (2005)
4. Coutinho, F.L., Morimoto, C.H.: Improving Head Movement Tolerance of Cross-Ratio Based Eye Trackers. International Journal of Computer Vision 101(3), 459–481 (2013)
5. Asteriadis, S., Nikolaidis, N., Pitas, I.: Facial feature detection using distance vector fields. Pattern Recognition 42(7), 1388–1398 (2009)
6. Timm, F., Barth, E.: Accurate Eye Centre Localisation by Means of Gradients. In: Proceeding of the International Conference on Computer Vision Theory and Applications, pp. 125–130 (2011)
7. Valenti, R., Gevers, T.: Accurate eye center location and tracking using isophote curvature. In: Proceeding of the IEEE International Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1–8 (2008)

8. Valenti, R., Gevers, T.: Accurate Eye Center Location through Invariant Isocentric Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(9), 1785–1798 (2012)
9. Kroon, B., Hanjalic, A., Maas, S.: Eye localization for face matching: is it always useful and under what conditions? In: Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval, CIVR, Niagara Falls, Canada, pp. 379–388 (2008)
10. Niu, Z., Shan, S., Yan, S., Chen, X., Gao, W.: 2D Cascaded AdaBoost for Eye Localization. In: Proceeding of the 18th International Conference on Pattern Recognition, ICPR, vol. 2, pp. 1216–1219 (2006)
11. Campadelli, P., Lanzarotti, R., Lipori, G.: Precise Eye and Mouth Localization. International Journal of Pattern Recognition and Artificial Intelligence 23(3), 359–377 (2009)
12. Cristinacce, D., Cootes, T., Scott, I.: A Multi-Stage Approach to Facial Feature Detection. In: Proceedings of the British Machine Conference, pp. 231–240 (2004)
13. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57, 137–154 (2004)
14. Prendergast, P.M.: Facial Proportions. Advanced Surgical Facial Rejuvenation, 15–22 (2012)
15. Maver, J.: Self-Similarity and Points of Interest. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(7), 1211–1226 (2010)
16. BioID: Technology Research, the BioID Face Database (2001), `www.bioid.com`
17. Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: Robust face detection using the hausdorff distance. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 90–95. Springer, Heidelberg (2001)
18. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)