Robust Algorithm for Pupil-Glint Vector Detection in a Video-oculography Eyetracking System

Sonia Goñi, Javier Echeto, Arantxa Villanueva, Rafael Cabeza
Dpto. de Ingeniería Eléctrica y Electrónica
Universidad Pública de Navarra, Spain
Email contact: avilla@unavarra.es

Abstract
This paper presents a robust real time algorithm for an eye tracking system employing the well-known bright-pupil technique that performs an effective detection of the pupil and glint positions in the image. The accuracy in the processing is essential if a good determination of the eye gaze is desired. This algorithm is competent for unconstrained images and presents an unmatchable behaviour for users using glasses. The algorithm employs mostly commands from Matrox Imaging Library (MIL) that present a wide sort of functions for image processing and pattern recognition.

1. Introduction

Studies about position of the gaze have for long been a field full of possibilities in multidisciplinary applications [1, 2, 3]. The eye tracking techniques can be classified depending on the degree of invasion respect of the subject under study. Nowadays, the less invasive method in order to carry out the eye tracking detection is the video-oculography (VOG) [1, 3]. This procedure is based in the identification of certain special characteristics of an image that has been obtained with a video camera [1, 4, 5] such as the relative position of the pupil, the limit between the limbus and the sclera, the shape of the pupil, etc.

This kind of algorithm employs infrared lighting [6] coaxially disposed respect to the optical axis of the camera that is fitted under the screen see figure 1.

The obtained images, figure 1, have two special characteristics. The pupil looks brighter than most of the rest of the image and secondly it assures the existence of a bright point corresponding to the glint of the diode in the cornea. The position of both points is directly related with the gaze position. This technique known as bright pupil presents two concrete problems. On one hand, in the situation when there is a lot of external lighting in the scene, the pupil looses its contrast respect the rest of the image, what makes more difficult to determine its boundaries. On the other hand it appears multiple glints when the users use glasses or contact lenses what prevent the algorithm from determining which of them is the corneal reflection.

2. Algorithm description

The aim of the algorithm is to determine both important characteristics using the information recorded by the image acquisition system: that is to say, the centre of the pupil and the corneal glint. This detection must be as robust as possible respect to external lighting changes as well as to other spurious reflections.

The algorithm covers two phases: a preliminary search of both interesting points in the eyes and a subsequent refinement in the calculation of the centre of the pupil and the glint. The first phase leads to a suitable fit of the processing window size, obtaining a notable reduction in calculation time. Let’s describe in detail and separately each of these phases.

2.1 Preliminary Detection

Observing the images taken by the camera, it can be checked that whether the pupil as the glint has certain characteristics that make them be different from the rest of the objects in the image.

- Pupil
  - Average bright level higher than background’s and lower than glint’s.
  - Average area: bigger than glint’s and smaller than background’s.
  - High compactness
- Glint
  - High bright level
Small area
Close to the pupil

Taking into account these properties it’s time to get the estimated location of the pupil and the glint. For this purpose it was used the group of functions existing in the libraries MIL of Matrox for the detection and segmentation of blobs. The algorithm employs at different steps certain configuration values that have been adjusted and experimentally validated for the used system and for a standard working session. Anyway these can be easily fitted whether any of the characteristics of the tracking changes.

As a previous step to this detection, the image is binarized with two pre-established thresholds, wide enough as to cover all the possible situations of illumination in which the system could be used. These values are not critic due to, as we’ll see later, the system has the capacity to vary them dynamically as they are evolved. They are, therefore, two initial values that allow to assure that the limits of the pupil are within such segment and whose default values have been experimentally validated. On the other hand and in the worst case, they could be adjusted at the start of each session.

Once the image is binarized, an open and close operation is carried out [7] with the purpose of eliminating the noise and the possible artefacts present in the image. As result of the two previous steps a group of blobs is obtained, among which it’ll be necessary to distinguish the pupil. As first criterion of selection a filter by area is made. The result will be all those blobs whose sizes were within the two pre-established values (between 600-4000 pixels). The last step is a selection by compactness. Due to its circular shape, the pupil presents a high compactness respect the rest of blobs present in the image. Calculating the compactness index of the remaining blobs, the one corresponding to the pupil is finally obtained (see figure 2). From this blob its mass centre as well as the contact points window, that is the minimum rectangular window containing the pupil, are extracted.

To minimize the calculation time of this first step in the algorithm (that will be proportional to the image area to be processed), the information of the former image is used. More precisely the position of the glint is used, so that a window of 180x135 pixels is centred at the position of the glint of the mentioned image, based on the hypothesis that the pupil can not vary its positions too much from image to image. It has been demonstrated that this size assures that the pupil is found in the searching area allowing a low computation time. Obviously, for the initial image of the session the detection will be made in the whole image acquired by the camera.

Once the estimated position of the pupil is carried out, the glint is searched. For this purpose, first of all the image is binarized with the highest threshold employed in the previous step. The obtained blobs are filtered by area eliminating the elements with an area lower than 5 pixels and higher than 200 pixels. At last, to be able to discern among multiple reflections, the closest to the mass centre of the recently calculated pupil is chosen (see figure 3).

With the purpose to minimize the processing time, an increment of 40 pixels is added to the contact points window calculated for the pupil to perform the glint searching. It has been widely demonstrated that the glint is always found in such area.

2.2 Calculation of thresholds

In order to perform the exact calculation of the two pursued image characteristics, an exhaustive calculation of the threshold values is carried out.

A window is located containing not only the pupil but also the glint to obtain later a histogram from itself as most defined as possible. The contact points of each calculated blob at former steps are used to construct the window leaving a safety margin of 20 pixels to put away
possible approximation errors in previous steps. In this way it’s assured that the pupil and the glint are within the window.

The histogram of this window is extracted obtaining the thresholds that limit the grey levels of the pupil and the glint. In order to locate in this histogram the maximums corresponding to the background and the pupil the threshold for the pupil obtained in the former image is employed. It divides the grey level distribution in two areas clearly different: the right area corresponds with the pupil and the left area with the background and their maximums are identified as pupils and backgrounds peaks respectively.

The threshold of the pupil is to be determined at some value between the maximums previously located. However this determination could result awkward because of the ringing existing between those maximums. In order to palliate as far as possible this limitation, the ringing or intermediate area is estimated and the average level of such uncertainty area is fixed as the threshold of the pupil. The ringing interval is defined as the interval delimited by the grey levels whose values in the histogram are below a certain value. This limit is calculated as the arithmetic average between two histogram values. The first one will correspond to the lowest maximum calculated, i.e. pupils or backgrounds peak. The other value will be the minimum reached by the histogram within the interval between those maximums. As it’s been already commented, the level of the pupil is chosen as the middle point of the ringing interval (see figure 4).

Figure 4. Typical histogram with its associated parameters

Once obtained the threshold of the pupil, the one for the glint is to be got. Whereas the background and the pupil present a very clear peak in the histogram, the glint doesn’t present any characteristic that makes it different because of its low number of pixels. Its only characteristic reflected in the histogram is that its level is higher than pupil’s. In this case, the criterion is simple: it is chosen the grey level with next property: the threshold and the next two levels are, at most, represented each one by two pixels in the image. The first level of those three will be chosen as the threshold of the glint. This method has shown to be effective for glint thresholding with the working conditions of the system.

2.3 Precisely determination of the centres

Once the thresholds of the pupil and the glint have been extracted, the centre of both objects is going to be obtained in a very precisely way. Each of these calculations will be restricted by a window containing only the object in question. In this way the processing time is reduced and the precision is increased.

For the calculation of the centre of the glint, it was employed the window obtained in the preliminary detection enlarged by 10 pixels. The image with the threshold calculated from the former section is binarized and the mass of the resulting blob centre is calculated. This one will be the definitive centre of the glint.

The procedure for the calculation of the centre of the pupil is more complicated. In this case the use of the mass centre is not feasible, due to the fact that the pupil can be strongly deformed by the presence of the glint either in its inside or in its perimeter. To avoid this problem the elliptic shape of the pupil is taking into account, from which we can assume a lineal behaviour in the distribution of the centre of the rows and columns that conform the pupil.

As initial step, it was obtained the pupil blob, binarizing the image with the threshold obtained in the previous section and applying an open and close operation, from the image of the window derived from the enlargement in 10 pixels of the contact points window. For this blob, the centre of each row and column is estimated and a linear adjustment is made for both distributions. Through an iterative method the most disperse centres from the obtained line are eliminated and replaced by the correspondent ones in the line. This method allows the detection of moved images (due to the interfacing of the even and odd fields) as well as eliminating those rows or columns that are in contact with the glint, deforming the pupil and falsifying its mass centre. The intersection of both regression lines will be taken as the centre of the pupil.

Once the centres of the pupil and glint are obtained, it’s possible to determine the direction of the user gaze, whether the whole system has been previously calibrated [8]. It’s important to point out the fact that the information obtained with success from the processing of an image, threshold and centres, is used as initial parameters for the following processing tasks, based on the hypothesis that the parameters involved in the process don’t vary too much from one image to other.
3. Experimental results

It’s been checked the functionality of the described algorithm through the processing of ten recorded sessions with the infrared video-oculography system. The images have been captured with a Hamamatsu C5999 camera and digitalized by a Matrox Meteor card with a resolution of 640x480 (RS-170). The diode used for the lighting has a spectrum centred at 850 nm. The user is located at a distance about 500mm from the screen. The whole system is controlled by a dual processor Pentium at 1.7 GHz with 256 MB of RAM. The development environment has been Visual C++ under Windows 2000 using the libraries MIL of Matrox.

One of the initial requirements for the algorithm is that it works in real-time. Considering that the camera speed is 30 images per second, no more than 30ms are available to process each one. It’s been checked that in the worst case, in which the whole image must be processed, the calculation time doesn’t exceed 20 ms. If it is possible to restrict the processing window using the information of the previous image, as it’s been described in the section before, the time is reduced drastically till 7 ms. In both cases the operation in real time is assured.

In order to check the robustness of the new algorithm its results have been compared with another existing one [9] whose efficiency is contrasted with several years of real use. The obtained results, shown in the table 1, express no doubt of the notable improves and especially in those users who need glasses, sessions G1 to G5. Only in the G5 session the results have been unfavourable due to a medium intensity reflection motivated by the lens of the glasses that appeared superimposed over the pupil, what made practically impossible its detection. From the processing results of these sessions it’s been also checked that the restoration of the new algorithm against tracking loss (inevitable with the user’s blinking) is instantaneous in the next image in which the pupil appears totally.

<table>
<thead>
<tr>
<th># images</th>
<th>Actual algorithm</th>
<th></th>
<th>Proposed algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>session</td>
<td># reject images</td>
<td>T (ms)</td>
<td># reject images</td>
</tr>
<tr>
<td>1</td>
<td>154</td>
<td>107 (69%)</td>
<td>4.35</td>
</tr>
<tr>
<td>2</td>
<td>150</td>
<td>93 (62%)</td>
<td>4.53</td>
</tr>
<tr>
<td>3</td>
<td>114</td>
<td>77 (67%)</td>
<td>3.95</td>
</tr>
<tr>
<td>4</td>
<td>172</td>
<td>54 (31%)</td>
<td>4.29</td>
</tr>
<tr>
<td>5</td>
<td>900</td>
<td>191 (21%)</td>
<td>4.29</td>
</tr>
<tr>
<td>G1</td>
<td>500</td>
<td>492 (98%)</td>
<td>5.28</td>
</tr>
<tr>
<td>G2</td>
<td>417</td>
<td>415 (99%)</td>
<td>6.76</td>
</tr>
<tr>
<td>G3</td>
<td>524</td>
<td>524 (100%)</td>
<td>8.92</td>
</tr>
<tr>
<td>G4</td>
<td>413</td>
<td>413 (100%)</td>
<td>9.35</td>
</tr>
<tr>
<td>G5</td>
<td>401</td>
<td>401 (100%)</td>
<td>7.60</td>
</tr>
</tbody>
</table>

4. Conclusions

In this article it’s been described a new image processing algorithm for the calculation of the position of the user’s pupil and the corneal reflection of the infrared lighting of a video-oculography system. The proposed algorithm rejects a reduced number of images even in the case that the user of the system wears glasses and its recovery against tracking loss is instantaneous. Its functionality has been contrasted with other system, showing its greater robustness and its operation in real-time.

5. References