

A Method for Gaze Fixation Point Localization through Layer by Layer Segmentation and Nine Point Localization

Yuan Xiaoxia^{1,†}, Wang Qingmin^{1,†}, Luo Tailai¹, Liu Qiuhong¹, Li Kehua^{1,*}

¹PLA Naval Medical Center, Naval Medical University, Shanghai, 200433, China

[†]First co-author: These authors contributed equally to this work

^{*}Corresponding author: leekehua@yeah.net

Keywords: Visual Tracking; Localization; Pupil Cornea; Nonlinear Matrix; Layer By Layer Segmentation

Abstract: This article investigates a visual tracking method for fixation point localization. By segmenting the observation area image layer by layer multiple times, nine feature points of each segmentation are selected using a layout centered around eight points, and a nonlinear homogeneous pupil corneal matrix is constructed to solve the local correspondence of the pupil corneal vector in each segmentation area, achieving real-time calculation of the coordinate position of the gaze point on the screen. The position calculation accuracy of this method will not decrease with the increase of the screen area, and it has good application value in the field of line of sight tracking with a large observation area and high measurement accuracy requirements.

1. Introduction

Eye tracking, also known as visual tracking, is a technology that uses mechanical, electronic, optical and other detection methods to obtain the current direction of the subject's visual attention. It is widely used in human-computer interaction, elderly and disabled assistance, vehicle driving, human factor analysis, virtual reality, and military fields [1]. The main methods for line of sight tracking include direct observation [2], post image method [3], mechanical recording method [4], current recording method [5], electromagnetic induction method [6], and optical recording method [7]. In the field of flight, studying the eye movement changes of pilots during flight is of great significance for optimizing the aircraft cockpit, allocating pilot attention, and evaluating and analyzing the physiological and psychological state of pilots. Fitts et al. found that visual tracking technology provides a valuable method for evaluating the importance of new instruments[8], the difficulty of instrument interpretation, and instrument panel layout by studying the eye movement changes of pilots during different flight processes [9].

In order to improve the accuracy of line of sight tracking, this paper proposes a layer by layer quartering method to segment the screen area, and then uses a layout centered around 8 points to select 9 feature points to calculate the local correspondence of each area. This can improve the tracking accuracy by pre sampling and calibrating multi-point data. This method to some extent overcomes the problems of poor global adaptability and low accuracy of the solution results for large screen areas caused by limitations and defects in related technologies. It has good applicability

for fixed areas and fields that require high-precision line of sight tracking, such as in the cockpit of aircraft pilots.

2. Layer by Layer Segmentation and Benchmark Determination Using the Quarter Method

Figure 1 shows the video screen S under the scene camera lens of the visual attention detection device in the cockpit. The method of dividing it into four parts layer by layer is as follows: first, we divided the screen area S into the first layer, evenly and symmetrically dividing it into four parts, namely top, bottom, left, and right, denoted as S_1 , respectively S_2 , S_3 , S_4 ; Next, we selected one of the regions $S_i (i=1,2,3,4)$, in sequence and perform the second layer of segmentation. We divided S_i into four parts, still uniformly and symmetrically distributed in the top, bottom, left, and right directions, denoted as S_{i1} in sequence, S_{i2} , S_{i3} , S_{i4} ; Then, we selected one of the regions S_{ij} in sequence ($i=1,2,3,4; j=1,2,3,4$), and performed the third segmentation, denoted as S_{ij1} in the same way S_{ij2} , S_{ij3} , S_{ij4} . Based on the size of the screen, further refinement and segmentation of the screen can be carried out. Here, we will only use three rounds of segmentation as an example to illustrate.

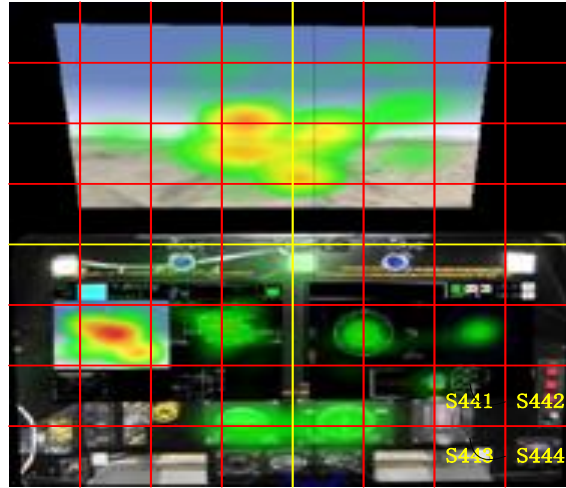


Figure 1: Video screen S .

Targeting a specific area $S_{ijk} (i=1,2,3,4; j=1,2,3,4; k=1,2,3,4)$, we selected 9 calibration reference points, denoted as $A_{ijkm} (i=1,2,3,4; j=1,2,3,4; k=1,2,3,4; m=1,2,3,4,5,6,7,8,9)$. The coordinates of each point are recorded as (a_{ijkmx}, a_{ijkmy}) , where, a_{ijkmx} is the horizontal coordinates of the screen, a_{ijkmy} is the vertical coordinates of the screen. The 9th point is the central position of each region, and the remaining 8 points are evenly spaced around it, as shown in Figure 2.

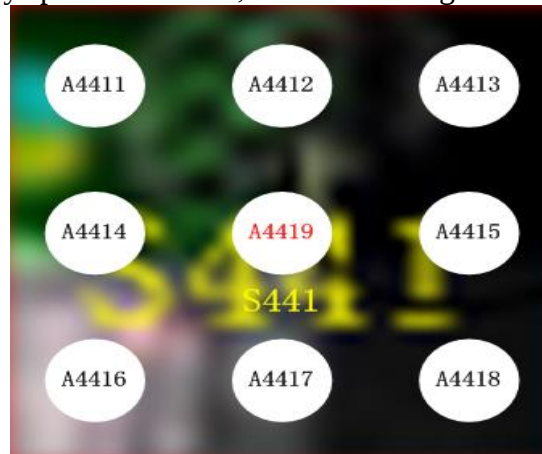


Figure 2: Schematic diagram of selecting basic points in sub regions.

3. Method for Constructing Pupil Corneal Vector Matrix

We selected the 1st to 8th points from the 9 calibration reference points after layer by layer segmentation, and combined the calibration factor vectors for the smallest calibration area. Extract 9 reference points around the subject's gaze on the screen area S_{111} (take S_{111} as an example), the pupil corneal vectors of A_{1111} , A_{1112} , A_{1113} , A_{1114} , A_{1115} , A_{1116} , A_{1117} , A_{1118} , A_{1119} are (x_{e1}, y_{e1}) , (x_{e2}, y_{e2}) , (x_{e3}, y_{e3}) , (x_{e4}, y_{e4}) , (x_{e5}, y_{e5}) , (x_{e6}, y_{e6}) , (x_{e7}, y_{e7}) , (x_{e8}, y_{e8}) , (x_{e9}, y_{e9}) .

We choosed 8 points of all 9 points as (a_{111jx}, a_{111jy}) ($j=1,2,3,4,5,6,7,8$), its corresponding pupil corneal vectors $aer(x_{ei}, y_{ei})$ ($i=1,2,3,4,5,6,7,8$), and constructed corresponding pupil corneal vector nonlinear homogeneous pupil corneal matrix G as follows:

$$G = \begin{bmatrix} 1 & x_{e1} & y_{e1} & \sqrt{x_{e1}y_{e1}} & x_{e1}^{1/3}y_{e1}^{2/3} & x_{e1}^{2/3}y_{e1}^{1/3} & x_{e1}^{2/5}y_{e1}^{3/5} & x_{e1}^{3/5}y_{e1}^{2/5} \\ 1 & x_{e2} & y_{e2} & \sqrt{x_{e2}y_{e2}} & x_{e2}^{1/3}y_{e2}^{2/3} & x_{e2}^{2/3}y_{e2}^{1/3} & x_{e2}^{2/5}y_{e2}^{3/5} & x_{e2}^{3/5}y_{e2}^{2/5} \\ 1 & x_{e3} & y_{e3} & \sqrt{x_{e3}y_{e3}} & x_{e3}^{1/3}y_{e3}^{2/3} & x_{e3}^{2/3}y_{e3}^{1/3} & x_{e3}^{2/5}y_{e3}^{3/5} & x_{e3}^{3/5}y_{e3}^{2/5} \\ 1 & x_{e4} & y_{e4} & \sqrt{x_{e4}y_{e4}} & x_{e4}^{1/3}y_{e4}^{2/3} & x_{e4}^{2/3}y_{e4}^{1/3} & x_{e4}^{2/5}y_{e4}^{3/5} & x_{e4}^{3/5}y_{e4}^{2/5} \\ 1 & x_{e5} & y_{e5} & \sqrt{x_{e5}y_{e5}} & x_{e5}^{1/3}y_{e5}^{2/3} & x_{e5}^{2/3}y_{e5}^{1/3} & x_{e5}^{2/5}y_{e5}^{3/5} & x_{e5}^{3/5}y_{e5}^{2/5} \\ 1 & x_{e6} & y_{e6} & \sqrt{x_{e6}y_{e6}} & x_{e6}^{1/3}y_{e6}^{2/3} & x_{e6}^{2/3}y_{e6}^{1/3} & x_{e6}^{2/5}y_{e6}^{3/5} & x_{e6}^{3/5}y_{e6}^{2/5} \\ 1 & x_{e7} & y_{e7} & \sqrt{x_{e7}y_{e7}} & x_{e7}^{1/3}y_{e7}^{2/3} & x_{e7}^{2/3}y_{e7}^{1/3} & x_{e7}^{2/5}y_{e7}^{3/5} & x_{e7}^{3/5}y_{e7}^{2/5} \\ 1 & x_{e8} & y_{e8} & \sqrt{x_{e8}y_{e8}} & x_{e8}^{1/3}y_{e8}^{2/3} & x_{e8}^{2/3}y_{e8}^{1/3} & x_{e8}^{2/5}y_{e8}^{3/5} & x_{e8}^{3/5}y_{e8}^{2/5} \end{bmatrix}$$

The characteristic of the above matrix G is that, except for the first column which is 1, the other 7 columns are all formal $x_{ei}^p y_{ei}^q$, and $p > 0$, $q > 0$ and $p+q=1$. When the value of p and q is large, it will result in uneven calculation results and poor data correlation, which is a major advantage of constructing a nonlinear homogeneous pupil corneal matrix G .

4. Calculation of Calibration Factor Vector

The calibration factor vector calculation establishes a mapping relationship from the pupil corneal vector to the fixation point coordinate position, based on the position coordinates of 8 points on the screen (a_{111jx}, a_{111jy}) ($j=1,2,3,4,5,6,7,8$), construct the screen coordinate vector D_x , D_y , where D_x , D_y is as follows:

$$D_x = [a_{1111x} \ a_{1112x} \ a_{1113x} \ a_{1114x} \ a_{1115x} \ a_{1116x} \ a_{1117x} \ a_{1118x}]^T,$$

$$D_y = [a_{1111y} \ a_{1112y} \ a_{1113y} \ a_{1114y} \ a_{1115y} \ a_{1116y} \ a_{1117y} \ a_{1118y}]^T$$

We solved the inverse matrix G^{-1} of the nonlinear homogeneous pupil corneal matrix G , and calculated the calibration factor vector E_{1111x} , E_{1111y} according to $E_{1111x} = G^{-1}D_x$, $E_{1111y} = G^{-1}D_y$. And then, we selected 8 out of 9 points (a_{111jx}, a_{111jy}) ($j=2,3,4,5,6,7,8,9$), and calculated the calibration factor vector E_{111jx} , E_{111jy} , using other combinations in sequence. Due to the selection of 9 combinations of 8 out of 9 points, 9 sets of calibration factor vectors E_{111jx} , E_{111jy} can ultimately be solved, Where $j=1,2,3,4,5,6,7,8,9$. Select the weighted average as the calibration factor vector S_{111} for region E_{111x} , E_{111y} . The calculation formula is as follows:

$$E_{111x} = \frac{1}{2}E_{1111x} + \frac{1}{16} \sum_{j=1}^8 E_{111jx}$$

$$E_{111y} = \frac{1}{2}E_{1111y} + \frac{1}{16} \sum_{j=1}^8 E_{111jy}$$

We obtained the final global calibration factor vector E_{ijkx} , $E_{ijk y}(i=1,2,3,4;j=1,2,3,4;k=1,2,3,4)$ through to solve the segmentation regions of each layer using the above method.

5. Mean Calculation of Pupil Corneal Vector

The mean of pupil corneal vector is calculated by using the mean of 9 points in the minimum segmentation screen sub area for the subject's pupil corneal vector obtained from eye movement sampling. Taking the screen area S_{112} , nine points are A_{1121} , A_{1122} , A_{1123} , A_{1124} , A_{1125} , A_{1126} , A_{1127} , A_{1128} , A_{1129} , and their corresponding gaze pupil corneal vectors are denoted as (x_{e1}, y_{e1}) , (x_{e2}, y_{e2}) , (x_{e3}, y_{e3}) , (x_{e4}, y_{e4}) , (x_{e5}, y_{e5}) , (x_{e6}, y_{e6}) , (x_{e7}, y_{e7}) , (x_{e8}, y_{e8}) , (x_{e9}, y_{e9}) . we find the mean corneal vector of the pupil in this area, denoted as (x_{112}, y_{112}) . The calculation formula is as follows:

$$x_{112} = \sum_{i=1}^9 x_{ei} / 9, \quad y_{112} = \sum_{i=1}^9 y_{ei} / 9$$

Similarly, the pupil corneal vectors of the other three regions out of the four sub regions can be obtained as S_{111} , S_{113} , S_{114} .

To calculate the pupil corneal vector in the layer above the minimum region, the mean of four segmentation regions is used to solve. The average pupil corneal vector of S_{11} is (x_{11}, y_{11}) , The calculation formula is as follows:

$$x_{11} = \sum_{i=1}^4 x_{11i} / 4, \quad y_{11} = \sum_{i=1}^4 y_{11i} / 4$$

By layer by layer recursion, the pupil corneal vectors of each segmented region can be obtained.

6. Real Time Line of Sight Tracking and Positioning Solution

In the real-time tracking process of the subject's focus on the screen area, the eye tracking device takes real-time photos and samples the local area of the human eye. After filtering, the real-time pupil corneal vector is extracted, denoted as (x, y) . This vector first relates to the four regions S_1 in the first layer S_2 , S_3 , and S_4 perform the first layer judgment, calculate the first layer pupil corneal error vector, and record them as e_1 , respectively e_2 , e_3 , e_4 . The calculation formula is:

$$e_1 = (x - x_1)^2 + (y - y_1)^2; e_2 = (x - x_2)^2 + (y - y_2)^2; \\ e_3 = (x - x_3)^2 + (y - y_3)^2; e_4 = (x - x_4)^2 + (y - y_4)^2;$$

We selected the smallest value by comparing e_1 , e_2 , e_3 , and e_4 . Without loss of generality, we solved the second layer of pupil corneal error vector for the four sub regions of S_2 corresponding to e_2 , assuming the smallest one is e_2 . And then, we selected the region with the smallest error to solve the third layer of pupil corneal error vector. Assuming the smallest one is e_{ijk} , the calibration factor vector E_{ijkx} , $E_{ijk y}(i=1,2,3,4)$ for the region S_{ijk} is ultimately selected and used to solve for real-time fixation point coordinates. The calculation formula for gaze point (a_x, a_y) is as follows:

$$a_x = E_{ijkx}(1) + E_{ijkx}(2)x_e + E_{ijkx}(3)y_e + E_{ijkx}(4)\sqrt{x_e y_e} + E_{ijkx}(5)x_e^{1/3}y_e^{2/3} \\ + E_{ijkx}(6)x_e^{2/3}y_e^{1/3} + E_{ijkx}(7)x_e^{2/5}y_e^{3/5} + E_{ijkx}(8)x_e^{3/5}y_e^{2/5} \\ a_y = E_{ijk y}(1) + E_{ijk y}(2)x_e + E_{ijk y}(3)y_e + E_{ijk y}(4)\sqrt{x_e y_e} + E_{ijk y}(5)x_e^{1/3}y_e^{2/3} \\ + E_{ijk y}(6)x_e^{2/3}y_e^{1/3} + E_{ijk y}(7)x_e^{2/5}y_e^{3/5} + E_{ijk y}(8)x_e^{3/5}y_e^{2/5}$$

The three-layer screen segmentation methods listed above can be similarly extended to 4, 5, or more layers. The real-time and accuracy requirements for line of sight tracking positioning are contradictory. Although the expansion of the layer by layer segmentation method mentioned above will increase the workload of selecting and calibrating calibration points, as the computing power of the hardware continues to grow, higher accuracy requirements can be achieved through the improvement of algorithm capabilities, while meeting real-time requirements.

7. Conclusion

The gaze localization algorithm is a core issue in the field of visual tracking. The visual tracking gaze localization method proposed in this paper uses 9 points and 18 parameter forms to solve the problem, which is more detailed than the traditional 9 points and 12 parameter forms to describe the non-linear mapping relationship between the local image of the human eye and the screen coordinates. Compared to the traditional method of unified calibration for the entire region, this paper adopts a layer by layer four division method for gaze screen segmentation, which has higher accuracy. Although its calibration process is complex, once pre calibrated, the accuracy of the results obtained will be very high. In practical usage scenarios, it is necessary to consider the selection of screen segmentation layers based on accuracy requirements, observation screen size, and computational hardware computing power intensity. This method is particularly suitable for applications where the screen area is large and the observation areas of interest are numerous and dense, or where accuracy is particularly high.

References

- [1] Hou Z Q, Han C Z. A survey of visual tracking [J]. *Acta automatica sinica*, 2006, 32(4): 603-617.
- [2] Ananos E. Eye tracker technology in elderly people: how integrated television content is paid attention to and processed [J]. *Comunicar*, 2015, 23(45): 75-83.
- [3] Kothari R S, Bailey R J, Kanan C, et al. Ell seg-gen, towards domain generalization for head-mounted eyetracking [C]// *Proceedings of the ACM on Human-Computer Interaction*. Rochester: Rochester institute of technology, 2022, 6: 1-17.
- [4] Tang L Z. Research on gaze tracking method based on wearable eye tracker [D]. Lanzhou: Lanzhou university, 2023.
- [5] Jiang T T, Wu Q, Xu Y P, et al. The application of eye tracking technology in information behavior research[J]. *Journal of the China Society for Scientific and Technical Information*, 2020, 39(2): 217-230.
- [6] Miles W. The peep-hole method for observing eye movements in reading [J]. *Journal of general psychology*, 1928, 1(2): 373-374.
- [7] Han Y C. The development history of eye tracking devices and eye tracking experimental methods [J]. *Psychological Science*, 2000, 23(4): 454-458.
- [8] Huang S, Hu Y, Li C. A CORBA-based computer support cooperative work for dynamic alliances [J]. *International journal of advanced manufacturing technology*, 2002, 19(10): 752-755.
- [9] Fitts P M, Jones R E, Milton J L. Eye movements of aircraft pilots during instrument landing approaches [J]. *aeronautical engineering review*, 2005, 18(24): 345-367.