Research and Engineering Implementation of Automatic Fundus Photography Algorithm for Ultra-Wide Angle Confocal Laser Fundus Scanning System

Haibo Ding [*], Feng Wang

Department of Biological Science & Medical Engineering, Southeast University, Nanjing, China

Abstract: To achieve automatic photo capture of fundus images in an ultra-wide-angle laser-focused fundus scanning system, a pupil detection and positioning algorithm based on confocal laser technology was proposed and deployed. First, image preprocessing methods such as automatic laser intensity enhancement and contrast adjustment were applied to reduce the impact of noise and interference on subsequent processing. Next, the pupil was quickly and accurately located using the proposed pupil detection and positioning algorithm based on confocal laser technology. Finally, the movement information was calculated based on the position of the pupil center and image center, and the three-axis stepper motor was controlled in a rapid closed-loop manner to achieve three-dimensional automatic tracking of the pupil. Once the shooting conditions were met, the photo was automatically captured. To validate the effectiveness and timeliness of this solution, automatic photo capture tests were conducted on different subjects using a prototype. The success rate of automatic photo capture was 95.6%, with an average time of about 8 seconds per single-eye capture. The fundus images captured by successful automatic photos all met the requirements for automatic capture.

1 Introduction

The fundus is the inner membrane of the eyeball, referring to the inner and posterior tissues of the eyeball, which mainly include the retina, macula, optic disc, and central retinal vein. It is the only part of the human body where blood vessels can be observed with the naked eye under non-invasive conditions [1]. Fundus imaging is a medical imaging technique used to assess the structure and pathological changes of the fundus, providing valuable insights into the health of the eyes and even the whole body, as many diseases leave traces in the fundus. For example, the cup-to-disc ratio from fundus images can suggest whether glaucoma screening should be conducted. Microvascular changes, hemorrhages, and exudates are common ocular complications in diabetic patients [2].

There are generally two types of fundus imaging techniques: (1) traditional optical-based fundus imaging, where white light is used as the light source to illuminate the retina and obtain static fundus images, which results in low patient comfort and poor image clarity and limited imaging range; (2) confocal laser-based fundus imaging, which uses a weak laser to scan the retina point by point, collecting light from the confocal plane to acquire images. This method offers better patient comfort and produces high-quality, dynamic images with a wide field of view [3-4]. In practical applications, fundus image capturing can be either manual or automatic. The manual method requires high cooperation from the patient and skilled medical personnel, while the automatic method involves detecting the pupil's position, calculating the offset, and controlling the motor to quickly align the pupil, offering advantages such as faster speed, simpler operation, and stable image quality.

Pupil detection is key to automation, and the accuracy of detection directly affects the precision of automation. Current pupil detection methods can be categorized into two main types: learning-based methods and non-learning-based methods. Learning-based methods [5-6] use supervised machine learning to precisely locate the pupil, offering high accuracy but at a high cost, requiring large amounts of labeled data and time, and being difficult to implement. Non-learning-based methods [7-8], such as Hough transform for circle detection, morphological detection, and ellipse fitting, rely on high-quality images to achieve high precision. However, these methods are simpler, cost-effective, and robust to noise.

To address the cost, speed, robustness, and accuracy requirements for automatic fundus photography in an ultra-wide angle laser confocal fundus scanning system, this paper proposes a pupil detection algorithm based on confocal laser technology and applies it to achieve automatic fundus photography in such systems. The system completes the task using a monocular camera, rapidly and accurately locating the pupil, calculating the offset, and controlling a three-axis motor to drive the optical system for automatic alignment and movement to the working distance, until the pupil region fills the entire image, triggering automatic photography. The research and engineering implementation of the automatic

---

916370799@qq.com

photography algorithm allow the system to be applied not only in specific examinations but also for rapid screening in health checkup centers or remote areas, reducing human factors during image capture, simplifying the process, and improving efficiency.

2 System Components

The system mainly consists of the eye frame, optical system, and three-axis stepper motor. The eye frame is used to support the subject's eyes, thereby minimizing the impact of head movement. Compared with the commonly used head frame in similar products, the eye frame better stabilizes the position of both eyes, greatly simplifying the subsequent automatic photography process. The optical system is used to capture fundus images. The acquired images are processed to obtain movement information, which is then used to control the movement of the three-axis motor. This drives the optical system to align the pupil, forming a closed loop, and finally captures the fundus image.

3 Algorithm Research and Engineering Implementation

3.1 Precise Pupil Location

The system extracts a single-frame image collected by the fundus camera and analyzes it using an automatic pupil alignment algorithm. First, the region containing valid information needs to be extracted. The original image captured by the system is a grayscale image of 2048×2076 pixels, but the actual valid information is confined to a circular region displayed on the interface. The ultimate goal of the automatic pupil alignment is to move the pupil region into this valid information area and fill it completely. The mask operation not only preserves the information within the valid region but also sets the pixel values outside the region to zero, thus eliminating some interference factors (such as the consending lens reflections at the corners of the image).

Next, the image contrast needs to be enhanced. When the subject has vitreous opacities or when the laser intensity is too low, the laser intensity entering the eye is insufficient, resulting in a dark image where the pupil region and the background are difficult to distinguish. To address this, histogram equalization is applied to enhance the contrast of the image, which helps average the distribution of pixel intensity values. This enhancement is beneficial for subsequent processing, and the specific steps are as follows:

Algorithm: Histogram Equalization
Input: image (grayscale image)

Output: enhanced_image (enhanced grayscale image) 1. Compute the histogram of the input image:
   - Create an array hist[] with 256 bins (for pixel values from 0 to 255).
   - For each pixel in the image:

    - Increment the corresponding bin in hist[] based on the pixel value.
2. Compute the cumulative distribution function (CDF):
  - Initialize CDF[] as an array of size 256.
  - Set CDF[0] = hist[0] (first bin).
  - For i = 1 to 255:
    -CDF[i] = CDF[i - 1] + hist[i] (cumulative sum of histogram values).
3. Normalize the CDF:
  - Find the total number of pixels N = width * height in the image.
  -For i = 0 to 255:
    -CDF_normalized[i] = round((CDF[i] - CDF[0]) / (N - CDF[0]) * 255)
4. Map the original pixel values to the equalized values:
  -For each pixel in the image:
    -Let original_pixel_value = image[x, y]
    -Set enhanced_image[x, y]= CDF_normalized[original_pixel_value]
5. Return the enhanced image.

Next, the enhanced image is binarized, and all connected components are identified. Since the pupil region in the enhanced image has a higher grayscale value, a threshold value Gv is selected, and each pixel is compared to this threshold to obtain the binarized image. All connected components in the binarized image are then identified. The binarization process is described by equation (1), where G denotes the grayscale value of the point (i, j) in the image:

$$Gij = \begin{cases} 0 & Gij \leq Gv \\ 255 & GIJ > Gv \end{cases} \quad (1)$$

Finally, based on the anatomical characteristics of the pupil, the pupil region is selected from the numerous connected components. Unlike other interfering regions which may exhibit irregular shapes due to environmental noise, the pupil area is approximated as a circle based on its distinct anatomical characteristics. Let a and b be the length and width of the bounding rectangle of a connected component, respectively. When the following two conditions are satisfied, the connected component is considered to meet the shape requirements of the pupil region:

1.a/b≈1

2.ab≈(bπ²)

Since the pupil area in the fundus image is approximately circular but not perfect circle, these conditions are approximated to increase tolerance. These two conditions help eliminate connected components that differ significantly from the pupil shape. The remaining connected components that satisfy these two shape conditions undergo the next area constraint. When the fundus camera is farther from the pupil, the pupil area is smaller and more likely to be interfered with by small circular light spots. However, when the camera is closer to the pupil, this interference is less likely, as the pupil area occupies most of the fundus image. Let S be the area of the connected component, and Smin be the minimum real

pupil area at a distance. When the following condition is met, the connected component is considered to meet the area requirements for the pupil region:

3. S≥Smin

Condition (3) helps eliminate interference from small circular light spots at longer distances. The remaining connected components that satisfy the area condition are then subject to the next grayscale value constraint. After applying the three aforementioned constraints, the connected component with the highest average grayscale value is considered the pupil region. Let G be the largest average grayscale value among the remaining connected components, which is preliminarily considered the average grayscale value of the pupil region, and G1 and G2 be the lower and upper bounds for the average grayscale value, respectively. When the following condition is met, the connected component is identified as the pupil region:

4. G1≤G≤G2

If the largest average grayscale value G is still lower than the lower bound G1, it is determined that the pupil is absent in this frame. If G exceeds the upper bound G2, it indicates overexposure, and normal pupil detection cannot be performed. Therefore, only the connected components that meet the average grayscale value criteria are finally identified as the pupil region, and subsequent operations are performed. Figure 1 shows the processes in this section, including image preprocessing and pupil detection.
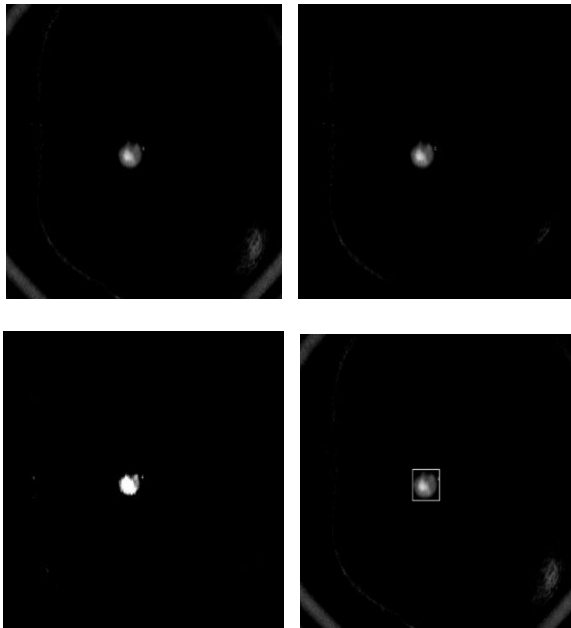

Figure 1. Pupil Detection Process

3.2 Pupil Alignment

After determining the pupil position, the pupil displacement information and pupil filling ratio are calculated, and the three-axis stepper motors are controlled to drive the optical system for automatic pupil alignment. The center coordinates of the image acquisition area are (Xcenter, Zcenter) = (1024, 1088), and the center coordinates of the pupil region are (Xpupil, Zpupil). The

radius of the circular image region extracted by the mask is R. The displacement ratios in the X and Z axes and the pupil filling ratio are calculated using equations (2), (3), and (4).

$$X_{delay} = \frac{\left| X_{pupil} - X_{center} \right|}{R} \qquad (2)$$

$$Z_{delay} = \frac{\left| Z_{pupil} - Z_{center} \right|}{R} \qquad (3)$$

$$Fill_{ratio} = \frac{S_{pupil}}{\pi R^2} \qquad (4)$$

Xdelay and Zdelay represent the displacement ratios of the pupil region center relative to the image acquisition area center along the X and Z axes. When both are equal to 0, it indicates that the pupil region center completely overlaps with the image acquisition area center. A displacement threshold of 5% is set in this paper, meaning that if the displacement ratio is less than 5%, no movement is required in that direction. If it is greater than 5%, the movement direction (positive or negative) is determined based on the pupil center's position relative to the image acquisition area center. The **Fill ratio** represents the ratio of the pupil region area to the circular image region area. A value of 1 means that the pupil region completely fills the circular image region. A filling ratio threshold of 98% is set in this paper, meaning that if the pupil filling ratio exceeds 98%, the optical system has reached the working distance (Y-axis), pupil alignment is complete, and the next step can be performed.

The movement of the optical system is driven by three-axis stepper motors, all of which are stepper motors. A stepper motor converts electrical pulse signals into corresponding angular or linear displacements. For each pulse input, the motor rotates by an angle ω. Therefore, the motor requires 360/ω pulses to complete one full rotation. The lead distance L of the motor refers to the length of the motor shaft's rotation per turn, i.e., the linear displacement produced by one full rotation. Hence, the linear displacement per pulse is l=L·ω/360

The motor operates in two modes: **speed mode** and **position mode**. In speed mode, the motor continues rotating at a set speed until a new speed command is received. In position mode, the motor accelerates, then moves at a constant speed, and finally decelerates to accurately stop at the set position. Since pupil alignment requires alignment along all three axes, and during Y-axis movement, the actual distance represented by each pixel in the image acquisition area decreases, resulting in varying ratios between image displacement and actual displacement, the stepper motors operate in speed mode.

The closed-loop control of the three-axis stepper motors is key to pupil alignment. However, due to hardware limitations, there is a two-frame delay between the image sent to the algorithm and the actual captured image, which affects closed-loop control. Therefore, the pupil alignment process is divided into two stages: **high-speed stage** and **low-speed stage**. In the high-speed stage, the stepper motors move at a higher speed, primarily driving the optical system to move most of the Y-axis travel and approach the working position, while also adjusting the XZ axes to bring the pupil region closer to

the circular image region area center. In the low-speed stage, the stepper motors move at a slower speed to minimize the impact of the frame of delay on closed-loop control and precise pupil alignment. The goal is to adjust the XZ axes so that the pupil region is located at the center of the circular image region area, while the Y-axis moves to the working position, ensuring the pupil region completely fills the circular image acquisition area, and the conditions for automatic photo capture are met. The operational process is shown in the figure 2 and the fundus image captured automatically is shown in the figure 3:
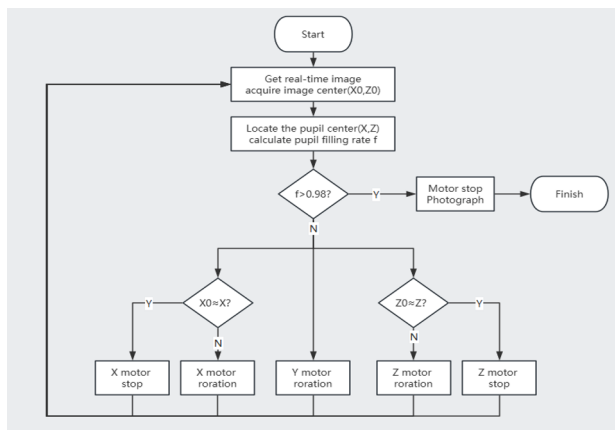

Figure 2. Overall Process Flow of Automatic Photography.

Figure 3. The fundus image captured automatically is shown in the figure.

4 Experimental Result

To verify the success rate, speed advantage, and repeatability of automatic photo capture based on the method presented in this paper, experiments were conducted on a total of 15 subjects, with each subject forming a group. For each subject, one manual photo capture was performed by a professional and three automatic photo captures were performed consecutively, ensuring that the pupil position remained constant during the process. The time spent for the capture, the XYZ positions of the motors at the end of the automatic capture, and whether the automatic capture was successful were also recorded.

In the 45 automatic captures performed during the experiment, 43 were successful and 2 failed, resulting in an automatic photo capture success rate of 95.6%. Figure 4 shows the average time spent on the three automatic photo captures for each group and the time spent on the

manual photo capture. Let Xpo1 is the position of the X-axis motor after the first automatic photo capture, Xpo2 and Xpo3 are the positions after the second and third automatic captures, and Xpo is the position of the X-axis motor after the manual photo capture, the average deviation of X motor in this experimental group is (Xpo1+Xpo2+Xpo3-3*Xpo)/3. Figure 5 shows the average deviation of the XYZ motors at the end of the three automatic photo captures compared to the manual photo capture for each group。From figure 4, it can be seen that the average time spent on automatic photo capture for each group is less than that of manual capture, indicating a clear speed advantage of the automatic photo capture based on the method in this paper. From figure 5, it can be observed that although the average deviation in motor positions at the end of the automatic and manual photo captures shows some fluctuations, the majority of deviations are smaller than 15000 and bigger than 3000, approximately 1 cm (approximately 1/50 of the three-axis length), indicating that the automatic photo capture based on this method has good repeatability.
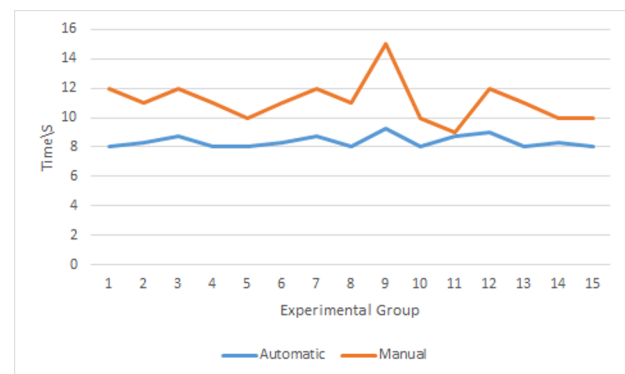

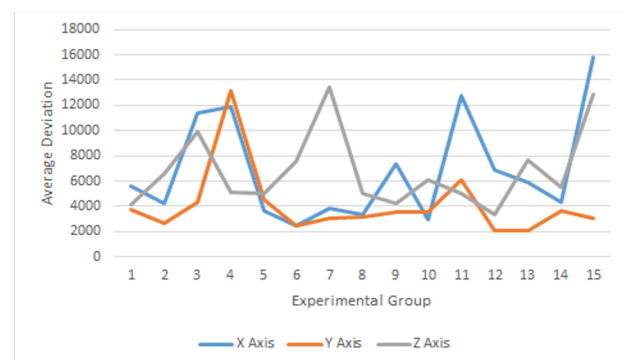**Figure 4.** Automatic and Manual Photography Time Consumption Experiment.

Figure 5. Automatic Photography Reproducibility Experiment.

5 Conclusion

This paper proposes a pupil detection algorithm based on confocal laser technology and applies this algorithm to achieve automatic photo capture of fundus images in an ultra-wide-angle laser-focusing fundus scanning system. The system completes the entire task using a monocular camera, quickly and accurately locates the pupil, calculates the offset information, and controls the three-

axis stepper motor to automatically align the optical unit with the pupil and move it to the working distance until the pupil area fills the entire image, at which point the fundus image is automatically captured. Experiments demonstrate that the method proposed in this paper can achieve automatic photo capture quickly, accurately, and with high repeatability. The success rate is 95.56%, with the time spent on each automatic capture being approximately 8-10 seconds. Additionally, the repeatability of multiple captures is good, with an average error of only 7000. From the experimental results, it can be concluded that this system has high success rates, fast speed, and good repeatability across multiple operations. The retention of manual operation allows the system to perform more detailed checks, addressing a small number of special requirements. The implementation of the automatic photography function allows even operators with limited experience to quickly master the operation process, meeting the majority of needs. It simplifies the photography workflow and improves efficiency. This makes CSLO not only suitable for detailed examinations of specific diseases but also applicable to large-scale retinal screening in populations. Furthermore, it facilitates the promotion and use of CSLO in areas with limited medical resources.

Certainly, there are several areas for future work to improve both accuracy and efficiency. For example, conducting more clinical validations of the algorithm in hospitals would enable it to accommodate a wider range of ophthalmic diseases. Additionally, exploring more effective three-axis motor control strategies could help reduce the impact of delay frames on speed.

Reference

1. R.P. Crick, P.T. Khaw, World Scientific (2003).
2. Shahina Parveen M., S. Hiremath, Springer, **519**, 579-588 (2023).
3. G. Lim, V. Bellemo, Y. Xie, Eye and Vision, **7**, 1-13 (2020).
4. L.P. Aiello, I. Odia, A.R. Glassman, A.R. Galssman, M. Melia, L.M. Jampol, N. Bressler, S. kiss, P.S. Silva, C.C. Wykoff, J.K. Sun, JAMA Ophthalmology, **137**, 65-73 (2019).
5. T. Yan, W. Zhen, H.J. Zhao, Chinese Journal of Liquid Crystal & Displays, **32**, 6 (2017)
6. C. Akinlar, H.K. Kucukkartal, C. Topal, Elsevier, **188**, 116004 (2022).
7. H.F. Wang, J.Z. Wang, K.M. Bai, S. Zhang, Transactions of Beijing Institute of Technology, **40**, 1111-1118 (2020)
8. P. Bonteanu, R.G. Bozomitu, A. Cracan, G. Bonteanu, IEEE 26th International Symposium for Design and Technology in Electronic Packaging, 98-101 (2020)