

## Research Article

# A Rapid Temporal Bone Localization Method Based on Machine Visual Detection Markers

Li Jian-jun <sup>1</sup>, Zhuo Jian-ye,<sup>1</sup> Wu Yue,<sup>1</sup> Wang Zuo,<sup>2</sup> Yao Zhen,<sup>1</sup> Jia Huan,<sup>3</sup> Wang Zhao-yan,<sup>3</sup> Chen Ying,<sup>3</sup> Zhou Dao-min,<sup>4</sup> and Yu Ming-zhou<sup>1</sup>

<sup>1</sup>China Jiliang University, Hangzhou 310018, China

<sup>2</sup>Maanshan University, Maanshan 243100, China

<sup>3</sup>Shanghai Ninth People's Hospital, Shanghai 200125, China

<sup>4</sup>Zhejiang Key Laboratory of Neuroelectronic and Brain Computer Interface Technology, Hangzhou, China

Correspondence should be addressed to Li Jian-jun; 13a0104121@cjlu.edu.cn

Received 25 July 2022; Revised 9 September 2022; Accepted 12 September 2022; Published 23 September 2022

Academic Editor: Lianhui Li

Copyright © 2022 Li Jian-jun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The number of hearing-impaired people is increasing year by year; robotic cochlear drilling surgery is one of the safest methods to treat deafness. Looking at the issue of low efficiency of temporal bone posture positioning in cochlear implantation robotic drilling, a novel auxiliary ring marker temporal bone positioning method was proposed to improve temporal bone posture positioning efficiency, optimize the operation time, and reduce auxiliary injuries caused by the surgery. First, the temporal bone visual positioning assistant ring was designed based on the requirements for cochlear robotic drilling surgery. The target detection was conducted on the auxiliary ring and image processing and feature point extraction methods were designed. Then, the three-dimensional coordinates of the measured feature points were obtained by binocular vision, and the auxiliary ring and temporal bone postures were estimated. Finally, the auxiliary ring and temporal bone localization methods were validated. The experiment results indicated that the temporal bone was located quickly and effectively in a total time of about 33 ms, which was faster and more accurate than traditional visual localization methods and could satisfy real-time temporal bone localization during surgery. This study can reduce the time of temporal bone visual positioning in cochlear implant drilling operations, greatly improving the robot's capabilities to extract visual information during the operation, which has a better auxiliary role for future research and applications of the cochlear implant drilling operation.

## 1. Introduction

Cochlear implant drilling is a new surgical procedure that optimizes the surgical method and reduces surgical trauma [1, 2]. As robot technology has developed, robot drilling surgery has gradually become more acceptable [3, 4]. Compared with human-performed operations, robotic operations for cochlear drilling have shown far more advantages than the former, which is conducive to an effective, quick, and safe methodology for cochlear drilling [5]. Human beings have innate advantages in visual information perception and can quickly extract and identify information and content during surgery; however, there is no perfect computer model in this field, and the weak visual

information processing ability is the key to the challenges facing robotic surgery for cochlear implant drilling [6]. In recent years, temporal bone visual localization as the core technology for robotic drilling surgery has garnered an increasing amount of attention. The goal is to map the positions of key tissue structures in the ear and plan the drilling path of surgery by using the postures of human temporal bones [7].

Currently, many scholars are studying the rapid temporal bone localization method, but most of them still abide by the traditional image detection marker method. Cho et al. proposed registering the locations of surgical wounds using a tripod visual calibration rod as a marker [8]. The calibration rod could provide reliable geometric feature

information and facilitate calculating wound location, but the registration method for implanting the calibration rod was cumbersome and would easily affect the operating space. Dillon et al. used titanium screw implantation on temporal bones to simplify the challenges inherent to marker implantation [9], but the locations where the titanium screws were implanted were random and riddled with uncertainties, and it is difficult for computers to obtain such random information. Jia et al. proposed a short-flow visual registration method of the malleus, which can effectively improve registration accuracy for intra-aural structures and reduce any damage caused by the registration process [10], but this method requires a surgeon to have a lot of clinical experience.

In conclusion, the temporal bone localization markers used now cannot satisfy the requirements for rapid detection in computer vision. To solve the aforementioned problems, this paper proposed a temporal bone localization method based on an auxiliary ring, named the Deep (DL-M), and functions based on a combination of deep learning target detection, computer vision, and medical requirements of cochlear implant drilling surgery. The DL-M reduces extraction time for feature points and matching calculation of irrelevant features through image processing. Compared with BM, SGBM, and other methods, DL-M is faster, and its overall duration is about 40 ms, which meets the requirement for real-time detection of more than 20FPS. The average detection accuracy of the auxiliary ring is less than  $\pm 0.63^\circ$ .

## 2. Principles of Auxiliary Temporal Ring Bone Localization

In the robotic drilling operation for a cochlear implant, a surgical approach from the mastoid surface of the temporal bone to the tympanic step needs to be drilled, so it is necessary to locate the temporal bone in the body to determine the drilling point and the direction of approach [11]. Since there is no fixed and easily identifiable feature information for the temporal bone, it is difficult to calculate its pose through vision. Therefore, it is necessary to plant external markers to establish the spatial relationship between the markers and the temporal bone and to detect and calculate the pose information for external markers to obtain the current pose information for the temporal bone. Therefore, during robotic cochlear implant drilling surgery, temporal bone position and pose information detection are obtained by detecting temporal bone visual markers.

*2.1. Temporal Bone Positioning Auxiliary Ring.* During cochlear implant drilling, the temporal bone with implant markers was initially scanned with high-precision CT, and its three-dimensional image was reconstructed. Then, the relative position of the tissue structure in the ear was calculated to plan the drilling path based on the surgical conditions, and the relative postural relationship between the path and markers was calculated. Finally, the robot drills the planned surgical approach through the relative postural

relationship between the postural information from the markers and the path [12, 13]. Therefore, temporal bone markers should not only have a clear shadow in a CT scan but also visual features that can be detected by machine vision. The traditional locating method with external markers is random and subjective, and the irregular pattern is not conducive to machine vision detection and analysis. There are three main effects of robotic drilling. First, the spatial location of marker planting is too single, which will affect the calculation of depth information, which will lead to a large relative position error between marker and tissue structure. Second, too many marker points will cause secondary damage to the temporal bone. Third, it is difficult for robots to detect and analyze the traditional irregular planting location distribution. To resolve these issues, a robotic drilling auxiliary ring for cochlear implantation was proposed and designed, as shown in Figure 1.

The auxiliary ring is structurally divided into three parts: the inner ring, the outer ring, and the attached titanium sphere. The inner ring is nested in the outer ring, and the titanium sphere is attached to the torus of the outer ring and is sequentially mapped. The effects of traditional markers and auxiliary ring implantation are shown in Table 1. By using the auxiliary ring as the temporal bone visual registration marker, up to three wounds on the temporal bone can be fixed, thus reducing temporal bone injuries in the implanting process. The use of a circular structure is more in line with human engineering properties. The titanium spheres attached to the outer ring can be used as markers in CT scan reconstruction instead of titanium nails. The inner and outer ring structure adds visual features that are easy to detect on the auxiliary ring surface. The feature points of colors, fixed shapes, and distribution laws are beneficial to computer vision processing and analysis.

*2.2. Binocular Vision Measurements.* In the cochlear implant drilling operation, the auxiliary ring spatial position should be calculated, and real-time feedback of image information during the operation is needed. Binocular vision was used to locate the auxiliary ring. Based on the parallax principle, binocular vision reconstructs the three-dimensional coordinates of the target in space according to the two-dimensional coordinates of the measured target in the left and right images combined with the transformation relationship between each coordinate [14]. Binocular ranging needs to calibrate the binocular camera, determine the transformation relationship between the world coordinate system, pixel coordinate system, and camera coordinate system, and finally calculate the rotation matrix and translation vector between the internal and external camera parameters and the left and right cameras. The calibration tool for the STEREO Camera Calibrator in MATLAB was adopted, and Chang's calibration method was employed to calibrate the internal and external parameters of the left and right cameras. The calibration results are shown in Figure 2.

According to the mapping relationship between the projections, external parameters, and internal parameters, the distortion coefficients were obtained after the left and

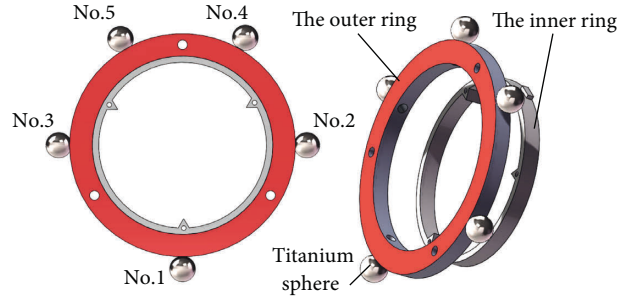


FIGURE 1: Auxiliary ring 3D model.

TABLE 1: Statistics of auxiliary ring model test results.

Target object identification	Real quantities	Quantity checks	Missed checks	Error checks
Outer ring	130	131	1	2
Titanium spheres	650	664	5	19

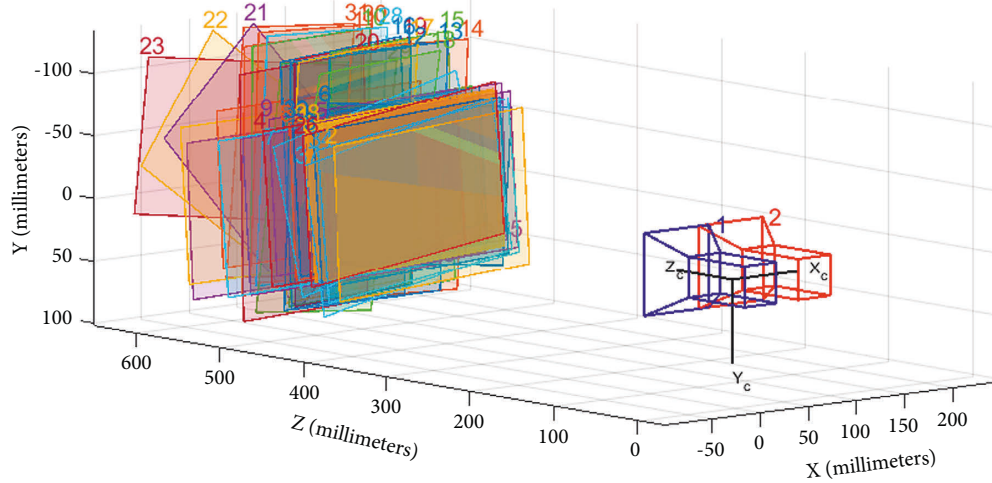


FIGURE 2: Binocular vision stereo calibration.

right cameras were calibrated, and the projection matrices of the cameras were  $M_l, M_r$ , respectively, allowing the world coordinates of any point in space to be  $(X, Y, Z)$ , then the following formula can be obtained:

$$\begin{cases} Z_{cr} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = M_r \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^r & m_{12}^r & m_{13}^r & m_{14}^r \\ m_{21}^r & m_{22}^r & m_{23}^r & m_{24}^r \\ m_{31}^r & m_{32}^r & m_{33}^r & m_{34}^r \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \\ Z_{cl} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = M_l \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^l & m_{12}^l & m_{13}^l & m_{14}^l \\ m_{21}^l & m_{22}^l & m_{23}^l & m_{24}^l \\ m_{31}^l & m_{32}^l & m_{33}^l & m_{34}^l \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \end{cases} \quad (1)$$

In the above formula,  $Z_{cr}, Z_{cl}$  is the distance from the projection of a point in space on the optical axis to the

optical center;  $M_r, M_l$  is the projection matrix of the camera; and  $(X, Y, Z)$  is the world coordinate of a point in space.

### 3. Auxiliary Ring Positioning Method

The temporal bone auxiliary ring positioning method DL-M proposed in this paper for cochlear implant drilling surgery can generally be divided into three parts: target region detection, feature point extraction, and position and pose a solution. The specific realization process is shown in Figure 3. First, the outer ring of the input image and the target region of the characteristic titanium spheres are detected by the trained deep learning model. Next, the feature points in the boundary box are extracted quickly, and the image coordinates are calculated. Then, binocular vision measures the three-dimensional coordinates of the feature points. Finally, the pose information for the auxiliary ring is solved according to the three-dimensional coordinates of the feature points.

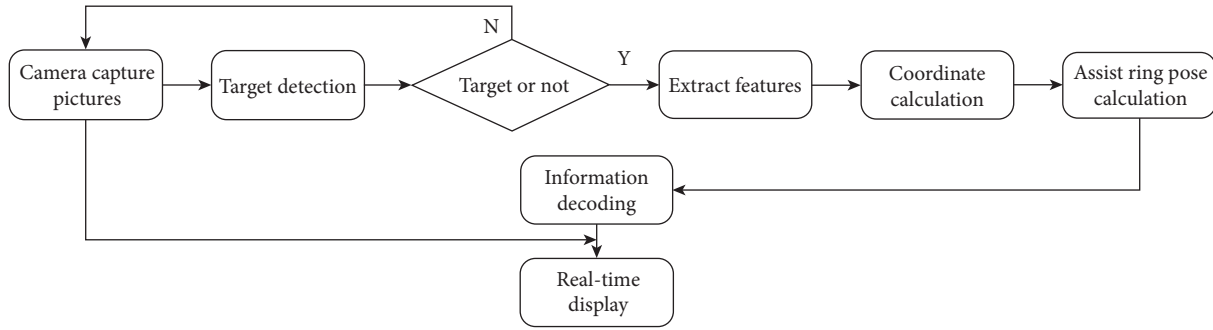


FIGURE 3: Auxiliary ring positioning method flow.

**3.1. Auxiliary Ring Target Detection.** How to efficiently identify an object to be measured is always one of the most important challenges for machine vision. Due to the influence of light, blood stains, instruments, and other factors, it is difficult to obtain the characteristic information of the auxiliary ring quickly and accurately through traditional detection methods in cochlear implant drilling surgery. Due to the development of the Convolutional Neural Network, deep learning algorithms based on the Convolutional Neural Network, it has gradually become the main method of target detection. Compared with traditional detection methods, deep learning algorithms are more advantageous in terms of speed, precision, and structure [15, 16]. In drilling surgery, the deep learning target detection algorithm can quickly and accurately identify the characteristic information of the auxiliary ring in complex surgical environments and accelerate auxiliary ring positioning.

To accurately and quickly detect the target auxiliary ring in a complex surgical environment, the YOLOV3 method was employed in this paper to detect the target related to the auxiliary ring [17]. The outer ring and titanium spheres were used as detection models for training auxiliary rings to obtain the boundary information of the outer ring and titanium spheres. The K-means clustering algorithm was used to calculate the prior parameters of anchor points in the data set, and the method for initiating random numbers in the clustering algorithm was changed. By analyzing the data set, 9 anchor points are given for the initial calculation, which can help the network better adjust the size of the bounding box while learning. In this experiment, the outer ring region and titanium sphere regions from about 650 images of the auxiliary rings with a different attitude in different environments were labeled, and the PASCAL VOC2007 dataset was established.

**3.2. Feature Extraction.** After the target features of the outer ring and titanium spheres were obtained by target detection, the target feature information in the boundary box should be further extracted, as shown in Figure 4. Firstly, the image of the attitude auxiliary ring was preprocessed by filtration and equalization. Secondly, the contour information in the boundary box of the outer ring was detected, screened, and fitted, and the fitted contour dataset was reconstructed by using relevant mathematical functions. Then, the

reconstructed ellipse dataset was matched with the original contour dataset to determine whether the contour was the target contour that met the conditions. Finally, the center of the fitted ellipse was calculated, which was the two-dimensional coordinate of the center of the auxiliary ring.

To reduce the image processing duration, ROI segmentation was conducted on the original image in the range of the bounding box information output after target detection, and image pre-processing was conducted within ROI to reduce the amount of image pre-processing calculations. Bilateral filtering removed the noise in images in the region of interest and the image edge information was retained. Then the target feature points were extracted by ellipse detection and a region growing algorithm. Finally, the feature points of the left and right cameras are matched to provide reliable matching points for binocular vision calculation.

**3.3. Posture Calculating.** After the matching target feature points were obtained, the characteristic coordinates of the auxiliary ring were obtained through the principles of binocular vision, and the auxiliary ring position and pose information were further calculated. The specific process is shown in Figure 5.

The feature points matched by the left and right images were substituted into (1), and irrelevant variables were eliminated. The least square fitting method solved the three-dimensional coordinates for the feature points.  $(x_i, y_i, z_i) i \in (1, 5)$ ,  $(q_x, q_y, q_z)$  represented the coordinates of the centers of the corresponding 1–5 titanium spheres and the corresponding inner ring. The obtained coordinates were the three-dimensional space coordinates for the auxiliary ring from the image plane coordinates. Given the world coordinates of the camera, the coordinate values of the auxiliary ring in the world coordinates can be obtained through coordinate transformation. To coordinate with each serial number corresponding to an individual sphere, this paper combined with the distribution rules of auxiliary ring attached titanium ball, and designed the coordinate collation of feature points by considering the intensive degree of space points. The intensive degree  $\rho_i$  of the coordinate distribution concerning the titanium cue ball was calculated by using the formula (5), the feature points were ordered by relative intension of  $\rho_i$ , and, thus, the titanium

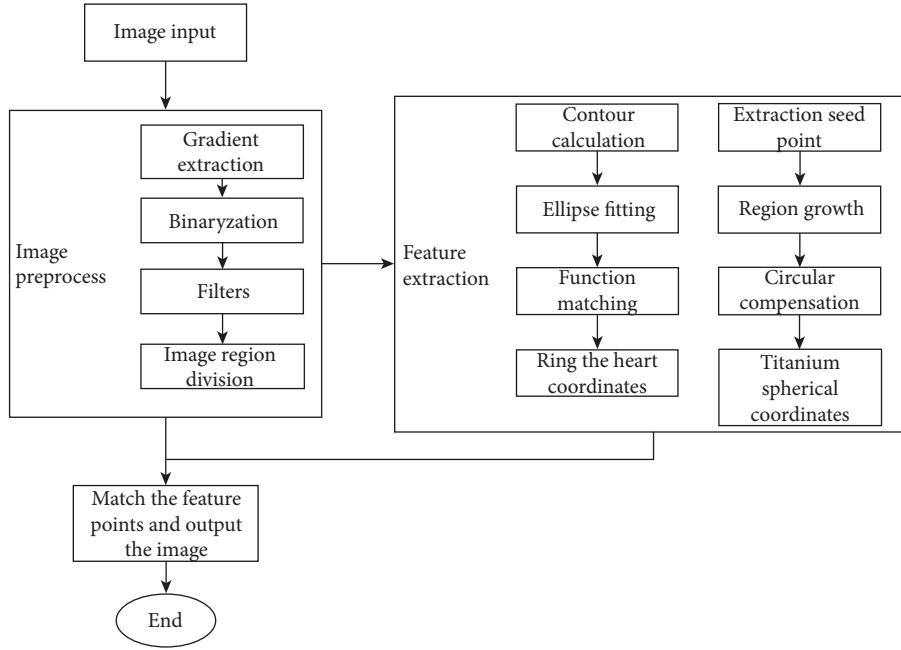


FIGURE 4: Feature point extraction process.

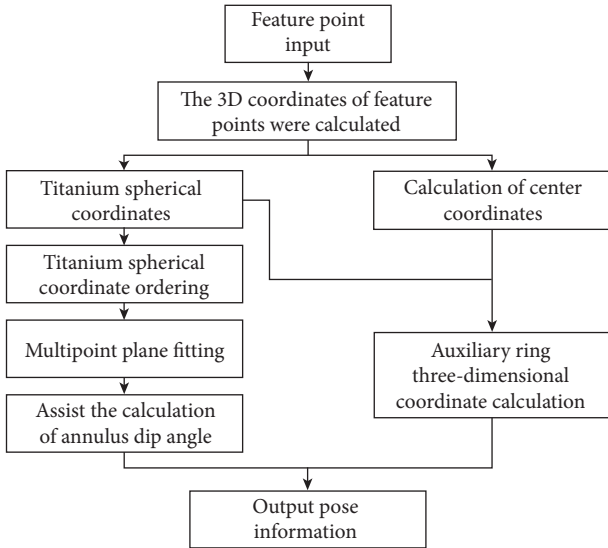


FIGURE 5: Pose and pose calculation process.

spherical coordinate in the image corresponded with the actual serial number of the ball. It can provide accurate characteristic correspondence information for the ring attitude solution.

According to the regular distribution of titanium spheres, a virtual average point of titanium spheres is:

$$p^{avg} = \frac{\sum(p^j)}{j} \quad (2)$$

Find the square of the absolute distance between each point and the average point.

$$\rho = \frac{1}{d^2} = (p^j - p^{avg})^2, \quad (3)$$

(2) and (3) are simultaneously established, and the irrelevant variables are eliminated to obtain:

$$\rho_i = \frac{1}{\sum_k \left( P_k^i - \frac{5}{j=1} P_k^j \right)^2} \{k = (x, y, z)\}, \quad (4)$$

Setting the auxiliary ring attitude as the initial attitude when the width coordinate of the line between the No. 1 titanium sphere and the ring center is perpendicular to the image. The rotation angles around the X, Y, and Z coordinate axes are defined as  $\theta_x, \theta_y, \theta_z$ , respectively. The initial angle is defined as when the No. 2 titanium ball and the No. 3 titanium ball are parallel to the X-axis of the spatial coordinates. The connection between the center of the No. 2 titanium ball and No. 3 titanium ball is parallel to X, so the rotation angle around the z-axis can be obtained only by calculating the tangent value of the connection and X, so  $\theta_z$  can be expressed as:

$$\theta_z = \tan^{-1} \left( \frac{y_3 - y_2}{x_3 - x_2} \right). \quad (5)$$

To reduce errors in angle calculation caused by visual detection and feature extraction, the three-dimensional coordinates of the spherical center were substituted into the spatial plane equation to fit the auxiliary ring plane by the spatial multipoint SVD plane fitting method to reduce the errors caused by binocular vision measurements and optimize the attitude calculation results.

$$AX = \begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ x_2 - \bar{x} & y_2 - \bar{y} & z_2 - \bar{z} \\ x_3 - \bar{x} & y_3 - \bar{y} & z_3 - \bar{z} \\ \dots & \dots & \dots \\ x_n - \bar{x} & y_n - \bar{y} & z_n - \bar{z} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0. \quad (6)$$

Taking the residual difference between the selected feature points and the fitting plane as the goal of minimization optimization, the fitting objective function is  $\min \|AX\|$ , with constraint  $\|X\| = 1$  through a singular value decomposition  $A = UDV^T$ . The feature vector corresponding to the minimum singular value was the coefficient vector of the fitting plane. The geometric meaning of  $a, b, c$  was the coordinate component values of the normal plan vector, so  $\theta_x, \theta_y$  can be expressed as:

$$\begin{cases} \theta_x = \tan^{-1}\left(\frac{a}{c}\right) \\ \theta_y = \tan^{-1}\left(\frac{b}{c}\right) \end{cases}. \quad (7)$$

Through the aforementioned coordinate calculation of feature points and plane fitting operation, the spatial coordinate position of the auxiliary ring under the camera plane was obtained. Since the spatial relationship between the temporal bone and the auxiliary ring was obtained by preoperative CT scanning, the spatial position of the temporal bone and structural tissue in the ear could be obtained by coordinate transformation.

#### 4. Rapid Extraction of Auxiliary Ring Features

In the temporal bone auxiliary ring visual positioning method, the key to pose detection is to extract the set features quickly and accurately. The more traditional BM or SGBM needs to calculate the pixel points of an entire image, and the algorithm takes a long time and is prone to producing parallax holes, which is not conducive to extracting specific feature points. In this paper, the DL-M feature point extraction scheme can quickly and accurately extract target features and, through improved ellipse matching, regional growth point selection and matching methods, accelerate target feature extraction.

*4.1. Ellipse Matching Based on Image Moments.* While detecting contour ellipse in the auxiliary ring, because the input contour was not screened in the Hough ellipse fitting, it is impossible to judge whether the original contours before fitting are an elliptic contour. To solve this problem, based on the concept behind the Hu moments image moment, this paper calculated the Hu-moment for the fitting contour and the original contour and judged it, which greatly eliminated any interference from nonelliptical contours. On this basis, a contour screening mechanism was added to reduce the computational burdens for ellipse recognition. After fitting the selected contour, the mathematical expression for the

fitting ellipse was established by using the rotation matrix information returned by the fitting, and the fitting ellipse pixel group was reconstructed. The calculation formula is:

$$V_{\text{ponit}}(x, y) = \begin{cases} P_i^x = x \cos \varphi - \sqrt{\left(1 - \frac{x^2}{a^2}\right)} b^2 \sin \varphi \\ P_i^y = \sqrt{\left(1 - \frac{x^2}{a^2}\right)} b^2 \cos \varphi + x \sin \varphi \\ \varphi = \frac{\text{angle} * \pi}{180} \end{cases}. \quad (8)$$

The geometric moments for the reconstructed elliptic contour and the original image contour were calculated, and the reciprocal deviations of the geometric moments were calculated to accumulate the moment errors in the original contour and reconstructed contours. The specific calculation formula is:

$$M_x = \sum_{i=1}^M \sum_{j=1}^N i^p j^q f(i, j), \quad (9)$$

$$I(A, B) = \sum_{i=1}^7 \left| \frac{1}{m_i^A} - \frac{1}{m_i^B} \right|.$$

The accumulated error function was then normalized. When the fitted ellipse contour was similar to the original contour,  $I(A, B)$  approached 0; otherwise, it approached 1. The maximum threshold was set to limit the errors between its geometric moments, and then the similarities between the two were judged.

*4.2. Obtaining the Planar Center of the Titanium Spheres.* While calculating the titanium sphere centers by the angle of view, part of the titanium sphere boundaries will be blocked by the outer ring plane, so they cannot be completely detected in the image. In this paper, the region-growing algorithm [18] and circle compensation were used to calculate the image coordinates of the titanium sphere centers. The challenge for the regional growth algorithm was selecting the seed starting point. Usually, the seed point is judged artificially, or the starting point of the seed is obtained by using a clustering algorithm adhering to a certain set of rules. This method was difficult to apply to the seed point selection for the titanium spheres because of their small size and indistinct features.

To solve the aforementioned problem, the region growth algorithm was designed in combination with the titanium sphere boundary box information obtained from YOLOV3 target detection. The specific process is shown in Figure 6. The midpoints of the titanium sphere beam boxes were utilized as the growth starting points, and the shape of the visible portions of the titanium spheres were described, and the field pixel values and growth sizes at the midpoints of the titanium sphere beam boxes was limited to exclude any incorrectly detected targets during target detection. The

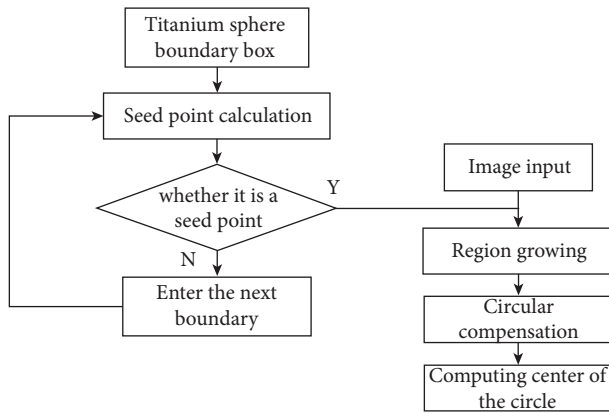


FIGURE 6: Calculation process for center coordinates.

minimum circumferential circle was used to complete the spherical surfaces of the titanium spheres, and their centers were calculated.

#### 4.3. Feature Point Matching Based on Polar Constraints.

After the feature points on the auxiliary rings from the left and right images were obtained through the aforementioned calculation, the order of the titanium spheres in the left and right images was different. Based on the concept of polar constraints, this paper matches the feature points of the titanium spheres, and the main process is:

- (1) Calculated the polar equation on the right image plane through the coordinates of the characteristic points of the centers of the titanium spheres on the left image plane.
- (2) Calculated the distances between the feature points in the right image and the polar line, and found the right feature points with the smallest distances after feature traversal. The smallest point was the matching point of the left image in the right image.
- (3) The matching feature information was stored, and the distance error was analyzed. If the error was greater than the set value, the point was brought into the region again for growth calculation.

With this method, the target feature points for the centers of the titanium spheres could be accurately matched. When compared with the traditional matching methods BM or SGBM based on polar line constraints, the calculation of useless coordinates during feature matching can be reduced. In addition to accelerating the matching speed, the precision of regional growth feature points was also calculated and analyzed.

## 5. Results and Discussion

In this paper, the method that calls for drilling an auxiliary ring for cochlear implants was used as the test object to detect features and calculate position and pose information. Firstly, the trained auxiliary ring model was tested on the test set, and the average detection rate, the missed detection rate,

and the false detection rate were calculated, and the effect of the model was evaluated by comparing it with the detection method. Then, the method of extracting feature points is verified to evaluate its reliability. Finally, the pose information of the auxiliary ring under different positions and postures is measured and the error is analyzed.

In the target detection experiment, 130 test pictures were tested, and some of the test results are shown in Figure 7. Among them, the number of auxiliary rings was 130, and the number of titanium spheres was 650. The number of detections, missed detections, and false detections for the model was statistically analyzed, and the results are shown in Table 1. It can be observed that the model presented in this paper has a high recognition rate for both the outer ring and the titanium spheres on the auxiliary ring, and the detection effect for the outer ring was better than that of the titanium sphere.

Using auxiliary ring structure characteristics of the feature point extraction scheme design, by combining deep learning target detection and image processing with a feature point detection method that incorporated a feature point matching method based on polar constraints, a large number of target detections were decreased, and the time needed for binocular image feature matching quickly extracted accurate target feature points. The specific process and effects are shown in Figure 8.

To better evaluate the current morphology and measurement data for the auxiliary ring during the operation, this paper used QT to design the image interactive interface. Multithreading displayed current image information in real time and controlled target detection, image processing, and information transmission programs, as shown in Figure 9. In the figure on the left, the image content obtained in real time by the left and right cameras was provided for real-time observation during the experiment. The image on the right is the target detection effect. The image in the lower right corner displays the current coordinate information of the auxiliary ring in real time. The interface of information transmission is also reserved in the program, which makes preparation for the interaction between the surgical robot and visual information.

The auxiliary ring was fixed on the experimental platform with a variable dip angle, and the auxiliary ring position and attitude were measured by changing the platform angle and setting the auxiliary ring dip angle. The binocular camera was placed about 25 cm above the experimental platform. To eliminate any error interference caused by the camera's own attitude, the camera was fixed, and the error between the actual and measured angle changes was calculated through the differences in experimental angle measurement. The position and attitude information of the auxiliary ring under random different attitudes were measured several times (greater than 50), and the standard deviations and maximum deviations obtained from all measurements were counted. The test results are shown in Table 2. Experimental data shows that the method has relatively high accuracy and small measurement fluctuations in the test environment, and the measurement results have good stability.

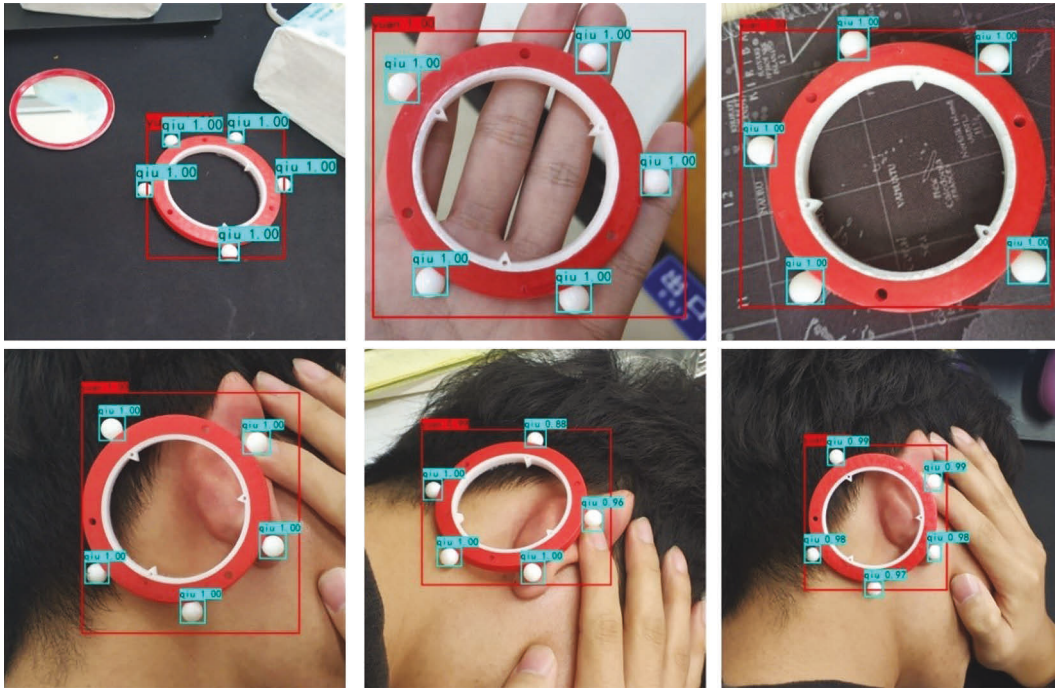


FIGURE 7: This is a figure. Schemes follow the same formatting.

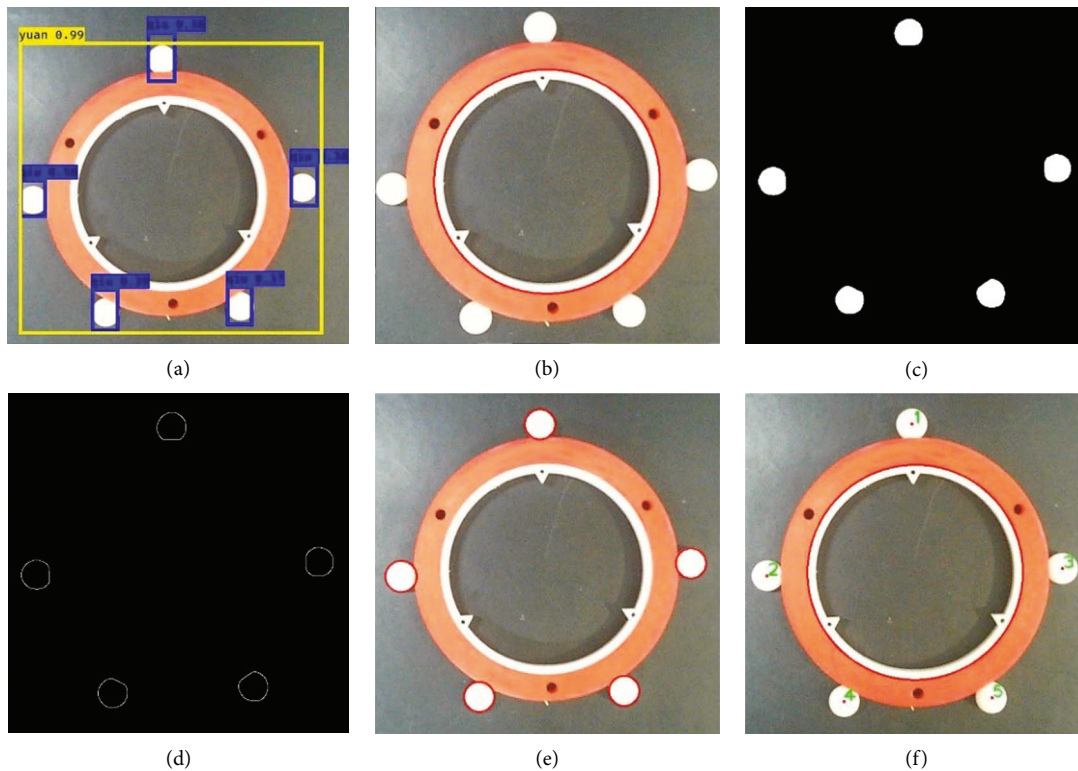


FIGURE 8: Feature Point Extraction Effect. (a) Target detection; (b) ellipse fitting; (c) region growing; (d) contour detection; (e) compensation outline; and (f) final extraction effect.

In robotic surgery, the use of titanium nails as markers does not give a fixed structure. The use of a triangulation bar usually requires the secondary calibration in the operation, which takes time and has a certain impact on the operating

space. The auxiliary ring of cochlear implant drilling used in this paper adopts the combination of simple shapes, colors, and materials, which not only meets the positioning requirements of the auxiliary robot cochlear implant drilling



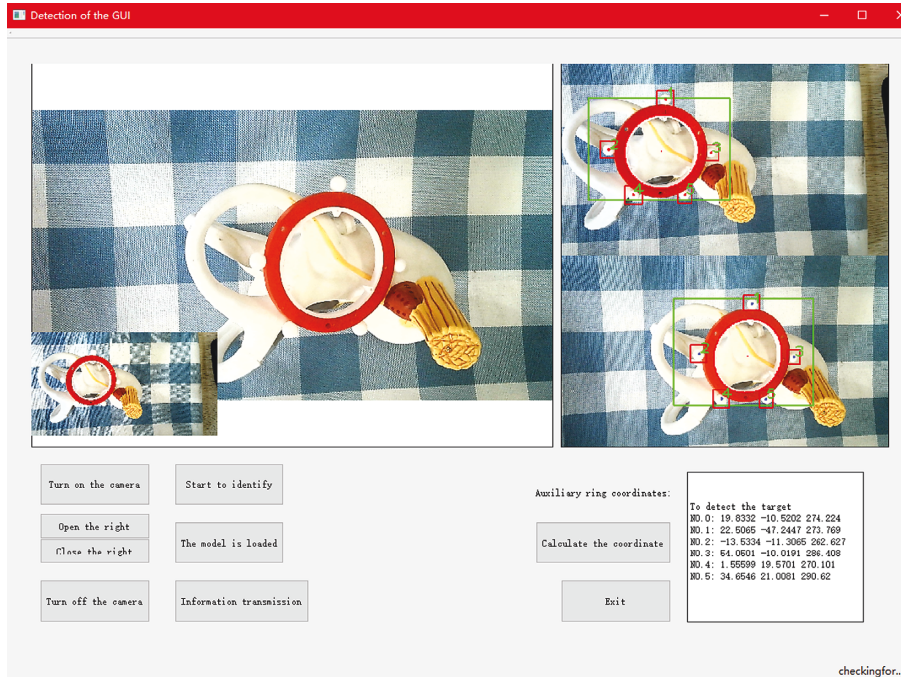


FIGURE 9: GUI interactive interface.

TABLE 2: Position and angle measurement results statistical table.

Measuring object	Max deviation	Standard deviation
X/Y	0.562 mm	0.287 mm
Z	1.331 mm	0.718 mm
$\theta_x/\theta_y$	0.632°	0.265°
$\theta_z$	0.312°	0.205°

TABLE 3: Comparisons of marker implanting effects.

Marker object	Volume	Number of wounds	Visual features	Feature fixed	Workspace impact
Titanium nail	Smallest	3~5	Shape	N	Minimal
Tripod visual calibration rod	Biggest	3~5	Shape	Y	Greatest
Auxiliary ring	Smaller	2~3	Shape, color	Y	Less

surgery but also provides enough operating space. It has certain advantages compared with titanium nail implantation or the use of a visual calibration rod. The effects of traditional markers and auxiliary ring implantation are shown in Table 3.

In this paper, the deep learning algorithm was combined with the binocular vision image processing auxiliary ring detection method DL-M. The neural network method is used to train the model of specific auxiliary rings in advance, and the deep learning target detection method is applied to medical-assisted detection, which is faster and more accurate than the traditional target detection method, and the detected perceptual region helps to reduce the computational effort and difficulty in the subsequent feature extraction. Compared with the traditional binocular vision matching algorithms, such as BM and SGBM, the detection speed and matching effect were improved. The DL-M method used in this paper is similar to the BM matching algorithm in speed, but the effect is more accurate. In this

paper, the DL-M method eliminated the duration of useless feature point matching and 3D coordinate calculation and utilized a fast image processing method to increase the accuracy of effective feature extraction for the auxiliary ring. In this paper, an image with a resolution of 1280 \* 720 was used to compare three visual matching algorithms, and the results are shown in Table 4.

Based on the aforementioned experiments, it can be seen that the DL-M method designed to detect the temporal bone in cochlear implant surgery using the auxiliary ring as the temporal bone marker in this paper can quickly and accurately detect the auxiliary ring, extract the features, solve the relative pose information, and then calculate the temporal bone pose. It reduces the duration spent on temporal bone localization while the cochlear implant robot drills, and the average detection and calculation efficiency is about 25FPS, which is a great improvement over traditional medical image navigation systems and binocular matching algorithms. The average attitude measurement accuracy is

TABLE 4: Comparison of method effects.

Method	Duration (ms)	Matching effect
BM	41	Poor
SGBM	305	Better
DL-M	40	Best

about  $\pm 0.6^\circ$ . The high frequency detection ability can further compensate for measurement errors and meet the requirements for the real-time acquisition of temporal bone postures in cochlear implant drilling surgery.

## 6. Conclusions

Combined with target detection, image processing, and binocular vision, this paper proposes a fast and real-time detection and calculation method for temporal bone marker localization in cochlear implant robot drilling surgery. The detection rate for auxiliary ring features is about 97%, and the overall detection and calculation time is about 40 ms. The average attitude measurement accuracy is  $\pm 0.63^\circ$ . The position and attitude information of the auxiliary ring can be obtained quickly, and the rapid calculation for the auxiliary ring position and pose can reduce the time of temporal bone visual positioning in cochlear implant drilling operations, greatly improving the robot's capabilities to extract visual information during the operation, which has a better auxiliary role for future research and applications of cochlear implant drilling operation. Since the accuracy of visual positioning is closely related to the binocular camera itself, the positioning accuracy of this method is still very limited. The accuracy of measurement may not yet fully meet the requirements of surgical positioning. The accuracy of the detection calculation is limited by camera performance, and the future research direction is to improve the accuracy of target detection and visual measurement, and our team will continue to conduct in-depth research in this direction. With the improvements in binocular vision technology and camera performance, this method will also make the research results more robust, and the use of this method will also obtain higher measurement accuracy, which our team will investigate here next.

## Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was funded by the National Natural Science Foundation of China (51705493) and the Natural Science Foundation of Zhejiang Province (LY17E050016).

## References

- [1] L. Bruschini, F. Forli, A. D. Vito, and S. Berrettini, "A new surgical approach for direct acoustic cochlear implant: a temporal bone study," *Clinical and experimental otorhinolaryngology*, vol. 9, no. 4, pp. 314–318, 2016.
- [2] P. Dai and Y. Jiang, "Precise and minimally invasive cochlear implant surgery," *Journal of Otolaryngology*, vol. 16, no. 6, pp. 9–13, 2018.
- [3] T. Williamson, X. Du, B. Bell et al., "Mechatronic feasibility of minimally invasive, atraumatic cochleostomy," *BioMed Research International*, vol. 2014, pp. 1–7, Article ID 181624, 2014.
- [4] Y. Y. Yuan and P. Dai, "Surgical strategy of minimally invasive cochlear implantation," *Zhonghua Yixue Zazhi*, vol. 101, no. 2, pp. 87–91, 2021.
- [5] H. Jia, J. X. Pan, Y. Li et al., "Preliminary application of robot-assisted electrode insertion in cochlear implantation," *Zhonghua er bi yan hou tou jing wai ke za zhi= Chinese Journal of Otorhinolaryngology Head and Neck Surgery*, vol. 55, no. 10, pp. 952–956, 2020.
- [6] Y. Zhao, B. M. Dawant, and J. H. Noble, "Automatic electrode configuration selection for image-guided cochlear implant programming," *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 9415 International Society for Optics and Photonics, Bellingham Washington USA.
- [7] J. Wang, H. Liu, J. Ke et al., "Image-guided cochlear access by non-invasive registration: a cadaveric feasibility study," *Scientific Reports*, vol. 10, no. 1, pp. 18318–18413, 2020.
- [8] B. Cho, N. Matsumoto, S. Komune, and M. Hashizume, "A surgical navigation system for guiding exact cochleostomy using auditory feedback: a clinical feasibility study," *BioMed Research International*, vol. 2014, Article ID 769659, 7 pages, 2014.
- [9] N. P. Dillon, R. Balachandran, and R. F. Labadie, "Accuracy of linear drilling in temporal bone using drill press system for minimally invasive cochlear implantation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 3, pp. 483–493, 2016.
- [10] J. Ke, S. Zhang, C. Li, Y. Zhu, L. Hu, and F. Ma, "Application of bonebed-malleus short process registration in minimally invasive cochlear implantation," *Computer Assisted Surgery*, vol. 21pp. 30–36, sup1, 2016.
- [11] S. Duret, C. Guigou, M. Grelat, and A. Bozorg-Grayeli, "Minimally invasive cochlear implantation assisted by intraoperative CT scan combined to neuronavigation," *Otology & Neurotology*, vol. 41, no. 4, pp. e441–e448, 2020.
- [12] B. Copson, S. Wijewickrema, X. Ma, Y. Zhou, J. M. Gerard, and S. O'Leary, "Surgical approach to the facial recess influences the acceptable trajectory of cochlear implantation electrodes," *European Archives of Oto-Rhino-Laryngology*, vol. 279, no. 1, pp. 137–147, 2021.
- [13] A. Karkas, N. M. d. Champfleury, A. Uziel, M. Mondain, J. L. Puel, and F. Venail, "Benefit of preoperative temporal bone CT for atraumatic cochlear implantation," *Otology & Neurotology*, vol. 39, no. 3, pp. 186–194, 2018.
- [14] X. Sun, Y. Jiang, Y. Ji et al., "Distance measurement system based on binocular stereo vision," *IOP Conference Series: Earth and Environmental Science*, vol. 252, no. 5, p. 052051, Article ID 052051, 2019.
- [15] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-

- supervised anomaly detection in videos,” *Journal of Imaging*, vol. 4, no. 2, 2018.
- [16] T. Fan, “Research and realization of video target detection system based on deep learning,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 18, Article ID 1941010, 01 pages, 2020.
- [17] J. Redmon and A. Farhadi, *YOLOv3: An Incremental Improvement*, arXiv e-prints, 2018, <https://doi.org/10.48550/arXiv.1804.02767>.
- [18] S. Kamdi and R. K. Krishna, “Image segmentation and region growing algorithm,” *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, vol. 2, no. 1, pp. 103–107, 2012.