Abstract: Three-dimensional (3D) triangulation based on active binocular vision has increasing amounts of applications in computer vision and robotics. An active binocular vision system with non-fixed cameras needs to calibrate the stereo extrinsic parameters online to perform 3D triangulation. However, the accuracy of stereo extrinsic parameters and disparity have a significant impact on 3D triangulation precision. We propose a novel eye gaze based 3D triangulation method that does not use stereo extrinsic parameters directly in order to reduce the impact. Instead, we drive both cameras to gaze at a 3D spatial point $P$ at the optical center through visual servoing. Subsequently, we can obtain the 3D coordinates of $P$ through the intersection of the two optical axes of both cameras. We have performed experiments to compare with previous disparity based work, named the integrated two-pose calibration (ITPC) method, using our robotic bionic eyes. The experiments show that our method achieves comparable results with ITPC.

Keywords: active binocular vision; 3D coordinates estimation; eye gaze; 3D triangulation; robotic bionic eyes

1. Introduction

Active binocular vision is a binocular vision system that could actively change its own view direction. It is beneficial for multiple applications such as manipulation [1], three-dimensional (3D) reconstruction [2], navigation [3], 3D mapping [4], and so on. 3D coordinates estimation based on active binocular vision attracts extensive research interests.

Active binocular vision systems can be divided into two categories: the first category has fixed cameras and the second category has non-fixed cameras. The vision systems of ASIMO [5], HRP-3 [6], HRP-4 [7], PR2 [8], ATLAS [9], and Walkman [10] belong to the first category. However, this category can not perceive objects that are very close to the cameras. The second category, which is similar as human vision system, can obviously improve flexibility and extend the field of view. The vision systems of the ESCHeR head [11], Yorick head [12], MAC-EYE robot eyes [13], iCub head [14], ARMAR-III head [15], Fiobi head [16], Zhang XL’s robot eyes [17], ARMAR-4 head [18], Muecas [19],
Romeo [20], SARA [21], and our designed anthropomorphic robotic bionic eyes [22–24] belong to the second category.

Usually, 3D triangulation can be performed using disparity and stereo extrinsic parameters [25–31]. The stereo extrinsic parameters are calibrated offline [32,33] or online [34–39]. Active vision systems of the first category only need offline calibration as the stereo extrinsic parameters are fixed, since one camera is static with respect to (w.r.t) the other. The stereo extrinsic parameters of the second category need to be calibrated online because the stereo extrinsic parameters may change all the time. In [40], Chen et al. proposed an integrated two-pose calibration (ITPC) method that calculated the stereo extrinsic parameters online using forward kinematics and the calibrated head-eye parameters.

However, in the disparity based triangulation method, the errors of the stereo extrinsic parameters and disparity have a significant impact on 3D triangulation [41]. In [41], Dang et al. analyzed that the stereo extrinsic parameters error can generate baseline error, pixel errors. The disparity error is generated by image detection and matching. The baseline error, pixel errors, and the disparity error will directly affect the 3D triangulation precision. In order to reduce the impact, this paper proposes a novel eye gaze based 3D triangulation method that does not use stereo extrinsic parameters directly. In our proposed method, the 2D image of a spatial point $P$ in each camera is kept at the principal point through eye gazing or visual servoing. Or, the point $P$ is always at the location of the fixation point of the two cameras. After that, we could obtain 3D coordinates of $P$ through intersection of the two optical axes. In real applications, the fixation point $P$ is represented by the middle point of the common perpendicular line segment of the two skew optical axes. Figure 1 shows the idea. The advantage of the eye gaze based triangulation is that we no longer need the stereo extrinsic parameters directly. The other advantage is that image detection error tolerance is higher than the disparity based triangulation method.

![Diagram](image.png)

Figure 1. In our proposed method, we use eye gazing or visual servoing to actively adjust the view direction of each camera to keep the 2D image of a three-dimensional (3D) point $P$ at the principal point. In other words, $P$ is located at the fixation point. 3D triangulation could be performed through the intersection of two cameras’ optical axes.

The main contribution of this paper is that keeping the target point lies on the optical axes of each camera through eye gazing, we can perform 3D triangulation through calculating the intersecting point of two optical axes of cameras instead of solving the projection equations of two cameras that rely on disparity and stereo extrinsic parameters.
2. System Configuration

The anthropomorphic robotic bionic eyes we designed (as shown in Figure 2) for human-like active stereo perception, is illustrated in detail in [24]. This is a revised version with seven degree of freedoms (DOFs). It has two eyes and one neck. Each of the eye has two DOFs: tilt and pan. The neck has three DOFs. Joints of the neck and eyes are serially connected. It means that the links are connected serially by the neck and eyes joints and form ordered open chains. In the ordered open chain, the next joint is the load of the previous joint. As shown in Figure 3, our designed robotic bionic eyes contain two serial open chains: frame $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$, frame $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 7 \rightarrow 8 \rightarrow 9$. Each eye is installed with a USB camera with resolution of $640 \times 480 @ 120$ fps. A common trigger signal is connected to both cameras in order to realize hardware synchronization.

![Figure 2. The anthropomorphic robotic bionic eyes.](image)

We define the coordinate systems that are based on the standard D-H convention. Figure 3 shows the frame definitions. Table 1 shows the D-H parameters of the robotic bionic eyes. The transformation matrix $i^{-1}T_i$ between Frame $i$ and Frame $i-1$ can be calculated according to the D-H parameters [42].

![Figure 3. The definition of the robot bionic eyes' coordinate system.](image)
Table 1. The D-H parameters of the robotic bionic eyes. (Here parameters \(d_i\), \(\theta_i\), \(a_i\), \(\alpha_i\) are the link offset, joint angle, link length and link twist, respectively. The \(i\)-th joint offset is the value of \(\theta_i\) in the initial state. \(L_1\) is 64.27 mm, \(L_2\) is 11.00 mm, \(L_3\) is 44.80 mm, \(L_3r\) is 47.20 mm, \(L_4\) is 13.80 mm, and \(L_5\) is 30.33 mm).

| \(i\) | \(d_i/\text{mm}\) | \(\theta_i/\text{rad}\) | \(a_i/\text{mm}\) | \(\alpha_i/\text{rad}\) | \(i\)-th joint offset/\(\text{rad}\) |
|---|---|---|---|---|---|
| 1 | 0 | \(\theta_1\) | 0 | 0 | 0 |
| 2 | \(L_1\) | \(\theta_2\) | 0 | \(-\pi/2\) | \(-\pi/2\) |
| 3 | \(L_2\) | \(\theta_3\) | 0 | \(\pi/2\) | \(\pi/2\) |
| 4 | \(-L_3\) | \(\theta_4\) | \(L_4\) | \(\pi\) | \(\pi\) |
| 5 | \(L_5\) | \(\theta_5\) | 0 | \(\pi/2\) | \(\pi/2\) |
| 6 | 0 | 0 | 0 | \(-\pi/2\) | \(-\pi/2\) |
| 7 | \(L_3r\) | \(\theta_7\) | \(L_4\) | 0 | \(\pi\) |
| 8 | \(L_5\) | \(\theta_8\) | 0 | \(-\pi/2\) | \(-\pi/2\) |
| 9 | 0 | 0 | 0 | \(-\pi/2\) | \(-\pi/2\) |

3. Eye Gaze Based 3D Triangulation Method

The main idea of the eye gaze [43–46] based 3D triangulation method is to control the view direction of both cameras to keep a 3D spatial point \(P\) lies on the optical axes of both cameras. We are able to calculate the exact 3D coordinates of \(P\) through calculating the intersection point of the two optical axes, as we could obtain the pose of each camera through forward kinematics. We also know that once \(P\) lies on the optical axis of a camera, the image of \(P\) will be at the principal point of the 2D image. Accordingly, we could perform image based visual servoing to change the pose of both cameras to keep the images of \(P\) at the principal points of two cameras and, consequently, \(P\) lies on the two optical axes.

In this section, we first introduce visual servoing based approach for eye gazing, and then we show the process of calculation of the intersection point.

3.1. Visual Servoing Based Eye Gazing

We use image based visual servoing to control the pose of each camera. We will first get relationship of the velocity of the image feature points and the velocity of the joints. After that, we will present a simple visual servoing control law in order to generate commands to drive the feature point error down to zero.

3.1.1. Relation Matrix

Here, we use a relation matrix to describe the velocity of the image feature points and the joint velocity. The relation matrix could be decomposed to be the multiplication of an Interaction matrix and a Jacobian matrix. An Interaction matrix is used to build the relationship of the velocity of an image feature point and the velocity of the camera. A Jacobian Matrix is used to describe the relationship of the velocity of the camera and joint velocity.

\[
p = J_{\text{image}}V_n,
\]

where \(p = [\dot{u} \ \dot{v}]^T\) is the velocity of an image feature point, \(V_n = [v_x \ v_y \ v_z \ \omega_x \ \omega_y \ \omega_z]^T\) is the velocity of a camera, and \(J_{\text{image}}\) is the Interaction matrix calculated, as follows:

\[
J_{\text{image}} = \begin{bmatrix}
-f_u/Z_c & 0 & \frac{u}{Z_c} & \frac{w}{f_v} & \frac{-f_u^2 + u^2}{f_v^2} & \frac{v}{f_v} \\
0 & -f_v/Z_c & \frac{v}{Z_c} & \frac{w}{f_u} & \frac{-f_v^2 + v^2}{f_u^2} & \frac{-u}{f_u}
\end{bmatrix},
\]

where \(f_u\) and \(f_v\) is the focal length in the \(X\) and \(Y\) directions of the camera, respectively, \(Z_c\) is the depth of the spatial point \(P\) in the camera frame.
The relationship between \( V_n \) and the joint velocity \( \dot{\theta} \) could be described, as follows:

\[
V_n = J_{\text{joint}} \dot{\theta},
\]

where \( J_{\text{joint}} \) is the Jacobian matrix w.r.t camera frame, and it is defined as:

\[
J_{\text{joint}} = \begin{bmatrix} R & 0 \\ 0 & R \end{bmatrix} J_{\theta},
\]

where \( R \) is the rotation matrix of the base frame w.r.t the camera frame, and \( J_{\theta} = [J_{\theta_1} \ J_{\theta_2} \ldots \ J_{\theta_n}]^T \) is the Jacobian matrix w.r.t the base frame. The elements \( J_{\theta_i} \) \((i = 1, 2, \ldots, n)\) can be calculated, as follows:

\[
J_{\theta_i} = \begin{bmatrix} Z_{i-1} \times (O_n - O_{i-1}) \\ Z_{i-1} \end{bmatrix},
\]

where \( Z_i \) and \( O_i \) are the direction vector of \( Z \) axis and the origin of Frame \( O_i - x_iy_iz_i \) w.r.t base frame.

Accordingly, the relationship between the velocity of an image feature point \( \dot{p} = [\dot{u} \ \dot{v}]^T \) and the joint velocity \( \dot{\theta} \) can be formulated, as follows:

\[
\dot{p} = J_{\text{image}} J_{\text{joint}} \dot{\theta}.
\]

### 3.1.2. Image Based Visual Servoing Control Law

We use a simple image based visual servoing control law to control the motion of the camera so as to reduce the error of the expected image point \( p^* = [u^* \ v^*]^T \) and the actual image point \( p = [u \ v]^T \) of 3D spatial point \( P \), where \( [u^* \ v^*]^T \) are set to the principal point \( [u_0 \ v_0]^T \), which can be achieved from the intrinsic parameters, and \( [u \ v]^T \) are the actual pixel coordinates of the spatial point \( P \) in the camera.

\[
\dot{e} = -Ke,
\]

where \( K = \text{diag}\{k_1, k_2\} \) is the gain matrix, and where \( k_1 \) and \( k_2 \) are the gain used to control how fast the pixel errors in the \( X \) and \( Y \) directions are regulated to zero independently [47], and \( e = p^* - p = [u^* - u \ v^* - v]^T \) is the error between \( p^* \) and \( p \). From Equation (7), we can infer that the error will decay to zero exponentially, where \( K \) is the exponential coefficient.

Because

\[
\dot{e} = \dot{p^*} - \dot{p} = -\dot{p},
\]

we could derive the following equation from Equations (6)–(8).

\[
J_{\text{image}} J_{\text{joint}} \dot{\theta} = K(p^* - p) = K \begin{bmatrix} u^* - u \\ v^* - v \end{bmatrix}.
\]

The above equation could be rewritten for the left eye and right eye, respectively. For the left eye, we could derive

\[
J_{\text{image}} J_{\text{joint}} \dot{\theta}_l = J_l \dot{\theta}_l = K(p^*_l - p_l) = K \begin{bmatrix} u^*_l - u_l \\ v^*_l - v_l \end{bmatrix},
\]

where \( p_l = [u_l \ v_l]^T \) and \( p^*_l = [u^*_l \ v^*_l]^T \) are the actual pixel coordinates and the expected pixel coordinates of the spatial point in the left camera, respectively, and \( \dot{\theta}_l = [\dot{\theta}_1 \ \dot{\theta}_2 \ \dot{\theta}_3 \ \dot{\theta}_4 \ \dot{\theta}_5]^T \) is the joint velocity.

We fix the neck joints and establish the base frame \( O_N - x_Ny_Nz_N \) at link 3, the joint velocity vector is rewritten as:

\[
\dot{\theta}_l = [0 \ 0 \ 0 \ \dot{\theta}_4 \ \dot{\theta}_5]^T.
\]
From Equations (10) and (11), the velocities of left eye joints can be calculated, as follows:

\[
\begin{bmatrix}
\dot{\theta}_4 \\
\dot{\theta}_5 \\
\end{bmatrix}
= 
\begin{bmatrix}
J_{l14} & J_{l15} \\
J_{l24} & J_{l25} \\
\end{bmatrix}
^{-1}
K(p_l^* - p_l) = K
\begin{bmatrix}
u_l^* - u_l \\
v_l^* - v_l \\
\end{bmatrix},
\]

(12)

where \(J_{lj}\) is the element on the \(i\)-th row and \(j\)-th column in the matrix \(J_l\).

Similarly, the velocities of the right eye joints can be calculated, as follows:

\[
\begin{bmatrix}
\dot{\theta}_7 \\
\dot{\theta}_8 \\
\end{bmatrix}
= 
\begin{bmatrix}
J_{r14} & J_{r15} \\
J_{r24} & J_{r25} \\
\end{bmatrix}
^{-1}
K(p_r^* - p_r) = K
\begin{bmatrix}
u_r^* - u_r \\
v_r^* - v_r \\
\end{bmatrix}.
\]

(13)

In each loop of the visual servoing, we send the generated eye joint velocity commands to drive the robotic bionic eyes to the view direction that reduces the error between \(p^*\) and \(p\) in both cameras.

### 3.2. 3D Triangulation by Calculation of Intersection Point

After visual servoing, the point \(P\) will lie on the optical axes of each camera, and theoretically the intersection point of the two optical axes is the representation of \(P\). We could calculate the intersection point geometrically if the equations of the two optical axes are obtained. We could use forward kinematics in order to derive the description of the two optical axes because each axis is the \(Z\) of the camera frame for each eye. In real situation, the two optical lines may not intersect each other, and so we use the middle point of the common perpendicular line segment of the two skew optical axes to be the representation of the intersection point instead.

#### 3.2.1. Calculation of the Coordinates of the Optical Axes

First of all, we will deduce the process to calculate the coordinates of the two optical axes \(L_l\) and \(L_r\). The optical axis of each camera is the coordinates of \(Z\) axis w.r.t the neck frame \(N\), and so we need to derive the transformation matrix of the frame of the two cameras w.r.t frame \(N\) through forward kinematics, or the pose of left camera \(N^T_{Cl}\) and the pose of right camera \(N^T_{Cr}\).

\[
N^T_{Cl} = N^T_6^6T_{Cl},
\]

(14)

where \(^6T_{Cl}\) is the head-eye parameters of the left camera, which can be calculated using the method in [40]. \(N^T_6\) is represented as the transformation matrix of the left eye frame w.r.t the neck frame \(N\), and it can be obtained using the D-H parameters of the robotic bionic eyes, as follows:

\[
N^T_6 = N^T_3^3T_4^4T_5^5T_6.
\]

(15)

It is also the case for the right camera.

\[
N^T_{Cr} = N^T_9^9T_{Cr},
\]

(16)

where \(^9T_{Cr}\) is head-eye parameters of the right camera. \(N^T_9\) is the transformation matrix of the right eye frame w.r.t the neck frame \(N\), it can be obtained, as follows:

\[
N^T_9 = N^T_3^3T_7^7T_8^8T_9.
\]

(17)

Accordingly, we can perform real-time calculation of \(L_l\) and \(L_r\) using Equations (14) and (16).

#### 3.2.2. Calculation of the Intersection Point

We are going to calculate the intersection point of the optical lines \(L_l\) and \(L_r\). Here, \(L_l\) and \(L_r\) are the ideal optical axes. The actual optical axes \(L'_l\) and \(L'_r\) will no longer intersect each other due to the
measurement errors. It means that $L_{l'}$ and $L_{r'}$ may not in the same plane. We use the middle point of their common perpendicular line segment $P_1P_2$ to be the representation of $P$ instead, which is shown in Figure 4.

![Figure 4. The actual model of our proposed method.](image)

In order to obtain $P$, we need to calculate $P_1$ and $P_2$. Here, $P_1$ is the intersection of line $L_{r'}$ and plane derived from $L_{l'}$ and $P_1P_2$. $P_2$ is the intersection of line $L_{l'}$ and plane derived from $L_{r'}$ and $P_1P_2$.

First of all, we would like to calculate the plane derived from $L_{l'}$ and $P_1P_2$. In order to do that, we will introduce two points $C$, $D$ on line $L_{r'}$ with coordinates of $C = (0, 0, 10, 1)^T$ and $D = (0, 0, 20, 1)^T$ in the right camera frame. We set the coordinates in frame $N$ as $N_C$ and $N_D$. And $N_C = N_T C$ and $N_D = N_T D$. We will also introduce two points $A$, $B$ on line $L_{l'}$. The $L_{r'}$ in frame $N$ (in the following statements, all of the parameters are represented in frame $N$) are defined, as follows:

$$X = C + \lambda(D - C). \tag{18}$$

We use $n_{L_{l'}}$, $n_{L_{r'}}$, $n_{P_1P_2}$, $n_{P_1P_2L_{l'}}$ to represent the direction of $L_{l'}$, $L_{r'}$, $P_1P_2$ and the norm of plane $P_1P_2L_{l'}$ respectively. Accordingly, $n_{P_1P_2} = n_{L_{l'}} \times n_{L_{l'}}$, $n_{P_1P_2L_{l'}} = n_{L_{l'}} \times n_{P_1P_2}$, and the plane $P_1P_2L_{l'}$ could be written as:

$$n_{P_1P_2L_{l'}} \cdot (X - A) = 0. \tag{21}$$

From Equations (18) and (21), we could get the intersection point $P_1$ between plane $P_1P_2L_{l'}$ and line $L_{r'}$.

$$P_1 = C + \frac{n_{P_1P_2L_{l'}} \cdot (A - C)}{n_{P_1P_2L_{l'}} \cdot (D - C)}(D - C). \tag{22}$$

Similarly, we could get $P_2$ like this:

$$P_2 = A + \frac{n_{P_1P_2L_{r'}} \cdot (C - A)}{n_{P_1P_2L_{r'}} \cdot (B - A)}(B - A). \tag{23}$$
Additionally, the fixation point $P$ could be obtained:

$$
P = \frac{P_1 + P_2}{2}.
$$

(24)

Hence, the eye gaze based 3D triangulation Algorithm 1 can be summarized, as follows:

**Algorithm 1: Eye Gaze Based 3D Triangulation Method.**

**Visual Servoing based Eye Gazing:**

- Gaze the 3D point $P$ at the principal point in the left and right cameras simultaneously through visual servoing based eye gazing. In each loop of the visual servoing:
  - Calculate the joint velocity commands of the robotic bionic eyes using Equations (12) and (13) according to pixel deviation in the left eye and right eye, respectively.
  - Send the joint velocity commands to the robotic bionic eyes and get the feedback of joint angles.

**3D Triangulation by Calculation of Intersection Point:**

- Calibrate the head-eye parameters $^{6}T_{C_l}$ and $^{9}T_{C_r}$ using method in [40].
- If visual servoing based eye gazing is successful:
  - Calculate the transformation matrices $^{N}T_{6}$ and $^{N}T_{9}$ using the forward kinematics with joint position feedback.
  - Obtain the optical axes of the left and right cameras using Equations (14) and (16).
  - Calculate the 3D coordinates of $P$ using Equations (22)–(24).

4. Propagation of Uncertainty

Uncertainty [48–50] is a quantification of the doubt about the validity of measurement results. The uncertainties of $P$ using our method and the ITPC method are mainly propagated from image detection uncertainty. For our method, because some times the image uncertainty is tiny, which may result in no motion of optical axes because of the joint feedback precision ($\pm 0.04^\circ$) is not as high as image coordinates. Subsequently, we could not measure the uncertainty directly, because there is no variation of the coordinates of $P$. In order to solve this problem, we use the law for the propagation of uncertainty [51,52] instead to calculate the theoretical uncertainties of point $P$ obtained from the two methods that are based on image detection uncertainty.

4.1. Image Detection Uncertainty

The image detection uncertainty is represented as $u_u$ and $u_v$. They can be calculated, as follows:

$$
u_u = \sqrt{\frac{\sum_{i=1}^{n} (u_i - \bar{u})^2}{n - 1}},
$$

(25)

$$
u_v = \sqrt{\frac{\sum_{i=1}^{n} (v_i - \bar{v})^2}{n - 1}},
$$

(26)

where $u_i$, $v_i$ are the $i$-th ($i = 1, 2, \ldots, n$) independent observation of the image coordinates $u$, $v$, respectively. $\bar{u}$, $\bar{v}$ are the average of the $n$ observations of $u$, $v$, respectively.
4.2. Uncertainty of $P$ Using Our Method

If eye gazing converges at time $t$, the image point of $P$ coincide with the principal point in both cameras. After time $t$, the eye joint angles $\theta_i$ ($i = 4, 5, 7, 8$) will vary according to the variation of the image coordinates of $P$ in both cameras through eye gazing. For the left eye, the joint angles $\theta_4, \theta_5$ can be formulated, as follows:

$$\theta_4 = \arctan\left(\frac{v_l - v_0}{f_{v_l}}\right),$$

$$\theta_5 = \arctan\left(\frac{u_l - u_0}{f_{u_l}}\right),$$

where $u_l, v_l$ are the image of $P$ in the left camera, $u_0, v_0$ are the principal point in the left camera, and $f_{u_l}, f_{v_l}$ are the focal length in the $X$ and $Y$ directions of the left camera, respectively.

The uncertainties of the eye joint angles $\theta_i$ ($i = 4, 5$) can be calculated by propagating the uncertainties of the image of $P$ in the left camera through partial derivatives of Equations (27) and (28), as follows:

$$u_{\theta_4} = \frac{\partial \theta_4}{\partial v_l} u_{v_l},$$

$$u_{\theta_5} = \frac{\partial \theta_5}{\partial u_l} u_{u_l},$$

where $u_{v_l}, u_{u_l}$ can be obtained using Equations (25) and (26). We use $\overline{v_l}, \overline{u_l}$ to be the replacement of $v_l$ and $u_l$ in $\frac{\partial \theta_4}{\partial v_l}$ and $\frac{\partial \theta_5}{\partial u_l}$. The uncertainties propagation of $\theta_7, \theta_8$ are similar to $\theta_4, \theta_5$.

From the previous Section, we derived that the 3D point $P$ can be triangulated by Equations (22)–(24). The uncertainties of $P$ can be calculated by propagating the uncertainties of the eye joint angles $\theta_i$ ($i = 4, 5, 7, 8$) through partial derivatives of Equation (24). The uncertainties of $P = (x_p, y_p, z_p)^T$ using our proposed method can be formulated, as follows:

$$u_{x_p} = \sqrt{\left(\frac{\partial x_p}{\partial \theta_4}\right)^2 u_{\theta_4}^2 + \left(\frac{\partial x_p}{\partial \theta_5}\right)^2 u_{\theta_5}^2 + \left(\frac{\partial x_p}{\partial \theta_7}\right)^2 u_{\theta_7}^2 + \left(\frac{\partial x_p}{\partial \theta_8}\right)^2 u_{\theta_8}^2},$$

$$u_{y_p} = \sqrt{\left(\frac{\partial y_p}{\partial \theta_4}\right)^2 u_{\theta_4}^2 + \left(\frac{\partial y_p}{\partial \theta_5}\right)^2 u_{\theta_5}^2 + \left(\frac{\partial y_p}{\partial \theta_7}\right)^2 u_{\theta_7}^2 + \left(\frac{\partial y_p}{\partial \theta_8}\right)^2 u_{\theta_8}^2},$$

$$u_{z_p} = \sqrt{\left(\frac{\partial z_p}{\partial \theta_4}\right)^2 u_{\theta_4}^2 + \left(\frac{\partial z_p}{\partial \theta_5}\right)^2 u_{\theta_5}^2 + \left(\frac{\partial z_p}{\partial \theta_7}\right)^2 u_{\theta_7}^2 + \left(\frac{\partial z_p}{\partial \theta_8}\right)^2 u_{\theta_8}^2}.$$

In the above equations, we use $\overline{\theta_{fi}}$ or the average of the $n$ independent repeated observations of the $i$-th ($i = 4, 5, 7, 8$) joint angle feedback as the value of $\theta_i$ in $\frac{\partial x_p}{\partial \theta_i}$, $\frac{\partial y_p}{\partial \theta_i}$, and $\frac{\partial z_p}{\partial \theta_i}$.

4.3. Uncertainty of $P$ Using ITPC Method

In the ITPC method, the eye joint angles $\theta_i$ ($i = 4, 5, 7, 8$) will not change after time $t$. The stereo extrinsic parameters $^C^T_{Cl}$ can be calculated using the eye joint angles $\theta_i$ ($i = 4, 5, 7, 8$). Here, $\theta_i = \theta_{fi}$, where $\theta_{fi}$ is the $i$-th joint angle feedback at time $t$. $P = (x_p, y_p, z_p)$ can be calculated while using the intrinsic parameters of both cameras, the stereo extrinsic parameters $^C^T_{Cl}$, and the image coordinates of $P$ in both cameras. The uncertainties of $P$ using the ITPC method can be calculated by propagating the uncertainties of the image coordinates in both cameras through partial derivatives:
we will do the precision experiments comparing with stereo system ZED mini with fixed cameras.

where 

\[ u_{l}, u_{u} \]

\[ v_{l}, v_{u} \]

Figure 5 shows the images of point \( P \). The expected pixel coordinates of point \( P \) can be calculated using Equation (25). 

\[ u_{l}, u_{u} \]

\[ v_{l}, v_{u} \]

\[ \begin{align*}
    u_{x} &= \sqrt{(\frac{\partial x_{p}}{\partial u_{l}})^2 u_{u}^2 + (\frac{\partial x_{p}}{\partial v_{l}})^2 v_{u}^2 + (\frac{\partial x_{p}}{\partial u_{r}})^2 u_{r}^2 + (\frac{\partial x_{p}}{\partial v_{r}})^2 v_{r}^2}, \\
    u_{y} &= \sqrt{(\frac{\partial y_{p}}{\partial u_{l}})^2 u_{u}^2 + (\frac{\partial y_{p}}{\partial v_{l}})^2 v_{u}^2 + (\frac{\partial y_{p}}{\partial u_{r}})^2 u_{r}^2 + (\frac{\partial y_{p}}{\partial v_{r}})^2 v_{r}^2}, \\
    u_{z} &= \sqrt{(\frac{\partial z_{p}}{\partial u_{l}})^2 u_{u}^2 + (\frac{\partial z_{p}}{\partial v_{l}})^2 v_{u}^2 + (\frac{\partial z_{p}}{\partial u_{r}})^2 u_{r}^2 + (\frac{\partial z_{p}}{\partial v_{r}})^2 v_{r}^2},
\end{align*} \]

5. Experiments and Results

We will first compare the triangulation performance of our method with ITPC method in simulated and physical experiments in order to evaluate the eye gaze based 3D triangulation method. After that, we will do the precision experiments comparing with stereo system ZED mini with fixed cameras. Finally, we will execute experiments to obtain the time response of our method.

5.1. Comparison Experiments with the ITPC Method

We will compare our proposed method with the ITPC method in both simulated and physical experiments. Static target points as well as moving target points will be triangulated to see the performance of the two methods.

5.1.1. Simulated Experiments

We will do the two comparison experiment in simulation environments. The first is triangulation of static target points and the other is triangulation of a moving target point.

Experimental Setup

In simulated experiments, we use GAZEBO [53] in order to simulate the kinematics of the robotic bionic eyes with the same mechanical model as the physical robotic bionic eyes. We also use the same joint controller under the ROS framework to send joint velocity commands to GAZEBO and get the joint angle feedback as well as image feedback from both eyes.

The image size of the left and right cameras is 1040 × 860 in pixel. The intrinsic parameters obtained while using Bouguet toolbox [54] are shown in Table 2. The checkerboard corner points are detected with sub-pixel position accuracy. We choose the point at the upper right corner as the spatial point \( P \). The expected pixel coordinates of point \( P \) in both cameras are fixed at the principal point (520.50, 430.50). Figure 5 shows the images of \( P \) in the left and right cameras.

| Camera | \( f_u/\text{Pixel} \) | \( f_v/\text{Pixel} \) | \( u_0/\text{Pixel} \) | \( v_0/\text{Pixel} \) |
|--------|----------------------|----------------------|----------------------|----------------------|
| Left   | 619.71               | 619.71               | 520.50               | 430.50               |
| Right  | 619.71               | 619.71               | 520.50               | 430.50               |

5.2. Experiments

5.2.1. Triangulation of Static Target Points

In this experiment, we get 21 different static target points by placing the simulated checkerboard at 21 different positions. At each position, we recorded the estimated 3D coordinates w.r.t the base frame \( N \) using our proposed method and ITPC method, respectively. The ground truth can be obtained from GAZEBO (as shown in Table 3).
Figure 5. The images of $P$ in the left and right simulated cameras. (a) left image; (b) right image.

Table 3. The ground truth of the 3D coordinates of the 21 static target points.

| Number | $x$/mm | $y$/mm | $z$/mm | Number | $x$/mm | $y$/mm | $z$/mm |
|--------|--------|--------|--------|--------|--------|--------|--------|
| 1      | 0.00   | -99.81 | 499.98 | 12     | 0.01   | -99.84 | 1599.99|
| 2      | 0.00   | -99.81 | 999.99 | 13     | 0.01   | -99.85 | 1699.99|
| 3      | 0.00   | -99.81 | 999.99 | 14     | 0.01   | -99.85 | 1799.99|
| 4      | 0.00   | -99.82 | 799.99 | 15     | 0.01   | -99.84 | 1899.99|
| 5      | 0.00   | -99.82 | 899.99 | 16     | 0.01   | -99.86 | 1999.99|
| 6      | 0.01   | -99.83 | 999.99 | 17     | 0.01   | -99.86 | 2099.99|
| 7      | 0.00   | -99.83 | 1099.99| 18     | 0.01   | -99.85 | 2199.99|
| 8      | 0.01   | -99.83 | 1199.99| 19     | 0.01   | -99.86 | 2299.99|
| 9      | 0.00   | -99.84 | 1299.99| 20     | 0.01   | -99.86 | 2399.99|
| 10     | 0.01   | -99.84 | 1399.99| 21     | 0.01   | -99.87 | 2499.99|
| 11     | 0.01   | -99.85 | 1499.99|        |        |        |        |

**Error Analysis:** we calculated the absolute errors of the mean of the 3D coordinates using our proposed method and ITPC method (as shown in Figure 6) w.r.t the ground truth, respectively. The absolute error of both methods in the simulated experiments are mainly from the error of D-H parameters, the error of intrinsic parameters and the error of head-eye parameters. The absolute errors in X axis of both methods are between 0.73 mm and 3.96 mm, as shown in Figure 6. The absolute errors in Y axis of both methods are between 1.14 mm and 6.18 mm. We can see that the minimum absolute errors in Z axis of both methods are only 0.02 mm and 0.04 mm, respectively, when Z ground truth is 1100 mm. The maximum absolute errors in Z axis of both methods only reach 0.5% of the ground truth depth. In conclusion, the absolute errors of our proposed method are very close to the ITPC method in the X, Y, and Z axes.

**Uncertainty Analysis:** the uncertainties of our proposed method (shown in Figure 7) were calculated using Equations (31)–(33). The uncertainties of the ITPC method (shown in Figure 7) were calculated using Equations (34)–(36). At each position, the times $n$ of independent repeated observations is 1150. For both of the left and right cameras, the image detection uncertainties in the X direction are between 0.006 pixels and 0.057 pixels. The image detection uncertainties in the Y direction are between 0.006 pixels and 0.057 pixels. The absolute uncertainties of our proposed method are very close to the ITPC method in the X, Y and Z axes, as shown in Figure 7. The absolute uncertainty in X axis of both methods are between 0.009 mm and 0.130 mm. The absolute uncertainty in Y axis of both methods are between 0.028 mm and 0.188 mm. The absolute uncertainty in Z axis of both methods are between 0.088 mm and 4.545 mm.
Figure 6. The absolute error of the mean of the estimated 3D coordinates using our proposed method and integrated two-pose calibration (ITPC) method w.r.t ground truth respectively. (a) absolute error in X axis; (b) absolute error in Y axis; and,(c) absolute error in Z axis.

Triangulation of a Moving Target Point

We want to verify the effectiveness of our proposed method in the case that the target point is moving.

We placed a checkerboard at 1000 mm on the Z axis w.r.t the base frame N. The trajectory of the target point on the X and Y axes w.r.t frame N is \( x = r \times \sin(t/10000) \) and \( y = -100 + r \times \cos(t/10000) \), respectively, where \( r = 100 \) mm. We recorded the 3D coordinates of the moving target points estimated by our proposed method and ITPC method, respectively.

Error Analysis: we calculated the absolute error of our proposed method and ITPC method w.r.t ground truth, respectively. The mean absolute error in X axis of both methods are 1.76 mm and 1.61 mm, respectively, as shown in Figure 8. The mean absolute error in Y axis of both methods are 2.03 mm and 2.17 mm, respectively. The mean absolute error in Z axis of both methods are 1.69 mm and 1.71 mm which reach only 0.17% of the ground truth depth. The mean absolute error in X axis of our proposed method is larger than the ITPC method and the mean absolute error in Y, Z axes of the proposed method are smaller than the ITPC method.
Figure 7. The uncertainties of our proposed method and ITPC method. (a) absolute uncertainty in $X$ axis; (b) absolute uncertainty in $Y$ axis; and, (c) absolute uncertainty in $Z$ axis.

**Uncertainty Analysis:** we calculated the standard deviation of the coordinates in $Z$ axis as the uncertainties of our proposed method and ITPC method. In the experiments, the number $n$ is 3780. The absolute uncertainties of our proposed method and the ITPC method in $Z$ axis are 0.30 mm and 0.98 mm, respectively. The absolute uncertainties in $Z$ axis of our proposed method are smaller than the ITPC method.

5.1.2. Physical Experiments

The physical experiments are performed on the real robotic bionic eyes. Triangulation on static and moving target points is performed. Comparisons are carried out between our proposed method and the ITPC method.

Experimental Setup

In physical experiments, the image size of the left and right cameras is $640 \times 480$ in pixel. The intrinsic parameters of both cameras are shown in Table 4. The distortion coefficients of the left and right cameras are $[-0.0437, 0.1425, 0.0005, -0.0012, 0.0000]$ and $[-0.0425, 0.1080, 0.0001, -0.0015, 0.0000]$. 
0.0000], respectively. The maximum horizontal and vertical field of view of the real robotic bionic eyes are 170° and 170°, respectively. The AprilTag [55] information is obtained through ViSP [56] library. We choose the center (point $P$) of the AprilTag as the fixation point. The expected image of point $P$ in the left and right cameras are set to the principal points (362.94, 222.53) and (388.09, 220.82), respectively. Figure 9 shows the images of $P$ in the left and right cameras.

![Image](image.png)

**Figure 8.** The absolute error of the mean of the estimated 3D coordinates using our proposed method and ITPC method w.r.t ground truth, respectively. (a) absolute error in $X$ axis; (b) absolute error in $Y$ axis; and, (c) absolute error in $Z$ axis.

**Table 4.** The intrinsic parameters of the left and right cameras of the designed robotic bionic eyes.

| Camera | $f_u$/Pixel | $f_v$/Pixel | $u_0$/Pixel | $v_0$/Pixel |
|--------|-------------|-------------|-------------|-------------|
| Left   | 521.66      | 521.83      | 362.94      | 222.53      |
| Right  | 521.20      | 521.51      | 388.09      | 220.82      |
Figure 9. The images of the left and right cameras. (a) left image; (b) right image.

Triangulation of Static Target Points

We placed the AprilTag at 21 different positions (as shown in Table 5). At each position, the estimated 3D coordinates using our proposed method and ITPC method were recorded, respectively.

| Number | x/mm  | y/mm  | z/mm  | Number | x/mm  | y/mm  | z/mm  |
|--------|-------|-------|-------|--------|-------|-------|-------|
| 1      | −208.73 | −177.64 | 508.04 | 12     | −721.54 | −180.49 | 1603.42 |
| 2      | −232.55 | −177.53 | 605.69 | 13     | −754.63 | −182.21 | 1704.31 |
| 3      | −255.58 | −176.02 | 699.53 | 14     | −779.88 | −181.77 | 1798.82 |
| 4      | −324.48 | −179.81 | 800.92 | 15     | −816.46 | −182.52 | 1909.67 |
| 5      | −407.39 | −178.42 | 899.88 | 16     | −842.48 | −182.31 | 2004.67 |
| 6      | −452.88 | −179.90 | 1001.11 | 17     | −909.75 | −171.71 | 2097.57 |
| 7      | −486.87 | −180.58 | 1102.26 | 18     | −984.94 | −168.96 | 2200.87 |
| 8      | −563.40 | −179.15 | 1198.49 | 19     | −1029.85 | −171.79 | 2296.25 |
| 9      | −607.89 | −181.18 | 1308.29 | 20     | −1144.02 | −168.11 | 2407.59 |
| 10     | −649.50 | −179.74 | 1404.37 | 21     | −1229.31 | −166.47 | 2506.59 |
| 11     | −688.94 | −182.54 | 1501.91 |        |        |        |       |

Table 5. The ground truth of the static target points.

Error Analysis: we calculated the absolute error of the mean of estimated 3D coordinates w.r.t ground truth using our proposed method and ITPC method, respectively (as shown in Figure 10). The absolute error of both methods in the physical experiments are mainly from errors of forward kinematics, errors of intrinsic parameter, errors of head-eye parameters and joint offset errors. As shown in Figure 10, our proposed method are closer to the ground truth than the ITPC methods in the X and Z axes in most of the 21 different positions, especially when Z ground truth are larger than 2000 mm. The absolute error in Y axis of our proposed method is between 0.55 mm and 12.28 mm. The absolute error in Y axis of the ITPC method are between 1.52 mm and 12.88 mm. The minimum absolute errors in Z axis of our proposed method and the ITPC method are 1.42 mm and 2.32 mm, which reach 0.23% and 0.38% of the ground truth depth, respectively. The maximum absolute error in Z axis of our proposed method and the ITPC method are 124.49 mm and 174.64 mm, which reach 4.97% and 6.98% of the ground truth depth, respectively. Our proposed method obtains smaller mean absolute errors in the X, Y, and Z axes.
Figure 10. The absolute error of the mean of the estimated 3D coordinates using our proposed method and ITPC method w.r.t ground truth respectively. (a) absolute error in X axis; (b) absolute error in Y axis; and, (c) absolute error in Z axis.

Uncertainty Analysis: the uncertainties of our proposed method (shown in Figure 11) were calculated using Equations (31)–(33). The uncertainties of the ITPC method (shown in Figure 11) were calculated using Equations (34)–(36). At each position, \( n = 1200 \). For the left camera, the image detection uncertainties in the X direction are between 0.027 pixels and 0.089 pixels. The image detection uncertainties in the Y direction are between 0.039 pixels and 0.138 pixels. For the right camera, the image detection uncertainties in the X direction are between 0.016 pixels and 0.067 pixels. The image detection uncertainties in the Y direction are between 0.046 pixels and 0.138 pixels. The absolute uncertainty in X axis of both methods are between 0.119 mm and 3.905 mm, as shown in Figure 11. The absolute uncertainty in Y axis of both methods are between 0.091 mm and 0.640 mm. The absolute uncertainty in Z axis of both methods are between 0.268 mm and 7.975 mm. In conclusion, the absolute uncertainties of our proposed method are very close to the ITPC method in the X, Y, and Z axes.
Figure 11. The uncertainties of our proposed method and ITPC method. (a) absolute uncertainty in X axis; (b) absolute uncertainty in Y axis; and, (c) absolute uncertainty in Z axis.

Triangulation of a Moving Target Point

We move the target point of the AprilTag from \((-278.60, -23.13, 957.84)\) to \((-390.02, -30.39, 1111.91)\) at the average speed of 0.01 m/s. Figure 12 shows the trajectory of the ground truth and the estimated \(P\) using our proposed method and ITPC method.

Figure 12. The trajectory of the ground truth and the estimated point \(P\) using our proposed method and ITPC method.
**Error Analysis:** we calculated the absolute error of our proposed method and ITPC method w.r.t ground truth respectively (shown in Figure 13). The mean absolute error in X axis of our proposed method and the ITPC method are 11.81 mm and 9.53 mm, as shown in Figure 13. The mean absolute error in Y axis of our proposed method and the ITPC method are 14.89 mm and 16.59 mm. The mean absolute error in Z axis of our proposed method and the ITPC method are 23.74 mm and 22.48 mm. The mean absolute error in X, Z axes of the proposed method are larger than the ITPC method, and the mean absolute error in Y axis of our proposed method is smaller than the ITPC method.

![Figure 13](image-url)  
**Figure 13.** The absolute error of our proposed method and ITPC method w.r.t ground truth respectively. (a) absolute error in X axis; (b) absolute error in Y axis; and, (c) absolute error in Z axis.

### 5.2. Comparison Experiments with Zed Mini

We compare the triangulation performance of our proposed method and ITPC method with stereo system ZED Mini while using fixed cameras (shown in Figure 2).

**Experimental Setup**

The image size of the left and right cameras of ZED mini are 1280 × 720 pixels. The intrinsic parameters including the focal length and principal point of both cameras are shown in Table 6.
The maximum horizontal and vertical field of view of ZED mini are $90^\circ$ and $60^\circ$, respectively. We placed the AprilTag at eight different positions to get 8 different spatial points. The estimated 3D coordinates using ZED mini are shown in Table 7.

**Error analysis:** we calculated the absolute error of our proposed method and ITPC method w.r.t ZED mini, respectively. As shown in Figure 14, the absolute errors of our proposed method are very close to the ITPC method in the X and Y axes. Our proposed method gets smaller absolute error in the Z axis at most of the eight different positions. The absolute error in X axis of both methods are between 25.17 mm and 68.63 mm. The absolute error in Y axis of both methods are between 35.26 mm and 55.18 mm. The minimum absolute errors in Z axis of our proposed method and ITPC method are 0.44 mm and 2.10 mm. The maximum absolute errors in Z axis of our proposed method and ITPC method are 86.40 mm and 92.90 mm. No camera movements exist in ZED mini. Our proposed method and ITPC method get a minimally larger errors when comparing with ZED mini mainly because of eye motion.

![Figure 14](image_url)

**Figure 14.** The absolute error of our proposed method and ITPC method w.r.t ZED Mini. (a) absolute error in X axis; (b) absolute error in Y axis; and, (c) absolute error in Z axis.
Table 6. The intrinsic parameters of the left and right cameras of ZED mini.

| Camera | \( f_u/\text{Pixel} \) | \( f_v/\text{Pixel} \) | \( u_0/\text{Pixel} \) | \( v_0/\text{Pixel} \) |
|--------|-------------------|-------------------|-------------------|-------------------|
| Left   | 676.21            | 676.21            | 645.09            | 368.27            |
| Right  | 676.21            | 676.21            | 645.09            | 368.27            |

Table 7. The estimated 3D coordinates of the spatial points w.r.t the base frame \( O_N-x_Ny_Nz_N \) using ZED mini.

| Point | \( x/\text{mm} \) | \( y/\text{mm} \) | \( z/\text{mm} \) | Point | \( x/\text{mm} \) | \( y/\text{mm} \) | \( z/\text{mm} \) |
|-------|-----------------|-----------------|-----------------|-------|-----------------|-----------------|-----------------|
| 1     | −52.97          | −52.35          | 995.14          | 5     | −53.83          | −55.49          | 1493.80         |
| 2     | −51.75          | −104.14         | 1004.50         | 6     | −47.30          | −93.07          | 1498.50         |
| 3     | −106.29         | −49.44          | 1002.60         | 7     | −99.29          | −55.39          | 1478.30         |
| 4     | −103.80         | −112.04         | 989.98          | 8     | −104.57         | −99.60          | 1495.60         |

5.3. Time Performance Experiments

5.3.1. Simulated Experiments

We perform experiments to evaluate how \( k_1 \) and \( k_2 \) in the constant coefficient matrix \( K \) affect the time it takes for eye gazing. We initially placed the simulated checkerboard to set the target point at \((-0.00, -100.00, 1000)\). A step signal with amplitude of 100 mm in \( Y \) direction was applied to the target point. We adjusted \( k_1 \) and \( k_2 \) from 0.5 to 4.0. The system overshot when \( k_1 \) or \( k_2 \) is larger than 4.0. Increasing \( k_1 \) and \( k_2 \) moderately can shorten the time it takes for eye gazing, as shown in Figure 15.

![Figure 15](image-url)  
Figure 15. The time it takes for eye gazing affected by \( k_1 \) and \( k_2 \).

5.3.2. Physical Experiments

In the physical experiments, we placed the AprilTag to set the target point at \((-0.00, -145.00, 1500)\). We set the robotic bionic eyes to the initial state. We set \( k_1 = 4.0, k_2 = 4.0 \) in \( K \). It takes 650 ms to move the target point \( P \) from \((385.98, 196.60)\) to \((362.94, 222.53)\) in the left image frame, and from \((361.09, 191.82)\) to \((388.09, 220.82)\) in the right image frame, respectively, through eye gazing.

6. Conclusions

In this paper, we have proposed an eye gaze based 3D triangulation method for our designed robotic bionic eyes. The eye gaze was realized through image based visual servoing in order to keep
the two optical axes pass through the target point $P$. We could obtain the 3D coordinates of $P$ through the intersection of the two optical axes of both cameras. The optical axes of both cameras could be derived from forward kinematics with head-eye parameters calibrated beforehand. In real applications, the two optical axes may not intersect each other due to visual servoing errors and model errors, and so we use the middle point of the common perpendicular line segment of the two skew optical axes as the representation of the intersection point $P$.

From the simulated and physical experiments, we can see that the proposed method achieves comparable results with the ITPC method in the absolute errors and the propagated uncertainties, and our proposed method gets smaller mean absolute errors with the triangulation of static target points in physical experiments. Our proposed method and ITPC method get larger errors w.r.t conventional stereo systems with fixed cameras such as ZED Mini due to model errors introduced by manufacturing which include link length error, coaxiality error and error due to link stiffness. The experiments show that at the beginning of the visual servoing process our method need several hundred milliseconds to locate a target point to its optical center. Selecting $k_1$ and $k_2$ in the gain matrix $K$ by using fuzzy PID could be potential solution to minimize the initial processing time to locate a target point to its optical center.

Although our method has only tiny improvement in triangulation precision compare with the ITPC method, it is a new bionic approach for triangulation using eye gazing through image based visual servoing. Our system has much larger field of view than traditional stereo pair, such as ZED Mini, and it does not rely on stereo extrinsic parameters directly. Another advantage is that image detection error tolerance is higher. In the future, we are going to reduce the model error and joint offset error introduced by manufacturing to obtain compatible precision as stereo pair with fixed cameras.

**Author Contributions:** Conceptualization, X.C.; Data curation, X.L.; Formal analysis, D.F. and X.C.; Investigation, F.M.; Methodology, D.F. and Y.L. (Yunhui Liu); Software, D.F. and Z.U.; Supervision, Y.L. (Yunhui Liu) and Q.H.; Validation, Y.L. (Yanyang Liu), F.M. and W.C.; Visualization, Y.L (Yanyang Liu), and X.L.; Writing—original draft, D.F.; Writing—review and editing, X.C., Z.U., Y.L. (Yunhui Liu) and Q.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (Grant No. 91748202); the Science and Technology Program of Beijing Municipal Science and Technology Commission (Z191100008019003); the Pre-research Project under Grant 41412040101; and the Open Project Fund of Hebei Static Traffic Technology Innovation Center, AIPARK (No.001/2020).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Sangeetha, G.; Kumar, N.; Hari, P.; Sasikumar, S. Implementation of a stereo vision based system for visual feedback control of robotic arm for space manipulations. *Procedia Comput. Sci.* 2018, 133, 1066–1073.

2. Lin, C.-Y.; Shih, S.-W.; Hung, Y.-P.; Tang, G.Y. A new approach to automatic reconstruction of a 3-d world using active stereo vision. *Comput. Vis. Image Underst.* 2002, 85, 117–143. [CrossRef]

3. Al-Mutib, K.N.; Mattar, E.A.; Alsuilaiman, M.M.; Ramdane, H. Stereo vision slam based indoor autonomous mobile robot navigation. In Proceedings of the 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), Bali, Indonesia, 5–10 December 2014; pp. 1584–1589.

4. Diebel, J.; Reutersward, K.; Thrun, S.; Davis, J.; Gupta, R. Simultaneous localization and mapping with active stereo vision. In Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No. 04CH37566), Sendai, Japan, 28 September–2 October 2004; Volume 4, pp. 3436–3443.

5. Sakagami, Y.; Watanabe, R.; Aoyama, C.; Matsunaga, S.; Higaki, N.; Fujimura, K. The intelligent asimo: System overview and integration. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Lausanne, Switzerland, 30 September–4 October 2002; Volume 3, pp. 2478–2483.

6. Kaneko, K.; Harada, K.; Kanehiro, F.; Miyamori, G.; Akachi, K. Humanoid robot hrp-3. In Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, 22–26 September 2008; pp. 2471–2478.
7. Zhang, T.; Uchiyama, E.; Nakamura, Y. Dense rgb-d slam for humanoid robots in the dynamic humans environment. In Proceedings of the 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids), Beijing, China, 6–9 November 2018; pp. 270–276.

8. Willow Garage. Available online: https://en.wikipedia.org/wiki/Willow_Garage (accessed on 1 September 2018).

9. Atlas (Robot). Available online: https://en.wikipedia.org/wiki/Atlas_(robot) (accessed on 1 September 2018).

10. Tsagarakis, N.G.; Caldwell, D.G.; Negrello, F.; Choi, W.; Baccelliere, L.; Loc, V.-G.; Noorden, J.; Muratore, L.; Margan, A.; Cardellino, A.; et al. Walk-man: A high-performance humanoid platform for realistic environments. J. Field Robot. 2017, 34, 1225–1259. [CrossRef]

11. Berthouze, L.; Bakker, P.; Kuniyoshi, Y. Learning of oculo-motor control: A Prelude to robotic imitation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS’96, Osaka, Japan, 8 November 1996; Volume 1, pp. 376–381.

12. Sharkey, P.M.; Murray, D.W.; McLauchlan, P.F.; Brooker, J.P. Hardware development of the yorick series of active vision systems. Microprocess. Microsyst. 1998, 21, 363–375. [CrossRef]

13. Biamino, D.; Cannata, G.; Maggiali, M.; Piazza, A. Mac-eye: A tendon driven fully embedded robot eye. In Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan, 5 December 2005; pp. 62–67.

14. Beira, R.; Lopes, M.; Praça, M.; Santos-Victor, J.; Bernardino, A.; Metta, G.; Becchi, F.; Saltarén, R. Design of the robot-cub (icub) head. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation, Orlando, FL, USA, 15–19 May 2006; pp. 94–100.

15. Asfour, T.; Welke, K.; Azad, P.; Ude, A.; Dillmann, R. The karlsruhe humanoid head. In Proceedings of the Humanoids 2008-8th IEEE-RAS International Conference on Humanoid Robots, Daejeon, Korea, 1–3 December 2008; pp. 447–453.

16. Lütkebohle, I.; Hegel, F.; Schulz, S.; Hackel, M.; Wrede, B.; Wachsmuth, S.; Sagerer, G. The bielefeld anthropomorphic robot head flobi. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 3–7 May 2010; pp. 3384–3391.

17. Song, Y.; Zhang, X. An active binocular integrated system for intelligent robot vision. In Proceedings of the 2012 IEEE International Conference on Intelligence and Security Informatics, Arlington, VA, USA, 11–14 June 2012; pp. 48–53.

18. Asfour, T.; Schill, J.; Peters, H.; Klas, C.; Bücker, J.; Sander, C.; Schulz, S.; Kargow, A.; Werner, T.; Bartenbach, V. Armar-4: A 63 dof torque controlled humanoid robot. In Proceedings of the 2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids), Atlanta, GA, USA, 15–17 October 2013; pp. 390–396.

19. Cid, F.; Moreno, J.; Bustos, P.; Núñez, P. Muecas: A multi-sensor robotic head for affective human robot interaction and imitation. Sensors 2014, 14, 7711–7737. [CrossRef]

20. Pateromichelakis, N.; Mazel, A.; Hache, M.; Koumpogiannis, T.; Gelin, R.; Maisonnier, B.; Berthoz, A. Head-eyes system and gaze analysis of the humanoid robot romeo. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14–18 September 2014; pp. 1374–1379.

21. Penčič, M.; Rackov, M.; Čavić, M.; Kiss, I.; Cioată, V. Social humanoid robot sara: Development of the wrist mechanism. In IOP Conference Series: Materials Science and Engineering; IOP Publishing: Bristol, UK, 2018; Volume 294, p. 012079.

22. Zhang, T.; Chen, X.; Owais, H.M.; Liu, G.; Fu, S.; Tian, Y.; Chen, X. Multi-loop stabilization control of a robotic bionic eyes. In Proceedings of the 2017 IEEE International Conference on Cyborg and Bionic Systems (CBS), Beijing, China, 17–19 October 2017; pp. 87–90.

23. Liu, G.; Owais, H.M.; Zhang, T.; Fu, S.; Tian, Y.; Chen, X. Reliable eyes pose measurement for robotic bionic eyes with mems gyroscope and akf filter. In Proceedings of the 2017 IEEE International Conference on Cyborg and Bionic Systems (CBS), Beijing, China, 17–19 October 2017; pp. 83–86.

24. Fan, D.; Chen, X.; Zhang, T.; Chen, X.; Liu, G.; Owais, H.M.; Kim, H.; Tian, Y.; Zhang, W.; Huang, Q. Design of anthropomorphic robot bionic eyes. In Proceedings of the 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), Macau, China, 5–8 December 2017; pp. 2050–2056.

25. Majumdar, J. Efficient parallel processing for depth calculation using stereo. Robot. Autonomous Syst. 1997, 20, 1–13. [CrossRef]

26. Nefiti-Meziani, S.; Manzoor, U.; Davis, S.; Pupala, S.K. 3D perception from binocular vision for a low cost humanoid robot nao. Robot. Autonomous Syst. 2015, 68, 129–139. [CrossRef]
27. Zhang, J.; Du, R.; Gao, R. Passive 3D reconstruction based on binocular vision. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP 2018), International Society for Optics and Photonics, Chengdu, China, 12–14 December 2018; Volume 11069, p. 110690Y.

28. Hartley, R.I.; Sturm, P. Triangulation. *Comput. Vision Image Underst.* 1997, 68, 146–157. [CrossRef]

29. Kanatani, K.; Sugaya, Y.; Niitsuma, H. Triangulation from two views revisited: Hartley-sturm vs. optimal correction. *Practice* 2008, 4, 5.

30. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.

31. Zhong, F.; Shao, X.; Quan, C. A comparative study of 3d reconstruction methods in stereo digital image correlation. *Opt. Lasers Eng.* 2019, 122, 142–150. [CrossRef]

32. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 2000, 22, 1330–1334. [CrossRef]

33. Cui, Y.; Zhou, F.; Wang, Y.; Liu, L.; Gao, H. Precise calibration of binocular vision system used for vision measurement. *Opt. Express* 2014, 22, 9134–9149. [CrossRef] [PubMed]

34. Neubert, J.; Ferrier, N.J. Robust active stereo calibration. In Proceedings of the 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292), Washington, DC, USA, 11–15 May 2002; Volume 3, pp. 2525–2531.

35. Xu, D.; Li, Y.F.; Tan, M.; Shen, Y. A new active visual system for humanoid robots. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* 2008, 38, 320–330.

36. Xu, D.; Wang, Q. A new vision measurement method based on active object gazing. *Int. J. Adv. Robot. Syst.* 2017, 14, 1729881417715984. [CrossRef]

37. Wang, Y.; Wang, X.; Wan, Z.; Zhang, J. A method for extrinsic parameter calibration of rotating binocular stereo vision using a single feature point. *Sensors* 2018, 18, 3666. [CrossRef]

38. Wang, Y.; Wang, X. An improved two-point calibration method for stereo vision with rotating cameras in large fov. *J. Modern Opt.* 2019, 66, 1106–1115. [CrossRef]

39. Li, M. Kinematic calibration of an active head-eye system. *IEEE Trans. Robot. Autom.* 1998, 14, 153–158.

40. Chen, X.; Wang, C.; Zhang, W.; Lan, K.; Huang, Q. An integrated two-pose calibration method for estimating head-eye parameters of a robotic bionic eye. *IEEE Trans. Instrum. Meas.* 2019, 68, 1664–1672. [CrossRef]

41. Dang, T.; Hoffmann, C.; Stiller, C. Continuous stereo self-calibration by camera parameter tracking. *IEEE Trans. Image Process.* 2009, 18, 1536–1550. [CrossRef]

42. Paul, R.P. *Robot Manipulators: Mathematics, Programming, and Control: The Computer Control of Robot Manipulators*; MIT Press: Cambridge, MA, USA, 1981.

43. Coombs, D.J.; Brown, C.M. Cooperative gaze holding in binocular vision. *IEEE Control Syst. Mag.* 1991, 11, 24–33.

44. Tanaka, M.; Maru, N.; Miyazaki, F. Binocular gaze holding of a moving object with the active stereo vision system. In Proceedings of the 1994 IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 5–7 December 1994; pp. 250–255.

45. Roca, X.; Vitria, J.; Vanrell, M.; Villanueva, J. Gaze control in a binocular robot systems. In Proceedings of the 1999 7th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA’99 (Cat. No. 99TH8467), Barcelona, Spain, 18–21 October 1999; Volume 1, pp. 479–485.

46. Satoh, Y.; Okatani, T.; Deguchi, K. Binocular motion tracking by gaze fixation control and three-dimensional shape reconstruction. *Adv. Robot.* 2003, 17, 1057–1072. [CrossRef]

47. Hutchinson, S.; Chaumette, F. Visual servo control, part I: Basic approaches. *IEEE Robot. Autom. Mag.* 2006, 13, 82–90.

48. Muelaner, J.E.; Wang, Z.; Martin, O.; Jamshidi, J.; Maropoulos, P.G. Estimation of uncertainty in three-dimensional coordinate measurement by comparison with calibrated points. *Meas. Sci. Technol.* 2010, 21, 025106. [CrossRef]

49. da Silva Hack, P.; ten Caten, C.S. Measurement uncertainty: Literature review and research trends. *IEEE Trans. Instrum. Meas.* 2012, 61, 2116–2124. [CrossRef]

50. BIPM, I.; IFCC, I.; IUPAC, I.; ISO, O. Evaluation of measurement data—Guide for the expression of uncertainty in measurement. Joint Committee for Guides in Metrology (JCGM) 100: 2008. *Citado en las* 2008, 18–21.
51. Leo, G.D.; Liguori, C.; Paolillo, A. Propagation of uncertainty through stereo triangulation. In Proceedings of the 2010 IEEE Instrumentation & Measurement Technology Conference Proceedings, Austin, TX, USA, 3–6 May 2010; pp. 12–17.

52. Leo, G.D.; Liguori, C.; Paolillo, A. Covariance propagation for the uncertainty estimation in stereo vision. *IEEE Trans. Instrum. Meas.* **2011**, *60*, 1664–1673. [CrossRef]

53. Koenig, N.; Howard, A. Design and use paradigms for gazebo, an open-source multi-robot simulator. In Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No. 04CH37566), Sendai, Japan, 28 September–2 October 2004; Volume 3, pp. 2149–2154.

54. Camera Calibration Toolbox for Matlab. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/ (accessed on 1 September 2018).

55. Olson, E. Apriltag: A robust and flexible visual fiducial system. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3400–3407.

56. Marchand, É.; Spindler, F.; Chaumette, F. Visp for visual servoing: A generic software platform with a wide class of robot control skills. *IEEE Robot. Autom. Mag.* **2005**, *12*, 40–52. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).