sensors

Article

A Method for Extrinsic Parameter Calibration of Rotating Binocular Stereo Vision Using a Single Feature Point

Yue Wang [1,2], **Xiangjun Wang** [1,2,*], **Zijing Wan** [1,2] and **Jiahao Zhang** [1,2]

[1] State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China; wy2012110040@163.com (Y.W.); zjwan@tju.edu.cn (Z.W.); q17695949504@163.com (J.Z.)

[2] Key Laboratory of MOEMS of the Ministry of Education, Tianjin University, Tianjin 300072, China

[*] Correspondence: tjuxjw@126.com; Tel.: +86-139-2066-9580

Abstract: Nowadays, binocular stereo vision (BSV) is extensively used in real-time 3D reconstruction, which requires cameras to quickly implement self-calibration. At present, the camera parameters are typically estimated through iterative optimization. The calibration accuracy is high, but the process is time consuming. Hence, a system of BSV with rotating and non-zooming cameras is established in this study, in which the cameras can rotate horizontally and vertically. The cameras' intrinsic parameters and initial position are estimated in advance by using Zhang's calibration method. Only the yaw rotation angle in the horizontal direction and pitch in the vertical direction for each camera should be obtained during rotation. Therefore, we present a novel self-calibration method by using a single feature point and transform the imaging model of the pitch and yaw into a quadratic equation of the tangent value of the pitch. The closed-form solutions of the pitch and yaw can be obtained with known approximate values, which avoid the iterative convergence problem. Computer simulation and physical experiments prove the feasibility of the proposed method. Additionally, we compare the proposed method with Zhang's method. Our experimental data indicate that the averages of the absolute errors of the Euler angles and translation vectors relative to the reference values are less than $0.21°$ and 6.6 mm, respectively, and the averages of the relative errors of 3D reconstruction coordinates do not exceed 4.2%.

Keywords: extrinsic parameter calibration; binocular stereo vision; single feature point

1. Introduction

In recent years, with the advancement in computer vision and image technology, binocular stereo vision has been extensively used in three-dimensional (3D) reconstruction, navigation, and video surveillance. This vision method requires that the cameras possess a higher calibration accuracy and better real-time calibration to satisfy the requirements of practical engineering applications. Therefore, research on camera calibration technology of binocular stereo vision is important both theoretically and practically.

Camera calibration technology is primarily divided according to traditional [1–3] and self-calibration [4–6] methods. Traditional calibration methods require precision-machined targets, which employ the known 3D world coordinates of control points and their image coordinates to calculate the cameras' intrinsic and extrinsic parameters. These methods include the direct linear transformation (DLT) [7], Tsai's radial alignment constraint (RAC) [8], Wen's iterative calibration [9], and double-plane calibration [10]. The traditional methods have high precision, but their algorithm is complex and the calibration process is time consuming and laborious. Hence, its practical application

will be significantly limited. In the 1990s, Faugeras and Maybank [11] first proposed the idea of self-calibrating cameras. The self-calibration methods only require the constraint relation from the image sequence without the aid of a calibration block, possibly allowing the camera parameters to be obtained online and in real time. This method can satisfy the requirements in some special occasions where the cameras' focal length should often be adjusted or the cameras' position will move according to the surrounding environment.

In 1992, Faugeras proposed a self-calibration method by directly solving the Kruppa equation [12], which is computationally complex and sensitive to noise. Given the difficulty in solving the Kruppa equation, Hartley et al. proposed a QR decomposition method in 1994 by using the hierarchical step-by-step calibration [13]. The projection matrix is decomposed by QR, but the method requires the initial value for an effective calibration. Triggs et al. proposed in 1997 a camera self-calibration method based on an absolute quadratic surface [14], which is more effective than the method based on the Kruppa equation with regard to inputting multiple images. The camera self-calibration method based on an active vision system, which controls the camera to perform special motions and uses multiple images captured from various positions to calibrate the camera, is represented using the linear method according to two groups of three-direction orthogonal motions proposed by Ma in 1996 [15]. Generally, the self-calibration methods are poor robustness, and calibration accuracy is lower than that of traditional methods. However, the self-calibration method can effectively utilize various constraints unrelated to the motion of the cameras and some prior knowledge, and render the algorithm more simple and practical. Therefore, the self-calibration method is more flexible and can be utilized in a wider range of applications.

In this work, we primarily investigate the implementation of the camera self-calibration. Hitherto, researchers have performed numerous related works. The typical self-calibration algorithms are to estimate camera parameters through iterative optimization with numerous matching features. For example, Zhang proposed a local–global hybrid iterative optimization method by using the bundle adjustment algorithm and SIFT points matching relationship [16]. Wu proposed a complete model for a pan–tilt–zoom (PTZ) camera. The camera parameters can be quickly and accurately estimated using a series of simple initialization steps followed by a nonlinear optimization from ten images [17]. Junejo provided a novel optimized calibration method for cameras with pure rotation or pan–tilt rotation from two images [18]. These methods presented good calibration accuracy, but were time consuming and unable to implement online calibration. Presently, most camera self-calibration algorithms utilize various constraints in natural scenes. Echigo presented a camera calibration method by using three sets of parallel lines in which the rotation parameters were decoupled from the translation parameters [19]. Song established a calibration method for a dynamic PTZ camera overlooking a traffic scene, which automatically used a set of parallel lane markings and the lane width to compute the focal length, tilt angle, and pan angle [20]. Schoepflin presented a new three-stage algorithm to calibrate roadside traffic management cameras by estimating the lane boundaries and the vanishing point of the lines along the roadway [21]. Kim and Hong proposed a nonlinear self-calibration method of rotating and zooming cameras by using inter-image homography from refined matching lines [22]. However, these methods are suitable for the camera calibration under static conditions, which will not work if the special markers like parallel lines are missing during the rotation of the cameras. In order to solve this problem, Muñoz Rodríguez performed binocular self-calibration by means of an adaptive genetic algorithm based on a laser line [23] and presented an online self-camera orientation for mobile vision based on laser metrology and computer algorithms [24], which avoided calibrated references and physical measurements. Self-calibration algorithms were also performed well by combining with the active vision measurement method. Ji introduced a new rotation-based camera self-calibration method that requires the camera to rotate around an unknown but fixed axis twice [25]. Cai proposed a simple method to obtain the camera intrinsic parameters by observing one planar pattern in at least three different orientations [26]. These methods improved traditional self-calibration algorithm, but could not satisfy the requirements of real-time calibration.

In order to achieve fast calibration of the camera, several methods simplified the camera model. Tang proposed a non-iterative self-calibration algorithm for upgrading the projective space to the Euclidean space [27], which combined the typically used metric constraints, including zero skew and unit aspect ratio constant principal points. Yu [28] presented a self-calibration method for moving stereo cameras and introduced some reasonable assumptions, such as principle points located at the image center, camera axis perpendicular to the camera plane, and constant focal length of stereo cameras during movement, to simplify the camera model. De Agapito proposed a linear self-calibration method for a stationary but rotating camera under the minimal assumption of zero skew, known pixel aspect ratio, and known principal point [29]. Once the camera model is simplified, the number of feature points for calibration can be reduced. Sun [30] proposed a novel and effective self-calibration approach for robot vision, which both the camera intrinsic parameters and the hand-eye transformation were estimated by using only two arbitrary feature points. Chen put forward a two-point calibration method for soccer cameras in a narrow field of view, in which only two point correspondences were required given the prior knowledge of base location and orientation of a PTZ camera [31]. The above two methods are rapid and avoid the convergence problems of iterative algorithms.

In this study, we present a novel self-calibration method of BSV with rotating and non-zooming cameras by using a single feature point. The intrinsic parameters of the left and right cameras are estimated in advance by using Zhang's method, as well as the rotation matrix and translation vector at the initial position. The left and right cameras can rotate in the horizontal and vertical directions. Thus, only the two rotation angles for each camera, denoted by pitch and yaw, must be calculated after the rotation. According to the homography of the feature point before and after the rotation, one quadratic equation for the tangent value of the pitch for each camera can be derived. The closed-form solutions of the pitch and yaw can be obtained with known approximate values of the pitch obtained using the SBG angle measuring system. Thus, the rotation angles of the left and right cameras in two directions can be calculated linearly, and then, the extrinsic parameters of the binocular stereo vision after rotation can be obtained. The proposed calibration algorithm is non-iterative and can quickly complete the extrinsic parameter calibration of the rotating cameras, rendering the possibility of estimating the 3D coordinates in real time for the dynamic stereo vision, which can be used for fast positioning of dynamic targets.

The remainder of this paper is organized as follows: Section 2 primarily describes the mathematical model of the BSV with rotating and non-zooming cameras and introduces the extrinsic parameter calibration algorithm by using a single feature point. Section 3 discusses the feasibility of the calibration method. Section 4 explains the virtual simulation and physical experiments to verify the performance of the proposed method. Finally, Section 5 elaborates the conclusions.

2. Principles and Methods

2.1. Camera model of the Binocular Stereo Vision

The binocular stereo vision (BSV) system with rotating and non-zooming cameras is established, as shown in Figure 1. We describe the intrinsic parameter matrices K_1 and K_2 of the left and right camera as Equation (1), where f_{uL}, f_{vL} and f_{uR}, f_{vR} represent the focal lengths in the column and row directions for the left and right cameras, respectively, and s denotes the skewness of the two image axes, which remains zero in this study:

$$K_1 = \begin{bmatrix} f_{uL} & s & 0 \\ 0 & f_{vL} & 0 \\ 0 & 0 & 1 \end{bmatrix}, K_2 = \begin{bmatrix} f_{uR} & s & 0 \\ 0 & f_{vR} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$
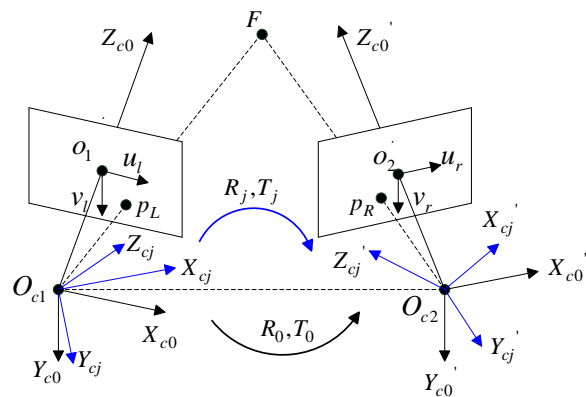
Figure 1. Mathematical model of binocular stereo vision.

We define the pixel coordinate of the principal point of the left and right cameras' image plane as (u_{0L}, v_{0L}) and (u_{0R}, v_{0R}), respectively. The distortion coefficient vector of a single camera is defined as $\boldsymbol{k_c} = [k_{c1}, k_{c2}, k_{c3}, k_{c4}, k_{c5}]$, where k_{c1}, k_{c2} and k_{c5} are respectively the two, four, and six order radial distortion coefficients, and k_{c3}, k_{c4} are tangential distortion coefficients. If we have the distortion coordinates (u_d, v_d) of one point on the imaging plane, the de-distortion coordinates (u, v) of this point can be obtained according to Equations (2) and (3):

$$u = f_u x_n + u_0, \; v = f_v y_n + v_0 \tag{2}$$

$$\begin{cases} x_d = \frac{u_d - u_0}{f_u} \\ y_d = \frac{v_d - v_0}{f_v} \\ x_d = x_n(1 + k_{c1}r^2 + k_{c2}r^4 + k_{c5}r^6) + 2k_{c3}x_n y_n + k_{c4}(r^2 + 2x_n{}^2) \\ y_d = y_n(1 + k_{c1}r^2 + k_{c2}r^4 + k_{c5}r^6) + k_{c3}(r^2 + 2y_n{}^2) + 2k_{c4}x_n y_n \\ r^2 = x_n{}^2 + y_n{}^2 \end{cases} \tag{3}$$

where (u_0, v_0) is the camera's principal point and f_u, f_v represent the focal lengths in the column and row directions, (x_n, y_n) and (x_d, y_d) are respectively the de-distortion coordinates and distortion coordinates on the imaging plane with normalized focal length.

The extrinsic parameters of BSV include rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{T}(\boldsymbol{T_x}, \boldsymbol{T_y}, \boldsymbol{T_z})$. The vector $\boldsymbol{om}$ is the Rodrigues representation of the rotation matrix, and the rotation matrix can be easily obtained using the Rodrigues transformation. In this study, we use three Euler angles rotating around the X-axis, Y-axis, and Z-axis (denoted by r_x, r_y, and r_z, respectively) to represent the rotation matrix $\boldsymbol{R}$ as shown in Equation (4). Subsequently, r_x, r_y, r_z can be converted by matrix $\boldsymbol{R}$ according to Equation (5), where $\boldsymbol{R_{ij}}$ ($i, j = 1, 2, 3$) represents the element of matrix $\boldsymbol{R}$ in the i-th row and j-th column:

$$R_{3\times3} = R_z.R_x.R_y \tag{4}$$

where $R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(r_x) & \sin(r_x) \\ 0 & -\sin(r_x) & \cos(r_x) \end{bmatrix}, R_y = \begin{bmatrix} \cos(r_y) & 0 & -\sin(r_y) \\ 0 & 1 & 0 \\ \sin(r_y) & 0 & \cos(r_y) \end{bmatrix}, R_z = \begin{bmatrix} \cos(r_z) & \sin(r_z) & 0 \\ -\sin(r_z) & \cos(r_z) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$

$$r_x = -\tan^{-1}\frac{R_{32}}{\sqrt{R_{31}{}^2 + R_{33}{}^2}}, \; r_y = \tan^{-1}\frac{R_{31}}{R_{33}}, \; r_z = \tan^{-1}\frac{R_{12}}{R_{22}} \tag{5}$$

The coordinate systems of the left and right cameras at the initial position are O_{c1}-$X_{c0}Y_{c0}Z_{c0}$ and O_{c2}-$X_{c0}'Y_{c0}'Z_{c0}'$, respectively. The cameras are stationary but can rotate in horizontal and vertical directions. The coordinate systems of the two cameras after the j-th rotation are denoted by O_{c1}-$X_{cj}Y_{cj}Z_{cj}$ and O_{c2}-$X_{cj}'Y_{cj}'Z_{cj}'$. We define the rotation matrix and the translation vector of the coordinate system O_{c1}-$X_{c0}Y_{c0}Z_{c0}$ relative to the coordinate system O_{c2}-$X_{c0}'Y_{c0}'Z_{c0}'$ as $\boldsymbol{R_0}$ and $\boldsymbol{T_0}$, as

shown in Equation (6). Similarly, the rotation matrix and translation vector after the j-th rotation is denoted by $\boldsymbol{R}_j$ and $\boldsymbol{T}_j$:

$$
\begin{bmatrix} X_{c0}' \\ Y_{c0}' \\ Z_{c0}' \end{bmatrix} = \boldsymbol{R}_0 \begin{bmatrix} X_{c0} \\ Y_{c0} \\ Z_{c0} \end{bmatrix} + \boldsymbol{T}_0 \tag{6}
$$

An arbitrary point in 3D is denoted by P, which has two projection points, namely, $p_L(u_L, v_L)$ and $p_R(u_R, v_R)$, on the left and right camera image planes after the j-th rotation, respectively. It should be emphasized that the pixel coordinates below are de-distortion coordinates. We define the left camera coordinate system as the world coordinate system, and the origin is the optical center of the left camera. According to the geometric properties of the optical imaging, we have equations Equations (7) and (8), where d_{xL}, d_{yL} and d_{xR}, d_{yR} represent the physical dimensions of the unit pixel of the left and right cameras in the column and row directions, respectively. Given that the two cameras of the same configuration are used in this study, we define $d_{xL} = d_{xR} = dx$, $d_{yL} = d_{yR} = dy$. According to Equations (7) and (8), the 3D coordinates of $P(X_{cj}, Y_{cj}, Z_{cj})$ in world coordinates can be solved by the least square method after cameras intrinsic and extrinsic parameters are obtained:

$$
Z_{cj} \begin{bmatrix} (u_L - u_{0L})d_{xL} \\ (v_L - v_{0L})d_{yL} \\ 1 \end{bmatrix} = \boldsymbol{K}_1 \begin{bmatrix} X_{cj} \\ Y_{cj} \\ Z_{cj} \end{bmatrix} \tag{7}
$$

$$
Z_{cj}' \begin{bmatrix} (u_R - u_{0R})d_{xR} \\ (v_R - v_{0R})d_{yR} \\ 1 \end{bmatrix} = \boldsymbol{K}_2 \boldsymbol{R}_j \begin{bmatrix} X_{cj} \\ Y_{cj} \\ Z_{cj} \end{bmatrix} + \boldsymbol{K}_2 \boldsymbol{T}_j \tag{8}
$$

2.2. Extrinsic Parameter Calibration Using a Single Feature Point

This study aims to calibrate the extrinsic parameters of the BSV system when the two cameras rotate. The intrinsic parameters of the left and right cameras and the translation vector of the two cameras at the initial position can be calibrated in advance offline. As shown in Figure 2a, the coordinate system of the left camera at the initial position is O_{c1}-$X_{c0}Y_{c0}Z_{c0}$, and point p in the image plane is the projection point of the feature point P in the 3D world coordinate system at initial time. Assuming the left camera rotates to the position of the blue dotted rectangle after the j-th rotation, the coordinate system of the camera at this moment is O_{c1}-$X_{cj}Y_{cj}Z_{cj}$, and point p' in the image plane is the projection point of the same feature point P at this time. The rotation angle of the camera in horizontal and vertical directions is denoted by yaw and pitch, respectively. The approximate values of these two attitude angles can be obtained in real time using the SBG angle measuring instrument, as shown in Figure 2b, in which the precision is $\pm 1°$. Let $P(X, Y, Z)$ represent the 3D coordinates of point P in O_{c1}-$X_{c0}Y_{c0}Z_{c0}$, and $p(u_0, v_0)$ and $p'(u_j, v_j)$ represent the image pixel coordinates in the image plane.

To simplify the notation, sin(yaw) and cos(yaw) are abbreviated as Sy and Cy, respectively. Meanwhile, sin(pitch) and cos(pitch) are abbreviated as Sp and Cp, respectively. According to the geometric imaging principle, Equations (9) and (10) are obtained for the left camera:

$$
Z_{c0} \begin{bmatrix} (u_0 - u_{0L})dx \\ (v_0 - v_{0L})dy \\ 1 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{9}
$$

$$
Z_{cj} \begin{bmatrix} (u_j - u_{0L})dx \\ (v_j - v_{0L})dy \\ 1 \end{bmatrix} = \boldsymbol{K}\boldsymbol{R}_j \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{10}
$$

where $K = \begin{bmatrix} \lambda f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $R_j = \begin{bmatrix} Cy & 0 & -Sy \\ SySp & Cp & CySp \\ SyCp & -Sp & CyCp \end{bmatrix}$, $f = f_{vL} \times dy$, $\lambda = \frac{f_{uL}}{f_{vL}}$.
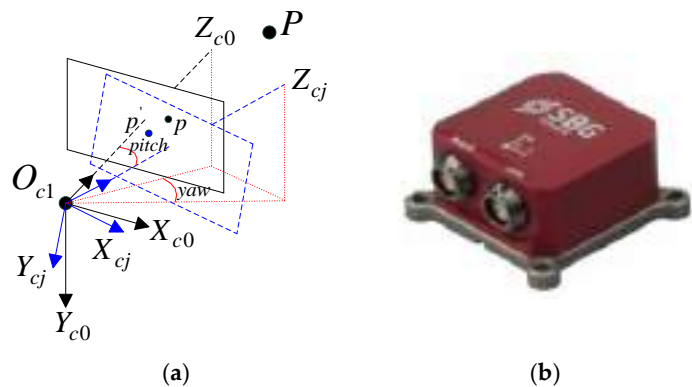

Figure 2. Dynamic mathematical model of the left camera and angle measuring instrument: (**a**) Rotation model of the left camera; (**b**) SBG angle measuring instrument.

Equations (9) and (10) are simply written as $Z_{c0}[U_0\ V_0\ 1]^T = K[X\ Y\ Z]^T$ and $Z_{cj}[U_j\ V_j\ 1]^T = KR_j[X\ Y\ Z]^T$. Thus, a mapping relationship exists between the projection points of the same single feature point in the image plane at the initial position and the image plane after the *j*-th rotation, expressed as Equation (11). This relationship is known as inter-image homography:

$$\mu \begin{bmatrix} U_j \\ V_j \\ 1 \end{bmatrix} = KR_jK^{-1} \begin{bmatrix} U_0 \\ V_0 \\ 1 \end{bmatrix} \tag{11}$$

where μ is the scale factor and $\mu = Z_{cj}/Z_{c0}$.

According to Equation (11), the following two linear equations for *Sy* and *Cy*, as shown in Equation (12), can be derived. Equation (12) can be simply represented by a matrix equation, that is, $A_{2\times2}[Sy\ Cy]^T = b_{2\times1}$. So we can convert the imaging model into a linear model with respect to sine and cosine of the yaw, where each element in the augmented coefficient matrix is a function of the single variable pitch. Subsequently, the determinant value of the matrix *A* in Equation (13) can be obtained:

$$\begin{cases} (U_jU_0Cp + \lambda^2 f^2)Sy + \lambda f(U_jCp - U_0)Cy = \lambda V_0U_jSp \\ (U_0V_jCp - fU_0Sp)Sy + \lambda f(V_jCp - fSp)Cy = \lambda V_jV_0Sp + \lambda fV_0Cp \end{cases} \tag{12}$$

$$det(A) = \begin{vmatrix} U_jU_0Cp + \lambda^2 f^2 & \lambda f(U_jCp - U_0) \\ (U_0V_jCp - fU_0Sp) & \lambda f(V_jCp - fSp) \end{vmatrix} = \lambda f(V_jCp - fSp)(\lambda^2 f^2 + U_0^2) \tag{13}$$

If $det(A) = 0$, then $V_jCp{-}fSp = 0$. Thus, $\tan(\text{pitch}) = V_j/f$. In the following, the tangent value of the angle pitch is simply denoted by *Tp*. If $det(A) \neq 0$, then the solution of *Sy* and *Cy*, expressed in Equations (14) and (15), can be obtained according to the linear equation principle:

$$Sy = \frac{\begin{vmatrix} \lambda V_0U_jSp & \lambda f(U_jCp - U_0) \\ \lambda V_jV_0Sp + \lambda fV_0Cp & \lambda f(V_jCp - fSp) \end{vmatrix}}{det(A)} = \lambda^2 fV_0 \frac{U_0(V_jSp + fCp) - fU_j}{det(A)} \tag{14}$$

$$Cy = \frac{\begin{vmatrix} U_1U_0Cp + \lambda^2 f^2 & \lambda V_0U_1Sp \\ U_0V_1Cp - fU_0Sp & \lambda V_1V_0Sp + \lambda fV_0Cp \end{vmatrix}}{det(A)} = \lambda V_0 f \frac{U_1U_0 + \lambda^2 f(V_1Sp + fCp)}{det(A)} \tag{15}$$

Given that $Sy^2 + Cy^2 = 1$, Equation (16) can be derived:

$$aSp^2 + bCp^2 + cSpCp + d = 0 \tag{16}$$

where $\begin{cases} a = \lambda^2 V_0^2 V_j^2 - f^2(\lambda^2 f^2 + U_0^2) \\ b = \lambda^2 V_0^2 f^2 - V_j^2(\lambda^2 f^2 + U_0^2) \\ c = 2V_j f(\lambda^2 V_0^2 + \lambda^2 f^2 + U_0^2) \\ d = V_0^2 U_j^2 \end{cases}$.

Subsequently, a quadratic equation of variable Tp, expressed as Equation (17), can be derived if Equation (16) is divided by the square of Cp. So the linear model of the pitch and yaw are transformed into a quadratic equation of the tangent value of the pitch:

$$mTp^2 + nTp + q = 0 \tag{17}$$

where $\begin{cases} m = V_0^2(\lambda^2 V_j^2 + U_j^2) - f^2(\lambda^2 f^2 + U_0^2) \\ n = 2V_j f[\lambda^2(V_0^2 + f^2) + U_0^2] \\ q = V_0^2(U_j^2 + \lambda^2 f^2) - V_j^2(\lambda^2 f^2 + U_0^2) \end{cases}$.

In the first case, if $m = 0$, then $Tp = -q/n$. In the second, if $m \neq 0$, then the solution of Tp can be obtained using Equation (18) because the quadratic equation must contain real roots:

$$Tp = \frac{-n \pm \sqrt{n^2 - 4mq}}{2m} \tag{18}$$

Notably, Equation (18) contains two solutions at the most. Hence, we utilize the SBG angular measurement instrument in this study to obtain the approximate value of the rotation angle pitch to eliminate the wrong solution. Therefore, in the two cases mentioned above, the correct solution of Tp can be uniquely determined. Given that the rotating angle pitch of the camera in the vertical direction satisfies $-90° <$ pitch $< 90°$, we have $Cp > 0$. Subsequently, the sine and cosine values of the angle pitch, namely, Sp and Cp, can be obtained using Equation (19) according to the trigonometric function theorem:

$$Cp = \sqrt{\frac{1}{1 + Tp^2}}, \ Sp = Tp \times Cp \tag{19}$$

Once the closed-form solutions of sine and cosine values of the angle pitch are obtained, the rotation angle yaw of the camera in the horizontal direction can be calculated linearly according to Equation (12), that is, $[Sy \ Cy]^T = A^{-1}b$. Thus, the rotation angles of the left camera in two directions after the rotation can be uniquely determined. Subsequently, the rotation matrix $\mathbf{R}_{lj}$ of the left camera relative to its initial position after the j-th rotation can be obtained. Similarly with the left camera, the closed-form solutions of the angle pitch and yaw of the right camera after the j-th rotation can also be calculated by using the same feature point with the aid of SBG system. Then the rotation matrix $\mathbf{R}_{rj}$ of the right camera relative to its initial position after the j-th rotation can be obtained. The calibration method is fast and linear, thereby avoiding the problem of the local optimal solution of iterative algorithms.

In this work, the pedestal of the left and right cameras are fixed, the coordinate systems O_{c1}-$X_{c0}Y_{c0}Z_{c0}$ and O_{c2}-$X_{c0}'Y_{c0}'Z_{c0}'$ of the left and right cameras at the initial position can be mapped to the coordinate systems O_{c1}-$X_{cj}Y_{cj}Z_{cj}$ and O_{c2}-$X_{cj}'Y_{cj}'Z_{cj}'$, respectively, after the j-th rotation according to the rotation theorem with a zero translation vector, expressed as Equation (20):

$$\begin{bmatrix} X_{cj} \\ Y_{cj} \\ Z_{cj} \end{bmatrix} = \mathbf{R}_{lj} \begin{bmatrix} X_{c0} \\ Y_{c0} \\ Z_{c0} \end{bmatrix}, \ \begin{bmatrix} X_{cj}' \\ Y_{cj}' \\ Z_{cj}' \end{bmatrix} = \mathbf{R}_{rj} \begin{bmatrix} X_{c0}' \\ Y_{c0}' \\ Z_{c0}' \end{bmatrix} \tag{20}$$

If the rotation matrix R_0 and translation vector T_0 of the BSV system at the initial position are known in advance, the mapping relationship between the coordinate system $O_{c1}\text{-}X_{cj}Y_{cj}Z_{cj}$ and $O_{c2}\text{-}X_{cj}'Y_{cj}'Z_{cj}'$, as shown in Equation (21), can be obtained by combining Equations (6) and (20). As can be seen from Equation (21), the rotating matrix R_j and translation vector $T_j(T_x, T_y, T_z)$ of the BSV system after the *j*-th rotation can be obtained as shown in Equation (22). The three Euler angles (i.e., r_x, r_y, and r_z) representing matrix R_j can be solved by Equation (5), and 3D coordinates (X, Y, Z) of the feature point can be estimated by Equations (7) and (8):

$$\begin{bmatrix} X_{cj}' \\ Y_{cj}' \\ Z_{cj}' \end{bmatrix} = R_{lj}R_0R_{rj}^{-1} \begin{bmatrix} X_{cj} \\ Y_{cj} \\ Z_{cj} \end{bmatrix} + R_{rj}T_0 \tag{21}$$

$$R_j = R_{lj}R_0R_{rj}^{-1}, \; T_j = \begin{bmatrix} T_x & T_y & T_z \end{bmatrix}^T = R_{rj}T_0 \tag{22}$$

3. Feasibility Analysis

In this work, the left and right cameras can rotate from $-45°$ to $45°$ in the horizontal direction and from $-10°$ to $10°$ in the vertical direction, respectively. To investigate the performance of the proposed method with respect to the posture of one arbitrary camera, we assume that the focal length of this camera is 16 mm and the image size is 1360 pixels $\times$ 600 pixels. We vary the angle pitch from $-10°$ to $10°$, and the angle yaw from $-45°$ to $45°$ with an interval of $1°$. We simulate 256 pairs of matched feature points between the camera's initial position and the position after the rotation for each posture. Hence, 256 repetitive trials are implemented and the Gaussian noise with zero mean and a standard deviation of 0.5 pixel are added to the feature points. Given that the key of the algorithm mentioned in Section 2 lies in the calculation of the angle pitch, we measure the root mean square (RMS) errors between the true values of the pitch and its estimated values to evaluate the calculation accuracy. As shown in Figure 3, the RMS increases with the increasing rotation angle but does not exceed $0.007°$, implying that the proposed algorithm is suitable for all poses of the left and right cameras.
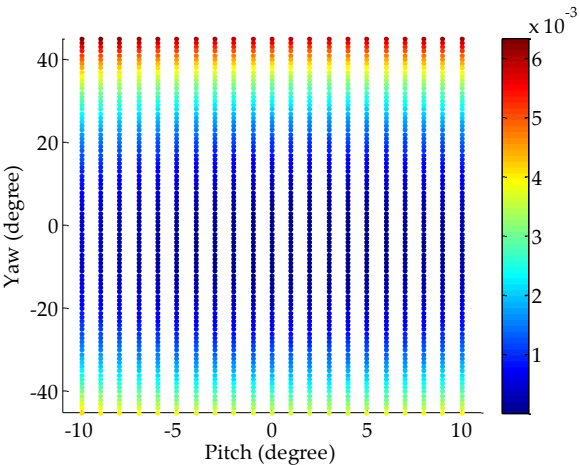

Figure 3. RMS versus the posture of the camera.

To investigate the performance with respect to the location of the pixel coordinates, we simulate a total of 8160 pixel points in the image at the initial position of the camera and simple at 10-pixel intervals. The corresponding matched pixel points after the rotation can also be simulated according to these pixel points. For each pixel coordinate location, 256 repetitive trials are implemented, and the Gaussian noise with zero mean and a standard deviation of 0.5 pixel are added to the feature points. The RMS mentioned above with respect to the pixel coordinates is shown in Figure 4. As presented in

the figure, the RMS value is less than $0.007°$, implying that the proposed method is feasible regardless of the pixel coordinates of the feature point.
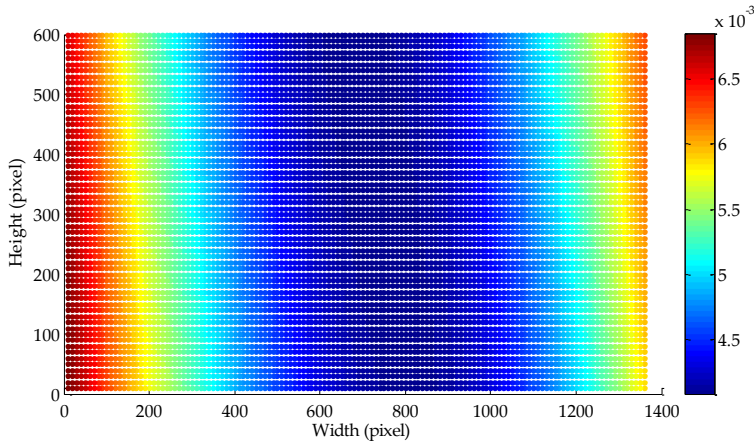

Figure 4. RMS versus the pixel coordinates of the feature point.

4. Experiments

4.1. Computer Simulation

The algorithm proposed in Section 2 is a self-calibration method by using a single feature point. The extraction of the pixel coordinates of the feature points significantly affects the calibration accuracy. To evaluate the performance of the proposed method, we simulate 216 pairs of matched feature points between the initial position of binocular stereo vision and the position after the rotation. The simulation assumptions are as follows. The image size of the left and right virtual cameras are 1360 pixels × 600 pixels, and the focal length is 16 mm and 18 mm, respectively. The Rodrigues vector *om* and translation vector T_0 at the initial position is *om* = [−0.00281, 0.05055, 0.03414] and T_0 = [−304.5563, −3.8191, 27.5258]^T. The Rodrigues vector *om′* and *T′* after the rotation is *om′* = [−0.02979, 0.13893, 0.04427] and *T′* = [−296.8178, −0.8696, −5.7945]^T. The units of the translation vector are millimeters. In each simulation experiment, a pair of feature points is randomly selected for the extrinsic parameters estimation, and the experiment is repeated 216 times in each noise level. The RMS values of the three Euler angles (i.e., r_x, r_y, and r_z), the translation vector $T = [T_x, T_y, T_z]^T$, and 3D reconstruction coordinates (*X*, *Y*, *Z*) are used to evaluate the calibration accuracy, as shown in Figure 5. The Gaussian noise with zero mean and standard deviation of 0.1–1 pixel with an interval of 0.1 pixel are added into the feature points. As shown in Figure 5, the RMS increases linearly with the increasing noise level, but the calibration error of the extrinsic parameters and the 3D reconstruction error are low when the noise level reaches 1 pixel.

4.2. Physical Experiment

To verify the proposed self-calibration method, the system of binocular stereo vision with rotating cameras is constructed, as shown in Figure 6. Two gigabit network cameras with the same configuration are installed on the horizontal platform. The two cameras can only rotate in the horizontal and vertical directions but is sufficient to adjust the field of view. The captured images are transmitted into the computer via a network cable. The image size of the left and right cameras is 1360 pixels × 600 pixels, and the physical size of the unit pixel in the column and row directions is 6.45 μm. Given that the left and right cameras are non-zooming, the intrinsic and extrinsic parameters of the cameras at the initial position can be calibrated in advance. Currently, Zhang's checkerboard calibration method is extensively used in binocular stereo vision because of its convenience, low cost, and high precision. The intrinsic parameters, including the focal length, principal point, and distortion coefficients, of

the left and right cameras obtained by using Zhang's method are shown in Table 1, as well as the Rodrigues rotation om and translation vector T_0 at the initial position.
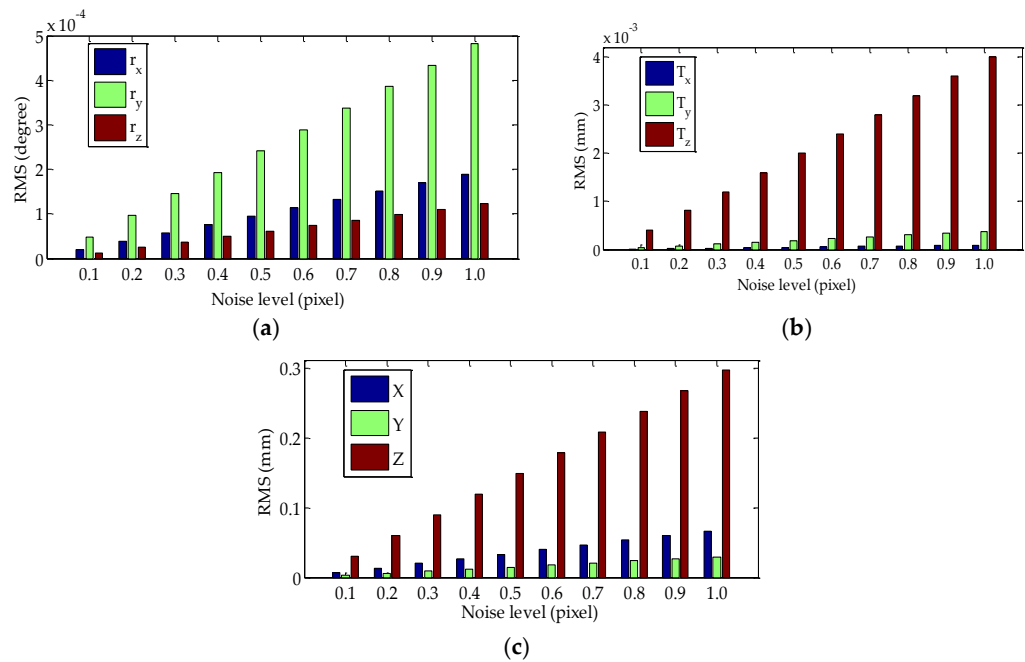

Figure 5. Performance of the proposed calibration method with respect to noise: (**a**) RMS of r_x, r_y, and r_z at various noise levels; (**b**) RMS of T_x, T_y, and T_z at various noise levels; (**c**) RMS of X, Y, and Z at various noise levels.
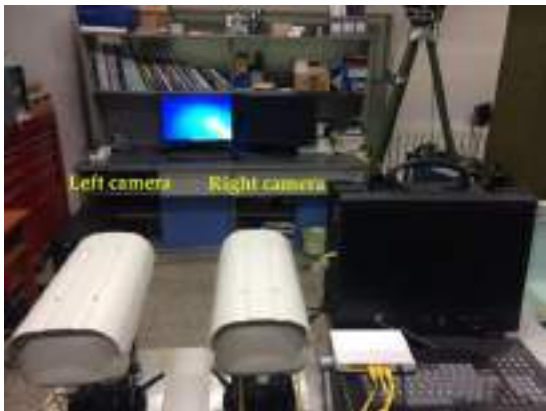

Figure 6. Experimental equipment of the binocular stereo vision.

Table 1. Cameras parameters.

Camera	f_u/pixels	f_v/pixels	u_0/pixels	v_0/pixels	k_c
Left	2493.09	2493.92	725.77	393.03	[−0.17, 0.18, 0.003, 0.0004, 0.00]
Right	2811.56	2811.54	692.04	371.96	[−0.13, 0.16, 0.001, 0.0003, 0.00]
om			[0.01085, 0.05695, 0.03387]		
T_0/mm			[−306.9049, −4.3956, 39.6172]		

In this experiment, the top-left corner point *A* of the display screen in Figure 7a is selected as the single feature point. The remaining three feature points are used for accuracy validation. The edge of the display screen is extracted using the Canny detection algorithm, as shown in Figure 7b. The four straight lines of the display screen's contour can be identified through the Hough line detection, as shown in Figure 7c. The intersections of these lines are the four corners of the display screen, and

the pixel coordinates of the corners can be extracted to the subpixel level. The feature point matching before and after the rotation is presented in Figure 7d.
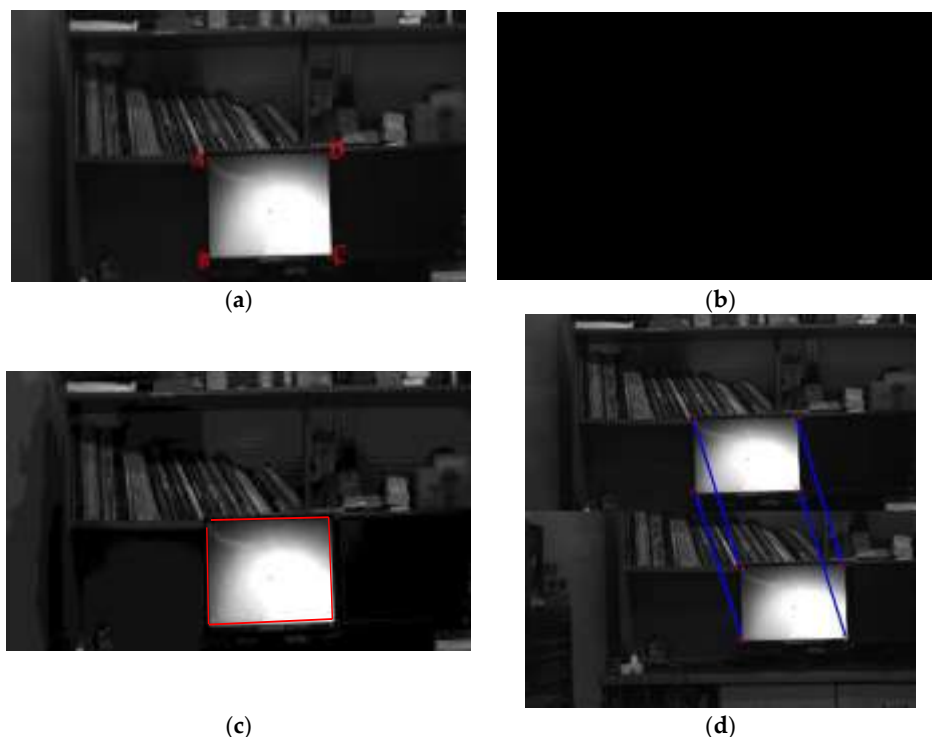

Figure 7. Image processing: (**a**) Source image; (**b**) Canny edge detection; (**c**) Hough line detection; (**d**) Feature point matching.

After the initial positions of the two cameras are determined, the first images of the display screen captured by the left and right cameras, are used as the reference frame. In each trial, the captured image is used as the current processing frame for the left and right camera after the cameras rotated to a certain position, and the output values of the SBG system are regarded as the approximate values of the rotation angle pitch and yaw. After the matching between the feature point in the current and reference images for the left and right cameras, respectively, is accomplished, the rotation angles of each camera in the two directions can be calculated using the proposed method in Section 2. Hence, the rotation matrix and the translation vector of the binocular stereo vision after the rotation can be obtained. In this work, Zhang's method is also used to implement the calibration after each rotation, and the calibration results are considered to be reference values and are compared with our data.

This experiment is repeated 12 times. Figure 8a presents the rotation angles of the left camera in the two directions obtained by using proposed method and the corresponding approximate values. Figure 8b presents that of the right camera. The angles in Figure 8 based on our calculations are close to the approximate values, thereby verifying the validity of the algorithm.

Figure 9 shows the three Euler angles representing the rotation matrix, namely, r_x, r_y, and r_z, obtained by using the proposed and Zhang's methods in each experimental trial. As shown in Figure 9, the trend of the three angles in proposed method is the same as that of Zhang's method. The absolute errors of the three Euler angles relative to the reference values are presented in Figure 10. The three averages of the absolute errors are less than $0.21°$, and the average error of r_x is the largest.
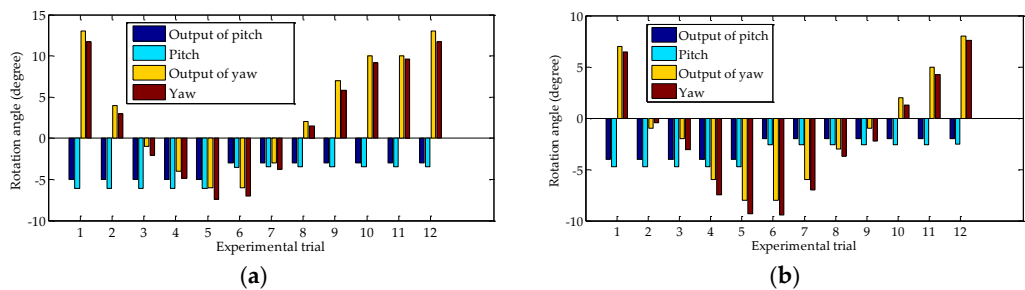
Figure 8. Our calculated rotation angles and the corresponding rough values: (**a**) Rotation angles of the left camera; (**b**) Rotation angles of the right camera.
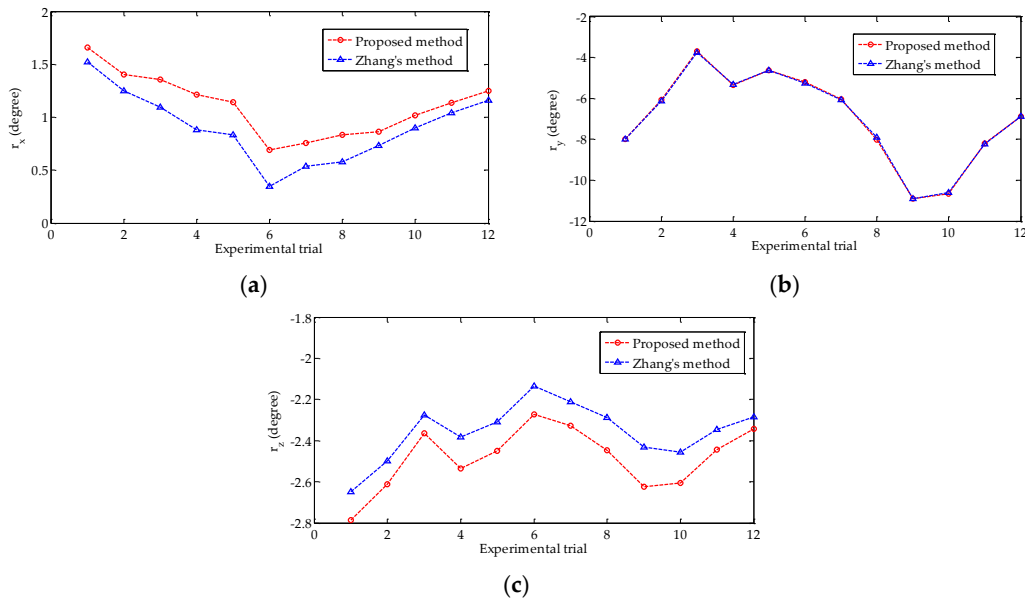

Figure 9. Euler angles estimated by using the proposed and Zhang's methods: (**a**) r_x of the two methods; (**b**) r_y of the two methods; (**c**) r_z of the two methods.
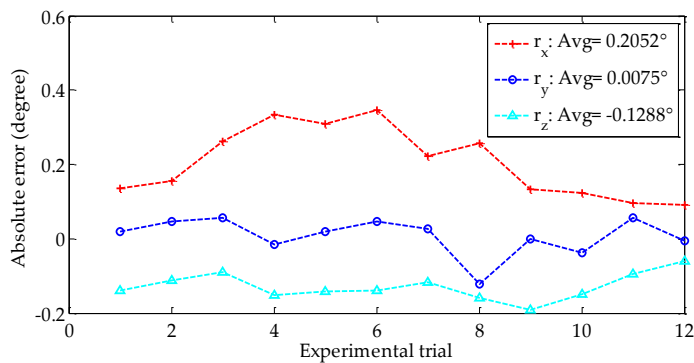

Figure 10. Absolute error of the three Euler angles.

Figure 11 shows the translation vector, namely, T_x, T_y, and T_z, obtained by using the proposed method and Zhang's methods in each experimental trial. Undoubtedly, the calibration accuracy of Zhang's method must be much higher than that of our self-calibration method by using a single feature point. However, the calibration results estimated by using the proposed method are close to those of Zhang's method, which proves the reliability of the proposed method. The absolute errors of the translation vector relative to the reference values are presented in Figure 12. The three averages of the absolute errors are less than 6.6 mm, and the average error of T_x is the largest.
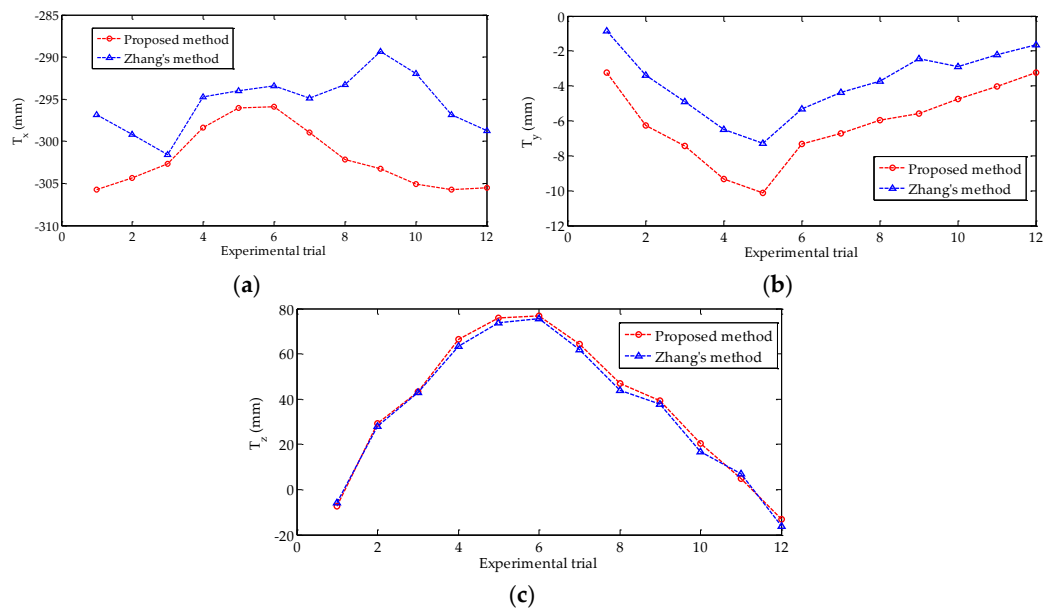
(a)

(b)

(c)

Figure 11. Translation vector estimated by using the proposed and Zhang's methods: (**a**) T_x of the two methods; (**b**) T_y of the two methods; (**c**) T_z of the two methods.
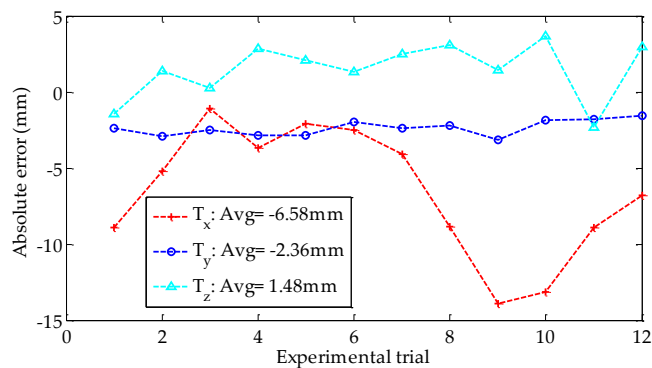

Figure 12. Absolute error of the translation vector.

In each trial, we calculate the 3D coordinates of the feature points B, C, and D on the display screen after the rotation, as shown in Figure 7a. The averages of the 3D coordinates are used for the evaluation of calibration accuracy. As shown in Figure 13, we compare the data obtained by using the proposed method with that calculated by using Zhang's method. The figure shows that the 3D reconstruction coordinates of the proposed method is close to that of Zhang's method, which proves the validity of the proposed method. The relative errors of the 3D coordinates with respect to the reference values are presented in Figure 14. The averages of the relative errors are less than 4.2%, and the relative error on Y-axis is the largest. The proposed method is suitable for the occasions where high accuracy of 3D reconstruction is not required.
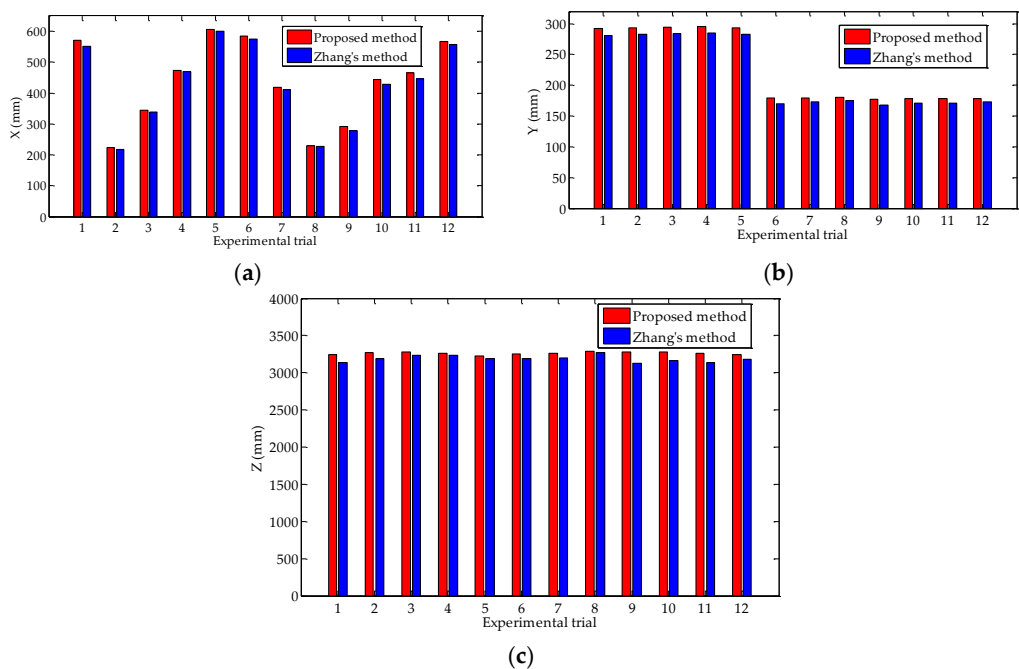
(**a**)

(**b**)

(**c**)

Figure 13. The 3D coordinates estimated by using the proposed method and Zhang's method: (**a**) X of the two methods; (**b**) Y of the two methods; (**c**) Z of the two methods.
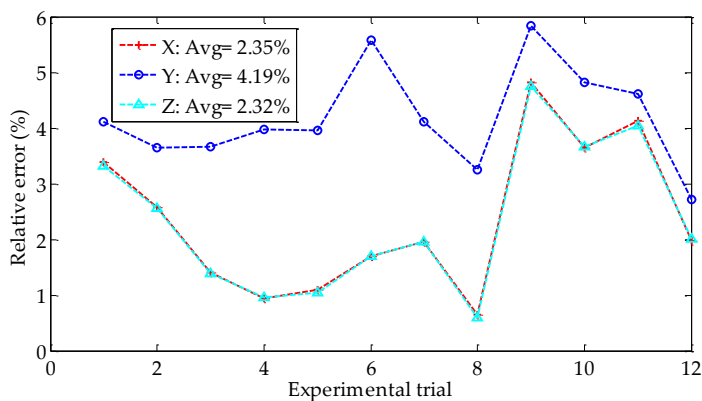

Figure 14. The relative error of the 3D coordinates.

5. Conclusions

In this study, we present a novel self-calibration method for extrinsic parameter estimation of a rotating binocular stereo vision by using a single feature point. This is achieved by assuming the intrinsic parameters of the left and right cameras are known in advance, as well as the rotation matrix and the translation vector at the initial position. Only the rotation angles of the left and right cameras in the vertical and horizontal directions, that is, pitch and yaw, must be calculated after the rotation. We transformed the geometric imaging model of the pitch and yaw into a quadratic equation of the tangent value of the pitch. The closed-form solutions of the pitch and yaw are obtained with the aid of SBG equipment. Once the pitch and yaw are uniquely determined, the rotation matrix of the BSV system and the three Euler angles representing rotation matrix can be calculated according to rotation theorem. The translation vector are estimated by the rotation matrix of the right camera and the initial translation vector.

The proposed method is non-iterative, thus addressing the problem of not obtaining a global optimal solution in iterative algorithms. Under extreme conditions, we limit the number of feature points for calibration to a minimum, remarkably shortening the consumption time of the feature point matching and allowing the possibility to calibrate the extrinsic parameters of the rotating binocular

stereo vision in real time. Computer simulations prove the feasibility of the proposed method, and the error of the extrinsic parameters and 3D coordinate reconstruction are minimal when the noise level was high. To validate the feasibility of the proposed method, we compare it with Zhang's method. Although Zhang's method exhibits a much higher precision, our calibration results from repeated experiments are close to the reference values estimated by Zhang's method. The primary contribution of this study is that the proposed method can be used for real time 3D coordinate estimation of dynamic binocular stereo vision in when an extremely high calibration accuracy is not required. In our future work, a real-time and highly accurate method will be simultaneously investigated.

Author Contributions: Y.W., X.W. conceived the article, conducted the experiments, and wrote the paper. Z.W. and J.Z. helped establish mathematical model.

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

References

1. Tsai, R.; Lenz, R.K. A technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans. Robot. Autom.* **1989**, *5*, 345–358. [CrossRef]
2. Tsai, R.Y. An efficient and accurate camera calibration technique for 3D machine vision. In Proceedings of the CVPR'86: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, USA, 22–26 June 1986; pp. 364–374.
3. Tian, S.-X.; Lu, S.; Liu, Z.-M. Levenberg-Marquardt algorithm based nonlinear optimization of camera calibration for relative measurement. In Proceedings of the 34th Chinese Control Conference, Hangzhou, China, 28–30 July 2015; pp. 4868–4872.
4. Habed, A.; Boufama, B. Camera self-calibration from bivariate polynomials derived from Kruppa's equations. *Pattern Recognit.* **2008**, *41*, 2484–2492. [CrossRef]
5. Wang, L.; Kang, S.-B.; Shum, H.-Y.; Xu, G.-Y. Error analysis of pure rotation based self-calibration. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; pp. 464–471.
6. Hemayed, E.E. A survey of camera calibration. In Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Miami, FL, USA, 21–22 July 2003; pp. 351–357.
7. Abdel-Aziz, Y.I.; Karara, H.M. Direct linear transformation into object space coordinates in close-range photogrammetry. *Photogramm. Eng. Remote Sens.* **2015**, *81*, 103–107. [CrossRef]
8. Zhang, L.; Wang, D. Automatic calibration of computer vision based on RAC calibration algorithm. *Metall. Min. Ind.* **2015**, *7*, 308–312.
9. Zhang, Z. Flexible camera calibration by viewing a plane from unknown orientations. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 666–673.
10. Martins, H.A.; Birk, J.R.; Kelley, R.B. Camera models based on data from two calibration planes. *Comput. Graph. Image Process.* **1981**, *17*, 173–180. [CrossRef]
11. Faugeras, O.D.; Luong, Q.-T.; Maybank, S.J. Camera self-calibration: Theory and experiments. In Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy, 19–22 May 1992; pp. 321–334.
12. Li, X.; Zheng, N.; Cheng, H. Camera linear self-calibration method based on the Kruppa equation. *J. Xi'an Jiaotong Univ.* **2003**, *37*, 820–823.
13. Hartley, R. Euclidean reconstruction and invariants from multiple images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 1036–1041. [CrossRef]
14. Triggs, B. Auto-calibration and the absolute quadric. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 609–614.
15. De Ma, S. A self-calibration technique for active vision system. *IEEE Trans. Robot. Autom.* **1996**, *12*, 114–120. [CrossRef]
16. Zhang, Z.; Tang, Q. Camera self-calibration based on multiple view images. In Proceedings of the Nicograph International (NicoInt), Hanzhou, China, 6–8 July 2016; pp. 88–91.

17. Wu, Z.; Radke, R.J. Keeping a pan-tilt-zoom camera calibrated. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1994–2007. [CrossRef] [PubMed]

18. Junejo, I.N.; Foroosh, H. Optimizing PTZ camera calibration from two images. *Mach. Vis. Appl.* **2012**, *23*, 375–389. [CrossRef]

19. Echigo, T. A camera calibration technique using three sets of parallel lines. *Mach. Vis. Appl.* **1990**, *3*, 159–167. [CrossRef]

20. Song, K.-T.; Tai, J.-C. Dynamic calibration of Pan-Tilt-Zoom cameras for traffic monitoring. *IEEE Trans. Syst. Man Cybern. Part B* **2006**, *36*, 1091–1103. [CrossRef]

21. Schoepflin, T.N.; Dailey, D.J. Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. *IEEE Trans. Intell. Transp. Syst.* **2003**, *4*, 90–98. [CrossRef]

22. Kim, H.; Hong, K.S. A practical self-calibration method of rotating and zooming cameras. In Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, 3–7 September 2000; pp. 354–357.

23. Rodríguez, J.A.M. Online self-camera orientation based on laser metrology and computer algorithms. *Opt. Commun.* **2011**, *284*, 5601–5612. [CrossRef]

24. Rodríguez, J.A.M.; Alanís, F.C.M. Binocular self-calibration performed via adaptive genetic algorithm based on laser line imaging. *J. Mod. Opt.* **2016**, *63*, 1219–1232. [CrossRef]

25. Ji, Q.; Dai, S. Self-calibration of a rotating camera with a translational offset. *IEEE Trans. Robot. Autom.* **2004**, *20*, 1–14. [CrossRef]

26. Cai, H.; Zhu, W.; Li, K.; Liu, M. A linear camera self-calibration approach from four points. In Proceedings of the 4th International Symposium on Computational Intelligence and Design, Hangzhou, China, 28–30 October 2011; pp. 202–205.

27. Tang, A.W.K.; Hung, Y.S. A Self-calibration algorithm based on a unified framework for constraints on multiple views. *J. Math. Imaging Vis.* **2012**, *44*, 432–448. [CrossRef]

28. Yu, H.; Wang, Y. An improved self-calibration method for active stereo camera. In Proceedings of the Sixth World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006; pp. 5186–5190.

29. De Agapito, L.; Hartley, R.I.; Hayman, E. Linear self-calibration of a rotating and zooming camera. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Fort Collins, CO, USA, 23–25 June 1999; pp. 15–21.

30. Sun, J.; Wang, P.; Qin, Z.; Qiao, H. Effective self-calibration for camera parameters and hand-eye geometry based on two feature points motions. *IEEE/CAA J. Autom. Sin.* **2017**, *4*, 370–380. [CrossRef]

31. Chen, J.; Zhu, F.; Little, J.J. A two-point method for PTZ camera calibration in sports. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 287–295.