



Article

A New Gaze Estimation Method Based on Homography Transformation Derived from Geometric Relationship

Kaiqing Luo ^{1,2,3,4}, Xuan Jia ^{1,2,3,4}, Hua Xiao ¹, Dongmei Liu ^{1,2,3,4}, Li Peng ^{1,2,3,4}, Jian Qiu ^{1,2,3,4} and Peng Han ^{1,2,3,4},*

- Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, China; kqluo@scnu.edu.cn (K.L.); jiax@m.scnu.edu.cn (X.J.); xiaoh@scnu.edu.cn (H.X.); dmliu@scnu.edu.cn (D.L.); pengli@m.scnu.edu.cn (L.P.); qiuj@scnu.edu.cn (J.Q.)
- Guangdong-Hong Kong Joint Laboratory of Quantum Matter, South China Normal University, Guangzhou 510006, China
- Guangdong Provincial Engineering Research Center for Optoelectronic Instrument, South China Normal University, Guangzhou 510006, China
- SCNU Qingyuan Institute of Science and Technology Innovation Co. Ltd., Qingyuan 511517, China
- * Correspondence: hanp@scnu.edu.cn

Received: 25 October 2020; Accepted: 15 December 2020; Published: 18 December 2020



Abstract: In recent years, the gaze estimation system, as a new type of human-computer interaction technology, has received extensive attention. The gaze estimation model is one of the main research contents of the system. The quality of the model will directly affect the accuracy of the entire gaze estimation system. To achieve higher accuracy even with simple devices, this paper proposes an improved mapping equation model based on homography transformation. In the process of experiment, the model mainly uses the "Zhang Zhengyou calibration method" to obtain the internal and external parameters of the camera to correct the distortion of the camera, and uses the LM(Levenberg-Marquardt) algorithm to solve the unknown parameters contained in the mapping equation. After all the parameters of the equation are determined, the gaze point is calculated. Different comparative experiments are designed to verify the experimental accuracy and fitting effect of this mapping equation. The results show that the method can achieve high experimental accuracy, and the basic accuracy is kept within 0.6°. The overall trend shows that the mapping method based on homography transformation has higher experimental accuracy, better fitting effect and stronger stability.

Keywords: gaze estimation; mapping equation; homography transformation

1. Introduction

With the development of computer technology and semiconductor technology, the gaze estimation system has attracted more and more attention. Eyes are the main way for people to obtain information from the outside world, and the direction of human vision often represents the region of interest. Therefore, eye tracking technology is widely used in all aspects of life and scientific research. Such as page analysis, human-computer interaction, intelligent instruments and military and other fields.

According to the current research, the gaze estimation system mainly includes three core modules: eye feature extraction module, calibration module and compensation module. The eye feature module is mainly used to obtain eye feature parameters. The calibration module is mainly divided into adjustment, calibration (test) and recording eye movement data, so as to obtain the eye feature

Appl. Sci. 2020, 10, 9079 2 of 18

parameters, and then substituted into the mapping equation obtained in the adjustmeng process to estimate the user's sight direction. The function of the compensation module is that the interference of influence factors (such as head movement, camera distortion, etc.) will occur during the calibration process, and the compensation algorithm will be added to correct the head posture and other problems.

At present, the methods of mapping equations are mainly divided into two categories: appearancebased methods [1–6] and feature-based gaze estimation methods [7–25]. The interpolation-based mapping method does not need to consider the geometric relationship of scene plane and camera calibration [1]. Neural network (NN) [2–4] and Gaussian process (GP) [5,6] interpolation are two of the most commonly used mapping algorithms. However, the main disadvantage of this method is that when it comes to head movement, it cannot combine head posture information with appearance in a robust way. The mapping method based on feature vector extraction is the most widely used method to extract the local features of the face, such as eye contour, eye corner or reflection area of eye diagram. It is mainly divided into two categories: 2D and 3D. The basic principle of the 2D model is to extract the eye's 2D eye feature parameters from the image as the eye feature quantity, and determine the required mapping function from the gaze parameters obtained in the calibration process, finally, the eye feature quantity is substituted into the mapping function to obtain the gaze point. The 3D model is based on the geometric model and spatial geometric model of the human eye, and calculates the 3D gaze direction from the eye feature parameters of the image and the location coordinates of the light source. The 2D eye tracking methods mainly include Pupil-canthus, Pupil-Cornea reflection, Cross-ratio, and homography normalization. The Pupil-canthus [26–28] is usually a single camera without light source system. Zhu et al. [26] established a 2D linear mapping model between the iris center-eye cornea vector and the gaze angle, this model is simple and can achieve higher accuracy, they obtained average accuracy of 1.4° . Xia et al. [27] simplified the mapping model into a linear mapping model between the coordinates of the pupil center and the gaze point on the screen, the robustness of the detection results is enhanced, the accuracy is 0.84° in horizontal direction and 0.56° in vertical direction. Both methods limit the free movement of the head. Shao et al. [28] proposed a method to estimate the gaze point based on the 2D of eye and screen, but more eye features and calibration points are needed than traditional second-order mapping models. Pupil-Cornea reflection [29–35] compared with Pupil-canthus, add a light source in the system. Blignaut et al. [29] proposed the best 2D mapping model when the number of calibration points is greater than or equal to 14, after comparing and analyzing the existing 2D mapping models, accuracy can be achieved within 0.5° and takes about 20 s. George et al. [30] proposed a more accurate iris center detection method to obtain the iris center parameters, which are brought into the polynomial to obtain more accurate gaze point coordinates. Tong et al. [31] conducted an in-depth research on the mapping equation and found a more appropriate high order odd polynomial fitting mapping equation, with an experimental accuracy of 0.87°. Zhang et al. [33] proposed a compensation model for eye feature parameters in a single camera dual ring light source system to derive the head movements, thus correcting the corresponding parameters to obtain accurate information about the gaze point. Morimoto et al. [34] verified that the quadratic polynomial is more suitable for solving the 2D sight direction through a variety of polynomial experiments. The Cross-ratio Invariance in projective geometry is applied in Cross-ratio [36,37]. Cheng et al. [36] proposed the idea of using dynamic virtual tangent plane, describe the relationship between reflection point and virtual point of light source with dynamic matrix, the sight accuracy can reach 0.7°. Dong et al. [37] proposed a gaze estimation method based on Cross-ratio under two cameras and five LED infrared light sources, which can obtain clearer eye images and extract more accurate pupil features. Homography normalization [38,39] is similar to the Cross-ratio method. Choi et al. [38] proposed that in the four infrared light source system, no additional light source. Due to the characteristics of the normalized space, it is not necessary to know the screen rectangle size formed, which is convenient for system implementation, and has stronger robustness, they obtained average accuracy of 1.03° – 1.17° .

Appl. Sci. 2020, 10, 9079 3 of 18

Based on the above literature, it can be found that when the hardware system is simple, the accuracy is usually low, when the accuracy is high, the hardware system is usually more complex, there are limitations and inaccuracies in the actual experiment process. Therefore, in order to reduce the complexity of the algorithm, save the hardware cost and improve the stability and accuracy of the system, this paper proposes a mapping equation based on homography transformation, and calculates the gaze point through adjustment, calibration and experimental verification, this method can achieve higher experimental accuracy.

This article will introduce the following aspects. Section 2 mainly compares our method with other methods. The derivation of the mapping equation will be introduced in the method section. Section 4 (Experiments) mainly introduces the experimental design and result analysis of the mapping equation. Section 5 (Discussion) is the discussion part, the practical significance of this equation is obtained by further polynomial surface fitting and analysis of characteristic parameters, and it is expected that it can be put into engineering applications. Finally, it introduces the conclusion of this paper and the significance of its existence.

2. Comparison Methods

2.1. Polynomial Fitting Method

Right now, the feature-based methods (also known as polynomial fitting mapping methods) [40] use specific parameters to map the image feature points to screen gaze point coordinates (gaze point coordinates can be 2D or 3D coordinates). It avoids complete hardware calibration steps, and has better simplification and operation. Therefore, it is a commonly used method to solve the mapping relationship between data. Depending on the complexity of the model, there will be existed different forms of polynomials, such as polynomial of first degree, quadratic polynomial, and even other high-order polynomials. Here, it is assumed that the coordinates of the pupil center position in the eyeball coordinate system is (x, y), and the gaze point coordinates in the scene plane coordinate system is (u, v). This section mainly uses the classical quadratic polynomial mapping model as an example to explain its principle.

In 1974, with continuous research Merchant et al. [41] finally proposed a video-based real-time gaze estimation method. The implementation process of this method mainly uses a pupil center and an optical lens combined with a photoelectric galvanometer to detect the gaze of human eye. They have proved the conclusion that the spatial distribution of pupil and gaze point is nonlinear through experimental research. Based on the above research, Morimoto et al. [42] finally verified through a variety of polynomial experiments that the classic quadratic polynomial fitting method is more suitable for solving viewing direction of 2d. The expression is as follows:

$$u = a_0 x^2 + a_1 xy + a_2 y^2 + a_3 x + a_4 y + a_5$$

$$v = b_0 x^2 + b_1 xy + b_2 y^2 + b_3 x + b_4 y + b_5$$
(1)

In the above formula, the eye feature vector is $F_e = [x, y]^T \epsilon R^{2 \times 1}$, Position vector of gaze point $G_g = [u, v]^T \epsilon R^{2 \times 1}$.

Morimoto uses one of the pupil corneal reflex methods, which requires a light source that forms a Purkinje image [43] through the corneal reflex, and the position of the Purkinje image on the cornea can be assumed to be stationary when head movements remain constant. With the Purkinje image as the datum, the drop of the sight on the screen can be obtained by bringing it into the mapping model. The pupil-corneal reflex method uses an infrared light source, and the phenomenon of "bright pupil" generated by it will make the video image easier to process. The production principle of bright pupil and dark pupil is shown in Figure 1.

Appl. Sci. 2020, 10, 9079 4 of 18

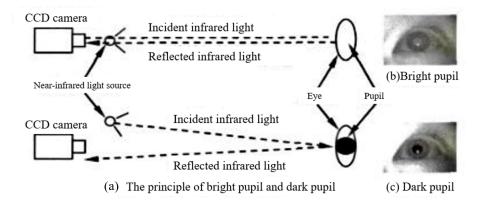


Figure 1. Pupil formation principle of light pupil and dark pupil [44].

Theoretically, the higher the function order, the higher the fitting accuracy. Therefore, our previous work [31] is to study the polynomial fitting equation in depth, in order to make the data fitting uncomplicated, the higher order power is discarded and the fifth order term is retained, and propose a higher-order odd polynomial fitting equation, the expression is shown in (2). However, the actual accuracy is not very ideal. In this paper, we continue to study and a mapping equation based on homography is proposed.

$$u = a_0 x + a_1 x y^2 + a_2 x^3 + a_3 x y^4 + a_4 x^3 y^2 + a_5 x^5$$

$$v = b_0 y + b_1 x^2 y + b_2 y^3 + b_3 x^4 y + b_4 x^2 y^3 + b_5 y^5$$
(2)

After the polynomial fitting method determining the mapping function equation, it needs to calibrate to solve the unknown parameters, a_i and b_i to obtain the appropriate gaze estimation regression model. The eye feature vector F_e is obtained from the pupil center coordinates and Purkinje image coordinates by the tester look directly at the screen and look at the calibration points at different positions in turn. The commonly used calibration points are 4, 5, 9 and 14. Blignaut et al. [29] found that when the number of calibration points is 14, the calibration accuracy is the highest, followed by 9 points. However, the more sample points there are, the greater the calculation amount will be, and the tester is prone to fatigue, which will lead to sight drift to interfere with the test. Therefore, nine calibration points are used for fixation calibration. The state information and data variables of the eye features were recorded by experiments when the subjects successively gazed at the calibration point. The obtained information is the data set required for calibration. Finally, regression analysis of the above formula by the least square method, i.e., to optimize the mean square deviation of the estimation error.

2.2. Homography Normalization

The hardware system based on the method of homography normalization [39] is configured as a camera and four light sources placed at the four corners of the screen. In this method, gaze point is estimated by the transformation relationship between two projections. The basic principle is to calculate the corresponding two homography matrices according to the projection mapping relationship between the imaging plane and the corneal reflection plane, and between the corneal reflection plane and the screen plane. When estimating the sight drop point, according to the position of the pupil center on the imaging plane, the homography matrix H_1^N formed by the imaging plane and the corneal reflection plane is used to convert the point P on the corneal reflection plane. However, since the system is not punctuated, the 3D coordinates of the corneal reflex plane are unknown. Therefore, it is assumed that the corneal reflex plane is a normalized plane, i.e., a unit square with

Appl. Sci. 2020, 10, 9079 5 of 18

predefined coordinates. Then, the homography matrix H_S^N formed by the corneal reflection plane and the screen plane is converted into the point S on the screen plane, which is the fixation point. H_1^N can be calculated from four CRs points generated by corneal reflex and the four corner points of predefined normalized plane. The projection matrix H_S^N composed of normalized plane and screen plane can be determined through the calibration process, and finally the projection transformation matrix from imaging plane to screen can be obtained:

$$H_1^S = H_1^N H_N^S (3)$$

Since the hardware condition based on the homography normalization method must be one camera and four light sources, there are many improvement methods for this problem. Ma [45] proposed the method by proposing three optional geometric transformations of corneal reflection plane, all of which are adaptive, enabling the examination to be carried out with two or three CRs and improved accuracy. Huang [46] can predict the change of homography when the head is in a new position by simulating head changes, so as to calibrate, thus improving the robustness of eye tracking.

The method proposed in this paper is different from the Homography Normalization method. In the coordinate system of 3D world, we use eye coordinate system, imaging plane coordinate system and screen coordinate system to establish the mapping relationship between imaging plane coordinates and pupil center coordinates, and between pupil center coordinates and screen gaze point coordinates through geometric relationship, thus forming the mapping equation based on homography transformation. In the relationship between the imaging plane coordinates and the pupil center coordinates, the pupil center coordinates are obtained by using the camera model(pin-hole imaging). Then, in the relationship between the pupil center coordinates and the screen gaze point coordinates, we use the similarity conditions of similar triangles to calculate the gaze point coordinates. In the homography normalization method, a normalized plane is constructed on the pupil through the four points of the Purkinje image to obtain two mapping matrices, so as to calculate the gaze point position, four CRs points are the basic condition. Although the improved method based on homography normalization method can improve the robustness, it also increases the complexity of the algorithm. Our proposed method can obtain the results only through simple mathematical geometric relationship. From the perspective of hardware, the proposed method only needs a camera and an infrared light to obtain the bright pupil image, instead of using four infrared lights to obtain four CRs points. To sum up, the proposed method has higher accuracy and reduces the cost of hardware, and the algorithm is more simple and feasible.

3. Method

According to the design of the experimental device, this section describes the geometric relationship among the eye pupil center coordinates in the world coordinate system, the corresponding coordinate points in the camera imaging plane coordinate system and the viewpoint coordinates in the scene plane coordinate system. First, as shown in Figure 2, the geometric relationship between eyeball coordinate system and the camera imaging coordinate system is constructed by the structure diagram of the head entrust stent, and the mathematical expression of homography mapping equation is constructed according to the geometric relationship: Suppose the eyeball is a standard sphere, the radius is set to R, and the center of the eyeball is no longer the origin of the eyeball coordinate system, let the position of any point in the eyeball coordinate system be the origin of coordinates, and its coordinate is (x_0, y_0, z_0) . Here, set the eyeball coordinate system as (x_e, y_e, z_e) , the eye image system coordinate system is (x_i, y_i) , the screen coordinate system is (U_s, V_s) , and the distance between the eyeball coordinate system and the screen coordinate system is set to L. The three coordinate systems are coaxial, so the homography relationship between the pupil center and eye image coordinate point and the homography relationship between the screen gaze points and the pupil center can be realized, thus establish the transformation relationship among the three, principle diagram is shown in Figure 3.

Appl. Sci. 2020, 10, 9079 6 of 18

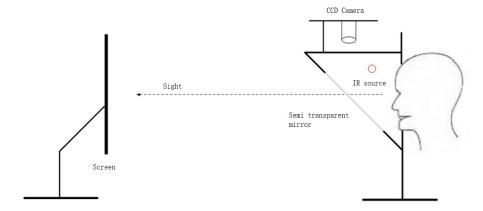


Figure 2. Structure diagram of the head entrust stent.

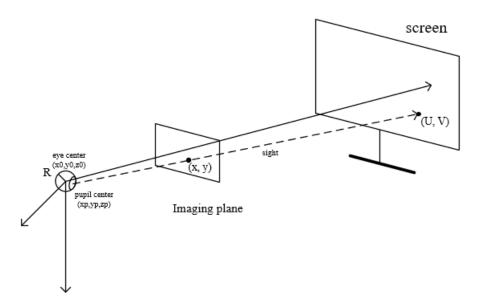


Figure 3. The principle diagram of mapping equation based on homography transformation. The gaze point of the screen is (U, V), the coordinate of the pupil center mapped to the eye diagram of the camera system is (x, y), the coordinates of pupil center is (x_p, y_p, z_p) .

3.1. Pupil Center Location Algorithm

The research of pupil center location has gone through different stages, such as ellipse fitting, Hough transform and other methods. Since the shape of pupil is similar to ellipse, the ellipse fitting algorithm based on least square method [47,48] is widely used. The method used in this paper is ellipse fitting algorithm based on least square method. Firstly, preprocess eye image area features, it includes steps such as grayscale, mean filter, image binarization and contour extraction, then the pupil center is located by ellipse fitting based on the binary image obtained above. It is demonstrated that the pupil center coordinate fitted by this algorithm is accurate.

The general expression of ellipse is:

$$Ax^{2} + Bxy + Cy^{2} + Dx + Ey + F = 0 (4)$$

The least square ellipse fitting method can minimize the sum of squares of measurement errors. The main purpose of this technique is to find a set of parameters to minimize the distance between data points and ellipses. According to the principle of least square method, the problem of curve fitting is transformed into the sum of squares of algebraic distance, so *G* in Equation (5) is the objective

Appl. Sci. 2020, 10, 9079 7 of 18

function. In order to avoid the occurrence of zero solution, constraints are required. In addition, if the fitting result must be ellipse rather than other conic curves, $4AC - B^2 > 0$ needs to be ensured. In this paper, use the constraint method $4AC - B^2 = 1$ [49].

$$G(A, B, C, D, E, F) = \sum_{i=1}^{n} (Ax_i^2 + Bx_iy_i + Cy_i^2 + Dx_i + Ey_i + F)$$
 (5)

The coefficient of the objective function is determined by the minimum value. According to the extreme value principle, the partial derivative of each coefficient is obtained respectively. When the partial derivative is 0, the minimum value of the function can be obtained.

$$\frac{\partial G}{\partial A} = \frac{\partial G}{\partial B} = \frac{\partial G}{\partial C} = \frac{\partial G}{\partial D} = \frac{\partial G}{\partial E} = 0 \tag{6}$$

Here, the center position of the ellipse is represented by (X_0, Y_0) . After the transformation of the expression, the central coordinates of the pupil can be represented by Equations (5) and (6).

$$\begin{cases} X_0 = \frac{BE - 2CD}{4AC - B^2} \\ Y_0 = \frac{BD - 2AE}{4AC - B^2} \end{cases}$$
 (7)

Combined with other constraints, the values of other different coefficients in the elliptic equation can be calculated, and then the center of the ellipse can be obtained.

3.2. Build the Mapping Equation

According to the principle of camera model [50], the principle of pin-hole imaging, it can be concluded that when the eye looks at a point (U, V) on the scene plane, the eye coordinate in the shooting system to which the 3D pupil center coordinates are mapped is (x, y).

Therefore, the conversion steps between pupil center coordinates (x_p, y_p, z_p) and eye diagram imaging coordinate (x, y) are as follows:

$$x_{pic} = \frac{f}{z_p} x_p$$

$$y_{pic} = \frac{f}{z_p} y_p$$
(8)

The relationship between the coordinates of the pupil center coordinate (x_p, y_p, z_p) in eyeball coordinate system and coordinate of the projection points (x_{pic}, y_{pic}) corresponding to the 2d imaging plane of the camera:

(1) The relationship between the 2d pupil center pixel coordinates (x, y) and the projection point coordinates of the imaging plane (x_{pic}, y_{pic}) :

$$x = s_x \cdot x_{pic} + c_x$$

$$y = s_y \cdot x_{pic} + c_y$$
(9)

(2) s_x and s_y respectively indicate the number of pixel points with unit physical distance in the direction x and y of the pixel plane coordinate system, and c_x , c_y represent the migration pixel coordinates generated by the coordinate origin of the projection imaging plane relative to the pixel plane.

Appl. Sci. 2020, 10, 9079 8 of 18

(3) Finally, substituting Equation (8) into Equation (9), we can get the expression between the pupil center coordinate (x_p, y_p, z_p) and the eye diagram imaging coordinate (x, y):

$$x = \frac{s_x \cdot f}{z_p} \cdot x_p + c_x = \frac{f_x}{z_p} \cdot x_p + c_x$$

$$y = \frac{s_y \cdot f}{z_p} \cdot y_p + c_y = \frac{f_y}{z_p} \cdot y_p + c_y$$
(10)

Or conversion form:

$$x_p = \frac{z_p}{f_x} \cdot (x - c_x)$$

$$y_p = \frac{z_p}{f_y} \cdot (y - c_y)$$
(11)

Let $f_x = s_x \cdot f$, $f_y = s_y \cdot f$, written in matrix form:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{z_p} \cdot \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_p \\ y_p \\ z_p \\ 1 \end{bmatrix}$$
(12)

After the camera model is determined, camera calibration is needed to obtain the parameters of the camera. According to the principle of "Zhang Zhengyou calibration method" [51], the chessboard is used as a movable calibration board to calibrate the camera. This paper uses 8×8 chessboard, each grid is 25 mm \times 25 mm square, use the camera to take nine images of the board at different angles, as shown in Figure 4.

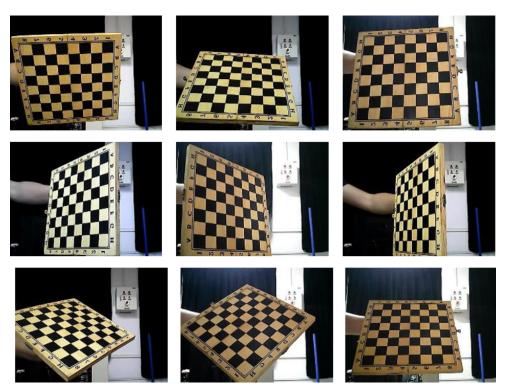


Figure 4. Camera capture chessboard image.

Use the function findChessboardCorners() in OpenCV to obtain approximate data values for the corner points of the board, then use the cornerSubPix() function to calculate the exact corners, and finally draw the corners by the drawChessboardCorners() function. After finding the corners of all Appl. Sci. 2020, 10, 9079 9 of 18

the chessboard charts taken by the camera, the calibrationCamera() function is called to calculate the camera's internal parameter matrix. The values of c_x , c_y , F_x , F_y are obtained according to the internal parameter matrix. These values will be used in the process of solving the mapping equation later.

Next, this paper will study the geometric relationship between the scene viewpoint plane coordinate point (U, V) and the pupil center coordinate (x_p, y_p, z_p) in the world coordinate system:

Suppose that the center of the sphere is converted to the origin of eyeball coordinate system, then the pupil center coordinates (x_p, y_p, z_p) and gaze coordinate points (U, V) will be converted into $(x_p - x_0, y_p - y_0, z_p - z_0)$ and $(U - x_0, V - y_0)$, and combined with the horizontal distance between the human eye and the plane of the viewpoint is L, so according to the similar conditions of the triangle, we can get:

$$\frac{x_p - x_0}{U - x_0} = \frac{y_p - y_0}{V - y_0} = \frac{z_p - z_0}{Z}$$
 (13)

The above formula converts both sides of the equation to calculate the following Formula (14):

$$U = \frac{L}{z_p - z_0} \cdot x_p + (1 - \frac{L}{z_p - z_0}) \cdot x_0$$

$$V = \frac{L}{z_p - z_0} \cdot y_p + (1 - \frac{L}{z_p - z_0}) \cdot y_0$$
(14)

The eyeball is a standard sphere, so it satisfies the standard equation expression of the sphere, and then combines the two conditions of Formula (11):

$$\begin{cases} x_p^2 + y_p^2 + z_p^2 = R^2 \\ x_p = \frac{z_p}{f_x} \cdot (x - c_x) \\ y_p = \frac{z_p}{f_y} \cdot (y - c_y) \end{cases}$$
 (15)

The final expression of z_p on the z-axis of the pupil center can be obtained as follows:

$$z_p = \frac{R \cdot f_x \cdot f_y}{\sqrt{f_y^2 \cdot (x - c_x)^2 + f_x^2 \cdot (y - c_y)^2 + f_x^2 \cdot f_y^2}}$$
(16)

Finally, substituting Equations (11) and (16) into Equation (14), the final expression of the mapping equation based on the homography transformation is:

$$U = \frac{R \cdot L \cdot f_{y} \cdot (x - c_{x}) - x_{0} \cdot L \cdot \sqrt{f_{y}^{2} \cdot (x - c_{x})^{2} + f_{x}^{2} \cdot (y - c_{y})^{2} + f_{x}^{2} \cdot f_{y}^{2}}}{R \cdot f_{x} \cdot f_{y} - z_{0} \sqrt{f_{y}^{2} \cdot (x - c_{x})^{2} + f_{x}^{2} \cdot (y - c_{y})^{2} + f_{x}^{2} \cdot f_{y}^{2}}} + x_{0}$$

$$V = \frac{R \cdot L \cdot f_{y} \cdot (y - c_{y}) - y_{0} \cdot L \cdot \sqrt{f_{y}^{2} \cdot (x - c_{x})^{2} + f_{x}^{2} \cdot (y - c_{y})^{2} + f_{x}^{2} \cdot f_{y}^{2}}}{R \cdot f_{x} \cdot f_{y} - z_{0} \sqrt{f_{y}^{2} \cdot (x - c_{x})^{2} + f_{x}^{2} \cdot (y - c_{y})^{2} + f_{x}^{2} \cdot f_{y}^{2}}} + y_{0}$$

$$(17)$$

This method is called the mapping equation based on homography transformation. In addition, c_x , c_y represent the migration pixel coordinates generated by the coordinate origin of the projection imaging plane relative to the pixel plane. Theoretically, the larger the origin of the migration imaging plane coordinate, the more obvious the function and effect of correcting camera distortion, and the smaller the error of the experimental results. f_x , f_y represents different focal lengths in the x direction and y direction of the pixel plane coordinate system. In addition, x_0 , y_0 , z_0 is an unknown parameter in the mapping equation, and L refers to the horizontal distance between the eyeball coordinate system and the scene screen coordinate system. In theory, the size of this distance will not produce large experimental error, but due to the limited range of human visual observation, this distance can only be limited in the

appropriate range. The coordinates of the origin of the coordinates in the eyeball coordinate system is (x_0, y_0, z_0) , for the above four parameters are solved. In this paper, the Levenberg-Marquardt fitting algorithm (L-M algorithm) is used to implement the process of fitting parameters between polynomials and then to obtained the optimization parameter value.

4. Experiment

The content of this section mainly verifies the accuracy of the above mapping equation by design experiments. Each tester must be required to sit in the limited range and try to keep their head still, as shown in Figure 5. In this paper uses conventional nine-point calibration and nine-point verification to verify the results of the equation, as shown in Figures 6 and 7, error analysis is shown in Table 1.

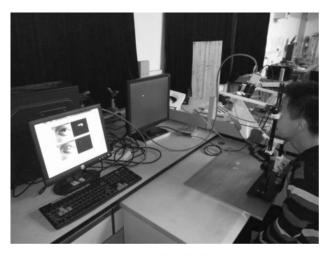


Figure 5. Eye tracker display system.

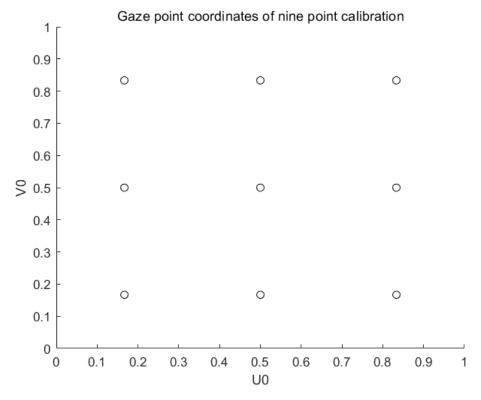


Figure 6. Nine-point calibration results of the standard.

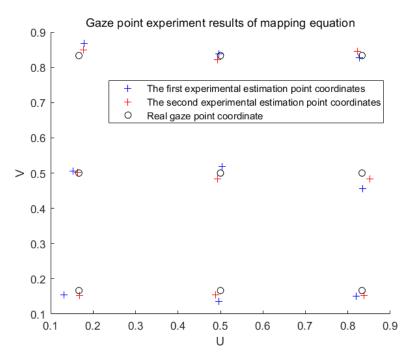


Figure 7. Nine-point verification results of the mapping equation.

Table 1. Error analysis of two ex	perimental results in the nine-	point calibration method (in degrees).

Fixation Point	1	2	3	4	5	6	7	8	9	Accuarcy
experiment 1 (+(blue))°	0.733	0.114	0.182	0.348	0.388	0.872	0.898	0.585	0.469	-
experiment 2 (+(red))°	0.387	0.292	0.371	0.087	0.354	0.554	0.257	0.379	0.287	-
mean value	0.560	0.202	0.276	0.218	0.371	0.713	0.577	0.482	0.378	0.420
variance	0.060	0.016	0.018	0.034	0.001	0.050	0.206	0.021	0.017	0.047

The experimental results show that this method achieves experimental accuracy within 0.5° , has good experimental results, and can be generalized to practical applications. In order to further prove the practicability of the mapping equation based on homography transformation, this article will continue to design comparative experiments and obtain experimental results to prove the advantages of this mapping equation.

According to the current calibration method, any number of calibration points and verification points with the same number can obtain high experimental accuracy. Therefore, this section will use different experimental designs such as nine-point calibration and 16-point verification to illustrate the advantages of this method, and further illustrate the feasibility of improving the mapping equation by comparing the experimental result data with the mapping equation of classical quadratic polynomial. The experimental results obtained are shown in Figure 8.

From the experimental results in Figure 8, it can be clearly seen that the result of advantages and disadvantages of the two method show that the accuracy of the classical quadratic polynomial mapping equation is 0.99° , which is close to 1° . Using the improved mapping method based on homography transformation, it can be seen that the accuracy is basically maintained at about 0.5° . This shows that the design of different calibration points and verification points is more suitable for mapping methods based on homography transformation.

Finally, according to the knowledge of the imaging principle of the projector, this article thought that this improved equation can have good experimental results in the design of small area calibration and large-scale verification. In addition, it is considered that the verification is performed outside

the calibration point, if the effect is ideal, it proves that this method has certain research significance. Therefore, this article sets a nine-point calibration in the center area of the display, and then uses the full-screen uniform distribution nine points for experimental verification, finally experimental results are shown in Figure 9.

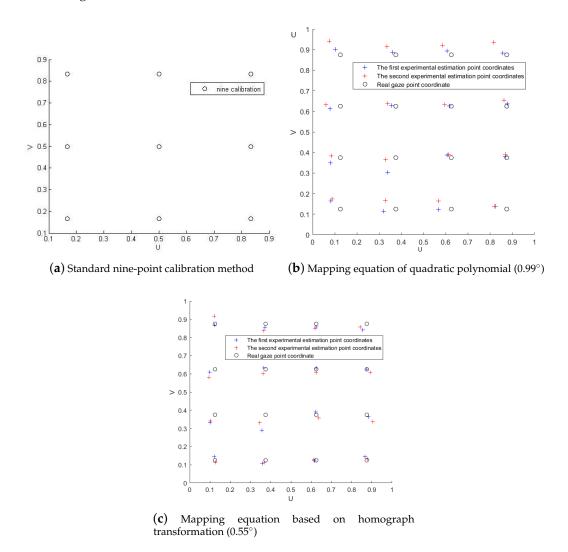


Figure 8. Comparative experimental results of the two methods.

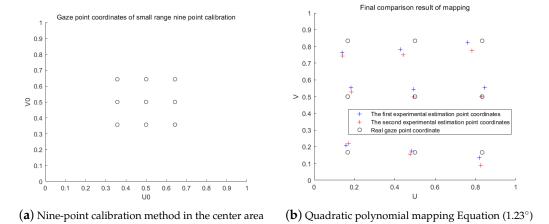
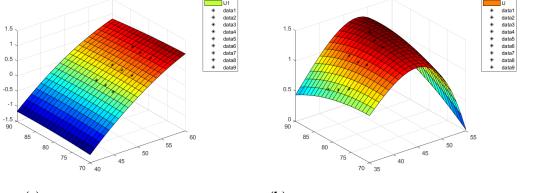
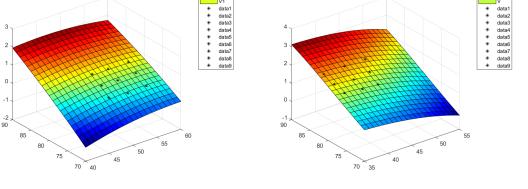


Figure 9. Cont.



(a) Fitting result of quadratic polynomial (b) Fitting results based on homography transformation

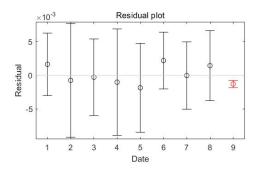


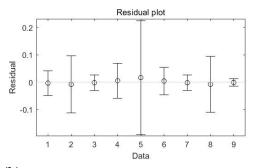
(a) Fitting result of quadratic polynomial (b) Fitting results based on homography transformation

Appl. Sci. 2020, 10, 9079 15 of 18

Secondly, F is F-statistic for short. Generally speaking, it tests the significance of the entire equation. The formula is expressed as the sum of squares divided by the sum of squared errors. This can indicate that the larger the mean square error of the classic quadratic polynomial mapping equation and the smaller the F value, and the lower the confidence of the estimated value.

Finally, we perform residual map regression plots on the two mapping models and judge whether there are exist abnormal points in the calibration data set. Among them, we only do regression diagnosis for abscissa of the two methods, and the analysis results of ordinate are similar. The results are shown in Figure 12.





- (a) Fitting result of quadratic polynomial
- (b) Fitting results based on homography transformation

Figure 12. Schematic diagram of gaze point ordinate surface fitting.

The main concepts involved in the 2d residual graph include: shape, amplitude, residual value, and confidence interval. The residual graph mainly depends on whether the shape of the experimental results is within the specified range and the magnitude of its amplitude is not specified. It can be analyzed from the residual diagram that when the confidence intervals of the residuals (the vertical lines pass through the zero point) contains zero points, it indicates that the regression model fits well, otherwise the worse.

The nine green circles in the figure represent the residual value, and the nine vertical lines represent the range of the residual confidence interval. If the confidence interval basically passed through the origin, it means that the equation fits well, otherwise, the fitting is poor. As can be seen from Figure 12a, the ninth gaze point in quadratic polynomial fails to pass through the origin, and it is a red vertical line, indicating that this point is an abnormal point, which will lead to poor fitting effect.

After many times of non-linear fitting of the data and analysis of the residual plot, the abscissa and ordinate residual plots in the classic quadratic polynomial model will appear abnormal points. The mapping equation based on the homography transformation fitting surface and residuals plot are basically kept in a stable range, and the model fitting effect well. Therefore, the prediction accuracy of this method is higher than the classical quadratic polynomial mapping equation.

By consulting the literature, we compare the accuracy of our method with other methods. The comparison results are shown in the table below.

From Table 3, we can see that our method has simple hardware and high accuracy, so it shows that our method can achieve good experimental accuracy.

Method	Number of Camera	Number of Light Sources	Accuracy
Our method	1	1	0.5°
Morimoto [42]	1	1	1.23°
Ma [46]	1	4	$0.72^{\circ} \pm 0.35^{\circ}$
Choi [39]	1	4	1.03°-1.17°
Cheng [36]	1	5	0.7°

Table 3. Comparison of experimental accuracy of different methods.

6. Conclusions

In this paper, the geometric relationship among eye diagram, camera imaging and scene plane coordinate system are constructed according to the experimental equipment, and a mapping method based on homography transformation is proposed. Because the former experimental device is simple, the precision is very low, and the high precision experimental device is more complex, so we proposes a mapping model based on homography transformation under the single camera and single light source device, the hardware system is simple and feasible. In the experimental part, the accuracy of the experimental results is verified by designing different comparative experiments. The mapping method based on homography transformation can achieve the experimental accuracy of 0.5°, so the experimental results show that the mapping equation can improve the accuracy of the experiment, and can be expected to be applied to practical engineering. Due to the limitations of the experiment, the testers must keep their head as still as possible. Therefore, it is hoped that the head posture compensation algorithm can be added in the following work, so as to realize the free movement of the tester's head.

Author Contributions: Conceptualization, K.L. and X.J.; methodology, K.L. and X.J.; software, X.J.; validation, K.L., X.J. and J.Q.; formal analysis, H.X. and K.L.; investigation, D.L. and X.J.; resources, P.H. and K.L.; data curation, L.P. and X.J.; writing—original draft preparation, X.J.; writing—review and editing, K.L.; visualization, K.L. and X.J.; supervision, P.H. and J.Q.; funding acquisition, P.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 numbers 61975058); Science and Technology Project of Guangzhou, China (grant numbers 201707010485, 201704020137); The National Natural Science Foundation of China Guangdong big data Science Center Project (grant numbers U1911401).

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

References

- 1. Hansen, D.W.; Ji, Q. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, 32, 478–500. [CrossRef] [PubMed]
- 2. Baluja, S.; Pomerleau, D. Non-Intrusive Gaze Tracking Using Artificial Neural Networks. In *Advances in Neural Information Processing Systems 6*; Technical Report; DTIC Document; ACM: New York, NY, USA, 1994.
- 3. Holl, C.; Komogortsev, O. Eye tracking on unmodified common tablets: Challenges and solutions. *Eye Track. Res. Appl.* **1994**, 1, 277–280.
- 4. Xu, L.; Machin, D.; Sheppard, P. A novel approach to real-time non-intrusive gaze finding. In Proceedings of the British Machine Vision Conference (BMVC), York, UK, 1 January 1998; pp. 1–10.
- 5. Hansen, D.W.; Hansen, J.P.; Nielsen, M.; Johansen, A.S.; Stegmann, M.B. Eye typing using Markov and active appearance models. In Proceedings of the IEEE Conference on Applications of Computer Vision Workshops, Orlando, FL, USA, 4 December 2002; pp. 132–136.
- 6. Sugano, Y.; Matsushita, Y.; Sato, Y. Appearance-based gaze estimation using visual saliency. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 329–341. [CrossRef]
- 7. White, K.P., Jr.; Hutchinson, T.E.; Carley, J.M. Spatially dynamic calibration of an eye-tracking system. *IEEE Trans. Syst. Man Cybern.* **1993**, 23, 1162–1168. [CrossRef]
- 8. Newman, R.; Matsumoto, Y.; Rougeaux, S.; Zelinsky, A. Real-time stereo tracking for head pose and gaze estimation. In Proceedings of the IEEE Conference on Automatic Face and Gesture Recognition, Grenoble, France, 28–30 March 2000; pp. 122–128.
- 9. Morimoto, C.H.; Amir, A.; Flickner, M. Detecting eye position and gaze from a single camera and 2 light sources. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002; pp. 314–317.
- 10. Noureddin, B.; Lawrence, P.D.; Man, C. A non-contact device for tracking gaze in a human computer interface. *Comput. Vis. Image Underst.* **2005**, *98*, 52–82. [CrossRef]
- 11. Wang, J.G.; Sung, E.; Venkateswarlu, R. Estimating the eye gaze from one eye. *Comput. Vis. Image Underst.* **2005**, *98*, 83–103.

12. Guestrin, E.D.; Eizenman, E. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 1124–1133. [CrossRef] [PubMed]

- 13. Meyer, A.; Böhme, M.; Martinetz, T.; Barth, E. A single-camera remote eye tracker. In *Perception and Interactive Technologies*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 208–211.
- Hennessey, C.; Noureddin, B.; Lawrence, P. A single camera eye-gaze tracking system with free head motion. In Proceedings of the Symposium on Eye Tracking Research and Applications, San Diego, CA, USA, 27–29 March 2006; pp. 87–94.
- 15. Eizenman, E. An automatic personal calibration procedure for advanced gaze estimation systems. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1031–1039.
- 16. Villanueva, A.; Cabeza, R.; SPorta, S. Eye tracking: Pupil orientation geometrical modeling. *Image Vis. Comput.* **2006**, 24, 663–679. [CrossRef]
- 17. Villanueva, A.; Cabeza, R.; SPorta, S. Gaze tracking system model based on physical parameters. *Int. J. Pattern Recognit. Artif. Intell.* **2007**, 21, 855–877. [CrossRef]
- 18. Nagamatsu, T.; Hiroe, M.; Rigoll, G. Corneal-Reflection-Based Wide Range Gaze Tracking for a Car. In Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019; pp. 385–400.
- 19. Brolly, X.L.; Mulligan, J.B. Implicit calibration of a remote gaze tracker. In Proceedings of the IEEE Conference on Applications of Computer Vision Workshop, Washington, DC, USA, 27 June–2 July 2004; p. 134.
- Ebisawa, Y.; Satoh, S.I. Effectiveness of pupil area detection technique using two light sources and image difference method. In Proceedings of the IEEE Conference on Engineering in Medicine and Biology Society, San Diego, CA, USA, 31 October 1993; pp. 1268–1269.
- 21. Hansen, D.W.; Pece, A.E. Eye tracking in the wild. Comput. Vis. Image Underst. 2005, 98, 155–181. [CrossRef]
- 22. Ji, Q.; Yang, X. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging* **2002**, *8*, 257–377. [CrossRef]
- 23. Williams, O.; Blake, A. Sparse and semi-supervised visual mapping with the s^3gp. *IEEE Conf. Comput. Vis. Pattern Recognit. Work* **2006**, *8*, 230–237.
- 24. Cerrolaza, J.J.; Villanueva, A.; Cabeza, R. Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems. In Proceedings of the Symposium on Eye Tracking Research and Applications, Savannah, GA, USA, 26–28 March 2008; pp. 259–266.
- 25. Cerrolaza, J.J.; Villanueva, A.; Cabeza, R. Error characterization and compensation in eye tracking systems. In Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, CA, USA, 28–30 March 2012; pp. 205–208.
- 26. Zhu, J.; Yang, J. Subpixel eye gaze tracking. In Proceedings of the Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 21 May 2002; pp. 124–129.
- 27. Xia, L.; Sheng, B.; Wu, W.; Ma, L.; Li, P. Accurate gaze tracking from single camera using gabor corner detector. *Multimed. Tools Appl.* **2016**, 75, 221–239. [CrossRef]
- 28. Shao, G.; Che, M.; Zhang, B.; Cen, K.; Gao, W. A novel simple 2d model of eye gaze estimation. In Proceedings of the IEEE Conference on Intelligent Human Machine Systems and Cybernetics, Nanjing, China, 26–28 August 2010; pp. 300–304.
- 29. Blignaut, P. Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker. *J. Eye Mov. Res.* **2000**, *18*, 331–335.
- 30. George, A.; Routray, A. Fast and accurate algorithmfor eye localisation for gaze tracking in low-resolution images. *IET Comput. Vis.* **2016**, *10*, 660–669. [CrossRef]
- Tong, Q.; Hua, X.; Qiu, J. A new mapping function in table-mounted eye tracker. In Proceedings of the 2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology, Tianjin, China, 12 January 2018; Volume 10620.
- 32. Li, W.; Che, M.; Li, F. Gaze Estimation Research with Single Camera. In Proceedings of the Conference on e Business Techology and Strategy, Tianjin. China, 1 January 2012; pp. 592–599.
- 33. Zhang, C.; Chi, J.N.; Zhang, Z.; Gao, X.; Hu, T.; Wang, Z. Gaze estimation in a gaze tracking system. *Sci. China Inf. Sci.* **2011**, *54*, 2295–2306. [CrossRef]
- 34. Morimoto, C.H.; Coutinho, F.L.; Hansen, D.W. Screen-Light Decomposition Framework for Point-of-Gaze Estimation Using a Single Uncalibrated Camera and Multiple Light Sources. *J. Math. Imaging Vis.* **2020**, 62, 585–605. [CrossRef]

Appl. Sci. 2020, 10, 9079 18 of 18

35. Wang, J.; Zhang, G.; Shi, J. 2D Gaze Estimation Based on Pupil-Glint Vector Using an Artificial Neural Network. *Appl. Sci.* **2016**, *6*, 174. [CrossRef]

- 36. Cheng, H.; Liu, Y.; Fu, W.; Ji, Y.; Yang, L.; Zhao, Y. Gazing point dependent eye gaze estimation. *Pattern Recognit.* **2017**, *71*, 36–44. [CrossRef]
- 37. Dong, H.Y.; Chung, M.J. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Comput. Vis. Image Underst.* **2005**, *98*, 25–51.
- 38. Choi, K.A.; Ma, C.; Ko, S.J. Improving the usability of remote eye gaze tracking for human-device interaction. *IEEE Trans.Consumer Electron.* **2014**, *60*, 493–498. [CrossRef]
- 39. Hansen, D.W.; Agustin, J.S.; Villanueva, A. Homography normalization for robust gaze estimation in uncalibrated setups. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Austin, TX, USA, 22–24 March 2010; pp. 13–20.
- 40. Zhu, Z.; Ji, Q. Novel eye gaze tracking techniques under natural head movement. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 2246–2259. [PubMed]
- 41. Merchant, J.; Morrissette, R.; Proterfield, J.L. Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Trans. Biomed. Eng.* **1974**, *21*, 309–317. [CrossRef] [PubMed]
- 42. Morimoto, C.H.; Koons, D.; Amir, A.; Flickner, M. Pupil detection and tracking multiple light sources. *Image Vis. Comput.* **2000**, *18*, 331–335. [CrossRef]
- 43. Morimoto, C.H.; Mimica, M. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.* **2005**, *98*, 4–24. [CrossRef]
- 44. Zhang, C.; Chi, J.N.; Zhang, Z.H.; Wang, Z.L. A novel eye gaze tracking technique based on pupil center cornea reflection technique. *Chin. J. Comput.* **2010**, *33*, 1273–1287. [CrossRef]
- 45. Ma, C.; Baek, S.J.; Choi, K.A.; Ko, S.J. Improved remote gaze estimation using corneal reflection-adaptive geometric transforms. *Opt. Eng.* **2014**, *53*, 053112. [CrossRef]
- 46. Huang, J.-B.; Cai, Q.; Liu, Z.; Ahuja, N.; Zhang, Z. Towards accurate and robust cross-ratio based gaze trackers through learning from simulation. In Proceedings of the Symposium on Eye Tracking Research and Applications, Safety Harbor, FL, USA, 26–28 March 2014; pp. 75–82.
- 47. Fitzgibbon, A.W.; Pilu, M.; Fisher, R.B. Direct least square fitting of ellipses. *IEEE Trans. Pattern Anal. Mach. Intell.* 1999, 21, 476–480. [CrossRef]
- 48. Ruan, Y.; Zhou, H.; Huang, J. An Improved Method for Human Eye State Detection Based on Least Square Ellipse Fitting Algorithm. In Proceedings of the 2018 13th World Congress on Intelligent Control and Automation (WCICA), Changsha, China, 4–8 July 2018.
- 49. Fitzgibbon, A.W.; Fisher, R.B. A Buyer's Guide to Conic Fitting. In Proceedings of the British Machine Vision Conference, Birmingham, UK, 1 September 1995.
- 50. Hartley, R.; Zisserman, A.F. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2000.
- 51. 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.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 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/).