A Straightforward Camera Calibration Method Based on a Single Low-cost Cubical Target

Yuezhen Cai[1], Linyuan Xia[1,*], Ting On Chan[1]

[1] School of Geography and Planning, Sun Yat-sen University, 510275 Guangzhou, China – caiyzh5@mail2.sysu.edu.cn, xialiny@mail.sysu.edu.cn (corresponding author), chantingon@mail.sysu.edu.cn

Keywords: Calibration, Cubical dice, Geometric Constraints, Lens Distortion, Camera Intrinsics.

Abstract

Calibrating optical sensors with common targets facilitates the efficient and convenient acquisition of the sensor's internal parameters. In this paper, we present a new method of camera calibration utilizing a low-cost foamy cube, in a form of dice, which is based on the fact that arrangement of pip and cubical die surfaces is mutually orthogonal. Initially, each face and pips are identified through the color information on the die's surfaces. Subsequently, the centers of pips are corrected using a circular projection model, and radial distortion coefficients are estimated based on centers' one-to-one correspondences. After that, the tangent information between pairs of pips on orthogonal dice faces are utilized to compute vanishing points, leading to estimation of intrinsic parameters. Experimental results demonstrate that our method has similar effects compared to well-known checkerboard calibration method, reaching an average relative error of 2.43%, simplifying the calibration process in practical applications and showcasing good practicality and robustness.

1. Introduction

Camera calibration is a crucial step in many computer vision and robotics applications, as it compensates internal geometric errors and restores camera orientations in most cases. . Serving as a critical prelude to advanced applications ranging from autonomous navigation systems to augmented reality, the process of calibrating a camera transcends mere adjustment of optical variables—it embodies the harmonization of the digital eye with the geometrical intricacies of the physical world.

Calibration methods can be broadly classified into three categories: target-based methods, self-calibration methods, and active vision-based methods (Song et al., 2013). The target-based calibration, the most widely used approach, relies on known geometric patterns. Traditional methods utilize specific calibration objects, such as checkerboards (Zhang, 2000), Concentric circle grids (Bu et al., 2021), or deltille grids (Ha et al., 2017). With the gradual popularization of deep learning technology, self-calibration methods enable automatic estimation of camera parameters without the need for external calibration equipment, significantly enhancing system flexibility and ease of use (Liao et al., 2023). In other hand, these methods rely on the accuracy and robustness of algorithms and require diverse scene content and motion, potentially underperforming in texture-poor or dynamic environments. Active vision-based methods require camera execute designed actions to actively encode calibration information (Chen et al., 2024), achieving high accuracy but requiring specialized hardware and precise synchronization.

Despite the advances in self-calibration and active vision-based methods, target-based calibration remains the most reliable and widely adopted approach. However, the reliance on specialized calibration objects poses certain limitations, including flexibility in calibration setup and the increased cost and effort to obtain precise calibration targets (Luhmann et al., 2016). To address this issue, researchers have explored the use of common objects as calibration targets, such as coplanar coins (Bergamasco et al., 2014), spheres (Roman-Rivera et al., 2022), or even human faces (Nasir and Rao, 2016). These common objects offer greater accessibility and flexibility compared to traditional calibration targets.

To enhance the robustness of calibration methods with common objects while retaining their convenience of use, building upon the basis of our previous research (Chan et al., 2023), we propose a straightforward calibration method based on a low-cost cubical target made of light material such as foam, a die-like object. Our method utilizes the unique and rich texture characteristics of dice, including unique distribution of pips and vibrant color information on the six faces, along with the mutual orthogonality of neighborhood dice, to increase the calibration process robustness. Additionally, the simplicity and ubiquity of dice make this method highly accessible and cost-effective, reducing time required for set-up and troubleshooting. Users only need to hold the imaging sensor and collect images around the dice, without the need to move calibration targets or perform specific movement patterns.

The proposed method does not require any prior professional knowledge, including the radius and relative distance of the pips on dice surface, or preset internal parameters of the camera. The significance of our method lies in its potential to make camera calibration more accessible and adaptable to various real-world situations, ultimately benefiting a wide range of applications that rely on accurate camera parameters. The rest of this paper is organized as follows: Section 2 presents the mathematical model and calibration principles based on the geometric characteristics of dice. Section 3 describes the experimental design to validate our calibration procedure, with experimental results analyzation. Finally, Section 4 concludes the paper and discusses future research directions.

2. Methodology

Figure 1 illustrates the workflow of our proposed method: Initially, the system identifies the dice regions from a sequence of images captured by the RGB sensor, which correspond to the pips on each dice face. Since these pips generally appear as ellipses in the captured images unless the dice faces are exactly

parallel to the image sensor plane, a circle-to-ellipse projection model is performed to correct the center of circles on image, and a distortion estimation model is employed with nine corresponding points. Once the distortions are corrected, the intrinsic parameters are computed by exploiting the geometric constraints provided by the mutually orthogonal faces of the dice. These constraints allow for the estimation of the intrinsic parameters based on the known geometry of the dice.
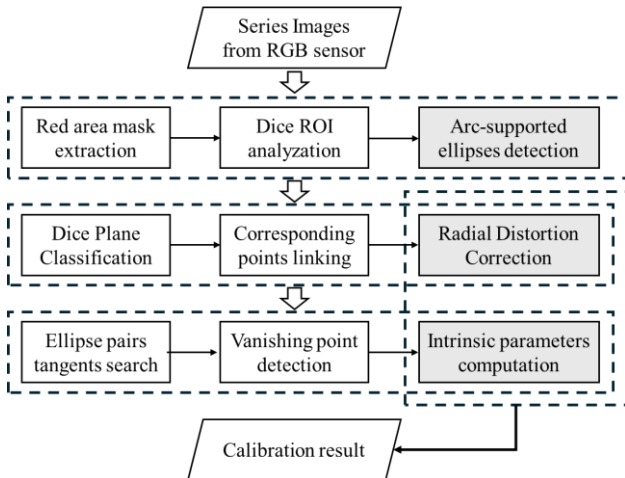

Figure 1. Workflow of our proposed method.

The innovative aspects of this method lie in its simplicity, flexibility, and the ability to achieve accurate camera calibration using common objects. By utilizing the inherent geometric properties of dice, the need for specialized calibration patterns or precise measurements is eliminated. The use of readily available objects, such as dice, makes the calibration process more accessible and practical in various settings. Furthermore, the incorporation of distortion correction and the exploitation of the orthogonality constraints of the dice faces enhance the accuracy and robustness of the calibration results.

Our proposed method does not require any prior knowledge of the intrinsic parameters. The abundant geometric features and correspondence information from common objects, combined with the availability of multiple viewpoints, allows for the accurate calibration of camera parameters without depending on prior knowledge of intrinsic parameters, as the problem is well-constrained. The subsequent sections of the paper will provide a more detailed explanation of each step in the proposed method, including the arc-supported ellipses detection for identifying dice regions, the radial distortion correction models, and the intrinsic parameters computation using vanishing points calculated by orthogonality constraints of dice faces.

2.1 Pip detection from dice

The HSV color model is used to detect the dice position and extract the pips, as it is more efficient for tracking objects with clear color properties under natural lighting conditions (He et al., 2023). The images are converted from RGB to HSV color space using the OpenCV library, and the red background color of the dice is defined by a specific range in the HSV space as follows:

$$\begin{cases} H \in [0,10] \cup [170,180] \\ S \in [120,255] \\ V \in [70,255] \end{cases} . \qquad (1)$$

The red regions are extracted, converted to a binary mask, and processed using contour detection to identify the continuous red regions. To improve the reliability of subsequent pip detection, morphological dilation and erosion operations were applied to the red regions to remove small holes, and regions with small area were filtered. After that, bounding boxes were generated for each valid contour to localize possible red die in the image.

Within each bounding box, the circular pips, which appear as ellipses due to perspective projection, are detected using a fast ellipse detection method based on arc-support line segments (ASLSs) (Lu et al., 2020). The region inside the bounding box is converted to grayscale, and Canny edge detection is applied. ASLSs are extracted from the edge map, and ellipses are fitted to the ASLS groups. The ellipses are filtered by size and aspect ratio, and the red die is localized to the bounding box with the most ellipses, each representing a pip on the die.

Figure 2 shows example intermediate results of a single red dice segmentation and white pips detection. By leveraging color information and geometrical constraints, the proposed method robustly extracts the die and pips.
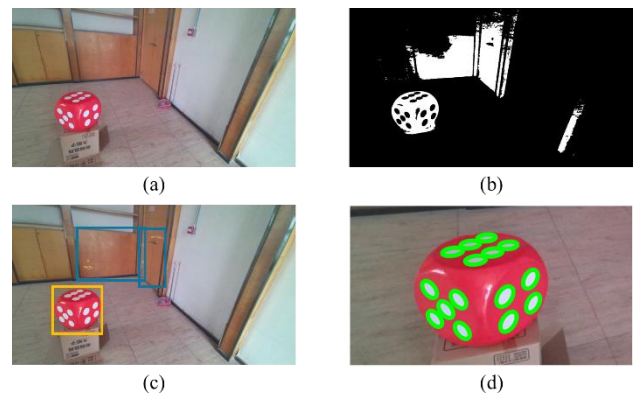

Figure 2. Detection progress of pips on a die. (a) original image; (b) red area mask; (c) dice region of interest detection; (d) edges of localized pips on dice.

2.2 Radial Distortion Correction

After detecting the elliptical pips on each die face, they were associated to different planes of the cube based on their geometric properties and spatial relationships. Then, a one-to-one correspondence could be established between the same pip captured in different images based on the relative position of each face.

However, due to the principle of perspective projection, the center of each circular pip did not coincide with the center of its projected ellipse, as illustrated in Figure 3. To find the corresponding points for radial distortion correction, the offset between the projected circle center and the ellipse center was computed using the circular projection model proposed by Matsuoka and Maruyama (2016). To be more specific, let (x_c, y_c, z_c) be the center of circle in camera coordinate system, and the ellipse $\mathbf{E}$ with center (x_e, y_e) on image coordinate system satisfy

$$\mathbf{v}^{\mathrm{T}} \mathbf{E} \mathbf{v} = 0, \qquad (2)$$

where

$$\begin{cases} \mathbf{E} = \begin{bmatrix} a & c & d \\ c & b & e \\ d & e & m \end{bmatrix}, \\ \mathbf{v} = \begin{bmatrix} x_e & y_e & f \end{bmatrix}^{\mathrm{T}} \end{cases} \qquad (3)$$

where a, b, c, d, e, m are the parameters of ellipse (which can be calculated by ellipse fitting with its edge points extracted above), and f stands for the focal length of camera. Since we do not use any prior knowledge of focal length, f is initially set to zero, and then replaced to computed result when it is iterative optimized. The offset of (x_e, y_e) can be computed as

$$\begin{cases} \delta x_e = -f \left(\dfrac{bd - ce}{ab - c^2} - \dfrac{x_c}{z_c} \right), \\ \delta y_e = -f \left(\dfrac{ae - cd}{ab - c^2} - \dfrac{y_c}{z_c} \right) \end{cases} \qquad (4)$$

if (x_c, y_c, z_c) is unknown, it can be solved by simultaneously solving a system of equations constructed by Matsuoka and Maruyama (2016).
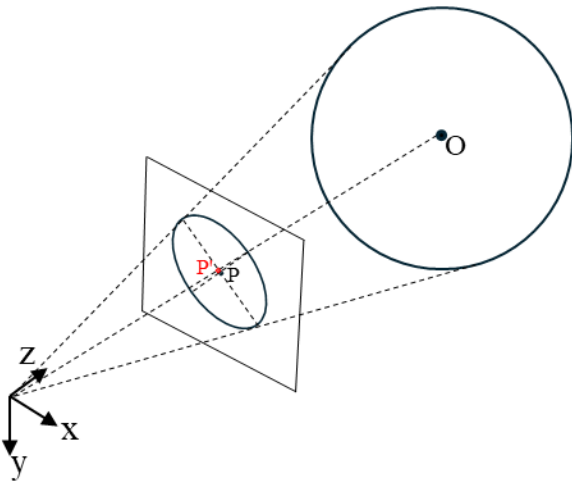

Figure 3. Circular projection and the center offset on image plane.

The corrected pip centers from multiple die faces provided point-to-point correspondences, which were then used to estimate the radial distortion coefficients following the method of Li and Harley (2005). We implement a one-parameter division model (Fitzgibbon, 2001) to describe the radial distortion:

$$\mathbf{p}_u = \frac{\mathbf{p}_d}{1 + k r_d^2}, \qquad (5)$$

where
    $\mathbf{p}_u$ = undistorted pixel coordinate
    $\mathbf{p}_c$ = distorted pixel coordinate
    k = coefficient of distortion model
    r_d = pixel radius from $\mathbf{p}_c$ to image center

Suppose $\mathbf{p}'_u$ and $\mathbf{p}_u$ be a correspondent pair in two images, their relationship is

$$\mathbf{p}'_u{}^{\mathrm{T}} \mathbf{F} \mathbf{p}_u = 0, \qquad (6)$$

where $\mathbf{F}$ represents fundamental matrix of camera. By using the Kronecker product $\otimes$ to rearrange the above equation, we get:

$$\left(\mathbf{p}'_u \otimes \mathbf{p}_u \right) \cdot \mathrm{vec}\left(\mathbf{F}^{\mathrm{T}} \right) = 0. \qquad (7)$$

Let $\mathbf{M} = \mathbf{p}'_u \otimes \mathbf{p}_u$, we can write down two nonlinear equations with two singularity conditions

$$\begin{cases} \det(\mathbf{M}) = 0 \\ \det\left(\mathrm{mat}\left(\mathrm{null}(\mathbf{M}) \right) \right)] = 0 \end{cases} \qquad (8)$$

where $\mathrm{null}(\cdot)$ = null-space operator
              $\mathrm{mat}(\cdot)$ = matrix operator, turning vector into matrix

We can get Equation (8) after finding correspondences, solving them will yield numerous solutions for different values of coefficient k. Hence, a kernel voting scheme was employed to determine the coefficient with the maximum likelihood. For each estimate, a Gaussian kernel was applied to model its contribution to the final coefficient. The kernel bandwidth was set adaptively based on the number and consistency of the point correspondences. The distortion coefficient with the highest voted confidence were selected as result for accurate radial distortion correction.

2.3 Camera Intrinsic Parameter Estimation

With the radial distortion coefficients estimated from the previous step, the distortions in the captured images could be corrected. This allowed for the estimation of the camera's intrinsic parameters by exploiting the orthogonal geometric constraints provided by a die.

The key idea to estimate intrinsic parameters was to utilize the tangent information between pairs of ellipses detected on orthogonal dice faces, which can be intersected at vanishing points. As illustrated in Figure 4, under perspective projection, the tangent lines of two ellipses originating from orthogonal circles would intersect at the vanishing point corresponding to the direction perpendicular to both circles. In this paper, we use methods from Da et al. (2012) to compute the tangent lines of two projected circles. By identifying multiple pairs of ellipses with orthogonal relationships, vanishing points in mutually orthogonal directions could be derived. To estimate the vanishing points, we use least-square method to get the intersection of tangent line segments. Based on the principles of perspective projection, the positions of the vanishing points in the image were directly related to the camera's focal length and the principal point location.
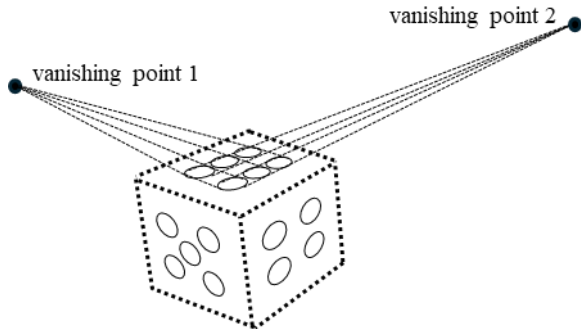
Figure 4. Vanishing point extraction based on geometry features on dice planes.

Specifically, the homogeneous coordinates of the vanishing points could be expressed as functions of the intrinsic parameters (Hartley and Zisserman, 2003). To explain in detail, let's identify the camera matrix in a common type

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

with focal lengths f_x and f_y, principal point (c_x, c_y), and skew coefficient s. Assume that we have orthogonal vanishing points $\mathbf{p}_1 = (x_1, y_1, 1)^T$ and $\mathbf{p}_2 = (x_2, y_2, 1)^T$ in homogeneous coordinates, they satisfy

$$\mathbf{p}_1 \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{p}_2 = 0. \tag{10}$$

Matrix $\mathbf{K}^{-T} \mathbf{K}^{-1}$ is symmetric, we can set

$$\mathbf{M} = \mathbf{K}^{-T} \mathbf{K}^{-1} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{12} & m_{22} & m_{23} \\ m_{13} & m_{23} & m_{33} \end{bmatrix}, \tag{11}$$

substituting $\mathbf{M}$ into Equation (10), we can get the following linear equation by expanding it

$$\begin{bmatrix} x_1 x_2 \\ x_1 y_2 + x_2 y_1 \\ x_1 + x_2 \\ y_1 y_2 \\ y_1 + y_2 \\ 1 \end{bmatrix}^T \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{22} \\ m_{23} \\ m_{33} \end{bmatrix} = 0 \tag{12}$$

By stacking more than six above functions, we can solve the coefficients in matrix $\mathbf{M}$. Then, camera matrix $\mathbf{K}$ can be computed by Cholesky factorization and matrix inversion, and we can get intrinsic parameters from it.

The proposed method leveraged the rich geometric information encoded in the dice faces to establish reliable constraints for intrinsic parameter estimation. By combining the estimates from multiple dice and images, a robust solution could be obtained even in the presence of noise and outliers.

3. Experiment

To validate the proposed method, experiments were conducted using an Azure Kinect sensor (Figure 5). The Azure Kinect is a high-resolution RGB-D camera developed by Microsoft, which provides synchronized color and depth images. It was chosen as the experimental device due to its obvious radial distortion in raw images and the availability of manufacturer-provided intrinsic parameters, which could serve as reference values for evaluating the accuracy of our estimation.

Figure 5. The Azure Kinect sensor.

In our experiment, we captured a set of images containing the common-used die from various viewpoints and distances using the Azure Kinect image sensor in 2160p (3840×2160) resolution. The dice were placed on a flat surface, and the camera was moved around to capture images from different angles. To compare the accuracy of our method, we select the wide-used checkerboard calibration method (Zhang, 2000) as a reference. Since we used a common object for calibration, we did not employ a professional calibration checkerboard; instead, we used a calibration checkerboard printed on A4 papers in advance, which is the method commonly used by most people. Moreover, this approach is also quite common in research areas such as SLAM (Simultaneous Localization and Mapping).

3.1 Result of radial distortion estimation

We used multiple photos to estimate the radial distortion value, and different combinations of photos can produce different estimates, after performing kernel density estimation on these estimates, we can see that the peak of the distortion estimate is quite pronounced. The corresponding kernel density is 0.3479, which can reflect the radial distortion parameter relatively accurately, as illustrated in Figure 6.
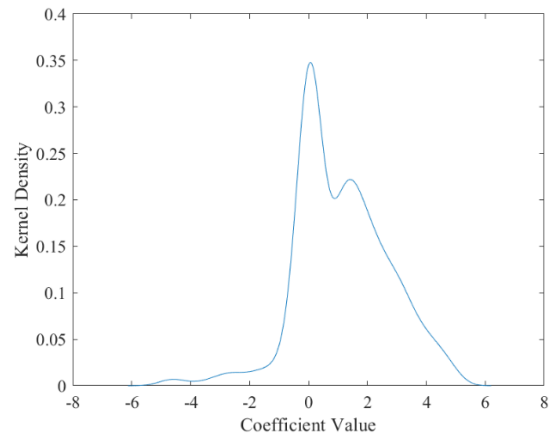

Figure 6. Estimating distortion parameter by kernel voting scheme.

Methods	Radial Distortion Parameters
Manufacture's initial value	k_1=0.479, k_2=-2.685, k_3=1.609, k_4=0.357, k_5=-2.505, k_6=1.530
Zhang's method	k_1=0.1031, k_2=-0.027
Our method	k=0.0487

Table 1. Radial distortion parameters estimated by different methods

Due to the use of different distortion models in various calibration methods, we are unable to quantitatively compare our results with theirs, as shown in Table 1. However, as shown in Figure 7, our distortion correction effect is comparable to the checkerboard calibration method provided by Zhang, demonstrating the effectiveness of our approach.

(a)   (b)

Figure 7. Undistorted images after radial distortion removal. (a) Zhang's checkerboard method; (b) our method.

3.2 Result of intrinsic parameters

The estimated intrinsic parameters, including focal lengths and coordinate of principal point, were compared with the manufacturer's initial values and those obtained by Zhang's method, which are tabulated in Table 2..

Estimated parameters		manufacturer's initial values	Zhang's method	Our method
Focal length	f_x	1799.6	1875.6	1825.8
	f_y	1799.7	1879.5	1796.0
Principal Point	c_x	1911.8	1900.4	2047.8
	c_y	1100.0	1095.1	1110.5

Table 2. intrinsic parameters (focal length, principal point) estimated by different methods

To comprehensively evaluate the performance of our method is estimating intrinsic parameters, we have employed relative error based on manufacturer's initial values, which is computed by the percentage difference relative to the initial values. Specifically, for the estimation of focal lengths f_x and f_y, our method yields results of 1825.8 and 1796.0 with 1.45% and 0.21% relative errors, respectively, which are closer to the manufacturer's initial values compared to Zhang's method. However, it should be noted that there is still a significant discrepancy for the principal point coordinates between our estimates and the initial values, particularly for the c_x coordinate with 7.11% relative error. In terms of overall precision, namely average relative error, we get a very close relative error compare to Zhang's method:2.43% for ours and 2.42% for Zhang's.

3.3 Discussion

Based on the data from Table 1 and Table 2, our method has shown a certain level of effectiveness in estimating radial distortion and focal length parameters. As we did not use a high-precision calibration board for calibration, it is likely that Zhang's method absorbed some of the error from feature points during the bundle adjustment, resulting in its calibration of focal length being less ideal than our method. To explain the poorer performance obtained in the calibration of the principal point, we believe this may be due to our estimation of matrix **K** lacking more constraints, as well as the relatively large residuals of some extracted vanishing points. This suggests that further improvements may be needed to more accurately determine the principal point coordinates, such as providing accurate initial value guess or offering upper and lower limitation boundaries for final result.

4. Conclusion

In this study, we propose a novel and straightforward calibration method for digital imaging sensors using a low-cost and easily accessible cubical target, a foamy die. By leveraging the unique texture features and geometric properties of dice, our method demonstrates robustness and flexibility in various environments while maintaining user-friendliness. The proposed workflow effectively estimates distortion parameters and intrinsic camera parameters without requiring prior professional knowledge or specific movement patterns. Experiments validate the accuracy and potential of our approach, with less than 1% average difference compared to well-known Zhang's checkerboard calibration method.

In future research, we will focus on enhancing the robustness and applicability of the proposed calibration method by handling more complex scenarios, integrating deep learning techniques for improved accuracy and efficiency, and conducting extensive evaluations and comparisons with state-of-the-art methods to facilitate wider adoption in practical applications.

References

Bergamasco, F., Cosmo, L., Albarelli, A., Torsello, A., 2014. Camera calibration from coplanar circles. *2014 22nd International Conference on Pattern Recognition*, 2137-2142. doi.org/10.1109/ICPR.2014.372.

Bu, L., Huo, H., Liu, X., Bu, F., 2021. Concentric circle grids for camera calibration with considering lens distortion. *Optics and Lasers in Engineering*, 140, 106527. doi.org/10.1016/j.optlaseng.2020.106527.

Chan, T.O., Xia, L., Lichti, D.D., Wang, X., Peng, X., Cai, Y., Li, M.H., 2023. A Method for Turning a Single Low-Cost Cube into a Reference Target for Point Cloud Registration. *Applied Sciences* 13, 1306. doi.org/10.3390/app13031306.

Chen, Z., Li, R., Tian, F., Liang, J., Si, X., Zhang, B., 2024. A Novel Method for Camera Focal Lengths Calibration Based on Active Vision. *International Conference on Advanced Unmanned Aerial Systems*, 185-195.

Da, F., Li, Q., Zhang, H., & Fang, X. (2012). Self-calibration using two same circles. *Optics & Laser Technology*, 44(6), 1924-1933. doi.org/10.1016/j.optlastec.2012.02.013.

Fitzgibbon, A. W., 2001. Simultaneous linear estimation of multiple view geometry and lens distortion. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1, I-I.

Ha, H., Perdoch, M., Alismail, H., So Kweon, I., Sheikh, Y., 2017. Deltille grids for geometric camera calibration. *Proceedings of the IEEE International Conference on Computer Vision*, 5344-5352. doi.org/10.1109/ICCV.2017.571.

Hartley, R., Zisserman, A., 2003. *Multiple view geometry in computer vision*. Cambridge university press.

He, W., Wu, G., Fan, F., Liu, Z., Zhou, S., 2023. A Robust Real-Time Ellipse Detection Method for Robot Applications. *Drones* 7(3), 209.

Li, H., Hartley, R., 2005. A non-iterative method for correcting lens distortion from nine point correspondences. *Omnidirectional Vision, Camera Networks, and Non-Classical Cameras,* 2(7), 2.

Liao, K., Nie, L., Huang, S., Lin, C., Zhang, J., Zhao, Y., Gabbouj, M., Tao, D. (2023). Deep learning for camera calibration and beyond: A survey. *arXiv preprint arXiv:2303.10559*.

Lu, C., Xia, S., Shao, M., Fu, Y., 2020. Arc-Support Line Segments Revisited: An Efficient High-Quality Ellipse Detection. *IEEE Trans. on Image Process*, 29, 768–781. doi.org/10.1109/TIP.2019.2934352.

Luhmann, T., Fraser, C., Maas, H.-G., 2016. Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 37–46. doi.org/10.1016/j.isprsjprs.2015.10.006.

Matsuoka, R., Maruyama, S., 2016. Eccentricity on an image caused by projection of a circle and a sphere. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 19-26. doi.org/10.5194/isprs-annals-III-5-19-2016.

Nasir, M., Rao, N. I., 2016. Camera calibration using human face as a 2D planar object. *2016 2nd International Conference on Robotics and Artificial Intelligence (ICRAI)*, 124-129. doi.org/10.1109/ICRAI.2016.7791240.

Roman-Rivera, L.-R., Sotelo-Rodríguez, I., Pedraza-Ortega, J.C., Aceves-Fernandez, M.A., Ramos-Arreguín, J.M., Gorrostieta-Hurtado, E., 2022. Reduced Calibration Strategy Using a Basketball for RGB-D Cameras. *Mathematics*, 10, 2085. doi.org/10.3390/math10122085.

Song, L., Wu, W., Guo, J., Li, X., 2013. Survey on camera calibration technique. *2013 5th International conference on intelligent human-machine systems and cybernetics*, 2, 389-392. doi.org/10.1109/IHMSC.2013.240.

Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 1330–1334. doi.org/10.1109/34.888718.