INFERRING CAMERA INTRINSICS BASED ON SURFACES OF REVOLUTION: A SINGLE IMAGE GEOMETRIC NETWORK APPROACH FOR CAMERA CALIBRATION

Christopher Walker, Yuxing Wang*, Yawen Lu*, Guoyu Lu*

Intelligent Vision and Sensing Lab, Rochester Institute of Technology, NY, USA

ABSTRACT

Camera calibration is a necessary prerequisite in many applications of robotics, especially in robot vision in order to obtain metric reconstruction from a 2D image. In this paper, we address the problem of calibrating from a single image of a surface of revolution (SOR) based on deep learning, in order to determine the camera intrinsic parameters. Geometric constraints based on the symmetry properties of the SOR structure are deployed to our proposed learning-based camera calibration framework. To enable the calibration from a single view, we also propose a learning-based conics detection model fitting the geometric primitive of a cylinder. The calibration from a single view can be completed by minimizing the geometric constraints of two conics detected by the learning-based model with cylinder images as input. Objects with a surface of revolution are commonly visible in daily life, such as cans, bottles, and bowls, making this research both significant and practical. Finally, traditional calibration techniques are compared against our single image calibration. Experiments conducted on newly generated dataset demonstrate the effectiveness and robustness of the proposed method.

Index Terms— Camera calibration, Surfaces of Revolution, Conic Detection

1. INTRODUCTION

Camera calibration is a crucial step in robotics for motion estimation and 3D metric reconstruction, which can determine the intrinsic parameters of cameras. The process is fundamental to many applications, such as scene reconstruction, depth estimation, autonomous driving, robotic navigation and self-localization. Accurate camera parameter estimation is the fundamental of extensive robotics tasks, which allows systems to overcome perspective distortion, lens distortion [1] due to fisheye lenses, telecentric lenses [2], or short focal length lenses, and to provide 3D scene reconstruction in real world measurement units [3]. Mainstream camera calibration can be roughly divided into two categories: photogrammetric calibration [4][5][6][7] and self-calibration [8][9][10]. In addition, camera calibration methods based on image geometric properties are proposed, such as vanishing point-based methods [11][12] and pure rotation-based methods [13][14].

In extensive scenarios, camera calibration targets (e.g., checkerboard) are not provided or not easily available. Despite remarkable progress in previous research, most methods are based on two views or multiple views to accomplish the process of the camera calibration. Therefore, it is necessary to propose a calibration method that can automatically calibrate the camera and obtain the camera intrinsics from a single image. In this paper, we present a unique learning-based approach to calibrating a camera from a single image using the symmetry properties of the Surfaces of Revolution (SOR) [15][16] [17] so that an easier camera calibration method is realized using single image existing conics geometry. The primary benefit of the SOR calibration is that it can be done from a single image using common SOR objects. We explore the geometric properties of the SOR to establish the neural network constraints that enable the trained network to directly estimate the camera intrinsics just based on one image. In order to utilize camera calibration using the calibration network, two conic shapes are required at the beginning of camera calibration, for which we develop a conics detection network to extract conics from the input image. The entire camera calibration framework explores geometry constraints of the SOR relying on the conic detection to fit the intrinsic parameter matrix.

To summarize, the main contributions of this work are as follows: 1) a deep camera calibration network to explore the camera intrinsic parameters; 2) a conic detection network is developed to extract meaningful geometric curves and shapes; 3) surfaces of revolution from a single view are explored in the calibration network, which, to the best of our knowledge, is the first deep calibration neural network relying on SORs; 4) we create a synthetic cylinder dataset using the Blender toolbox for the calibration network training and testing. The camera calibration framework is shown in Fig. 1.

2. SINGLE IMAGE CAMERA CALIBRATION SCHEME

2.1. Conic Detection Network

Current deep learning methods have been used to detect ellipses, lines, and other edge curves [18]. In the case of ellipse detection, there exists few stable, robust, and efficient methods to accurately detect elliptical edges and useful mathematical parameters. In regards to real-world images, occluded object edges and background noise make ellipse detection even more challenging. To overcome the challenge

*Equally contribution

978-1-6654-0540-9/22/$31.00 ©2022 IEEE

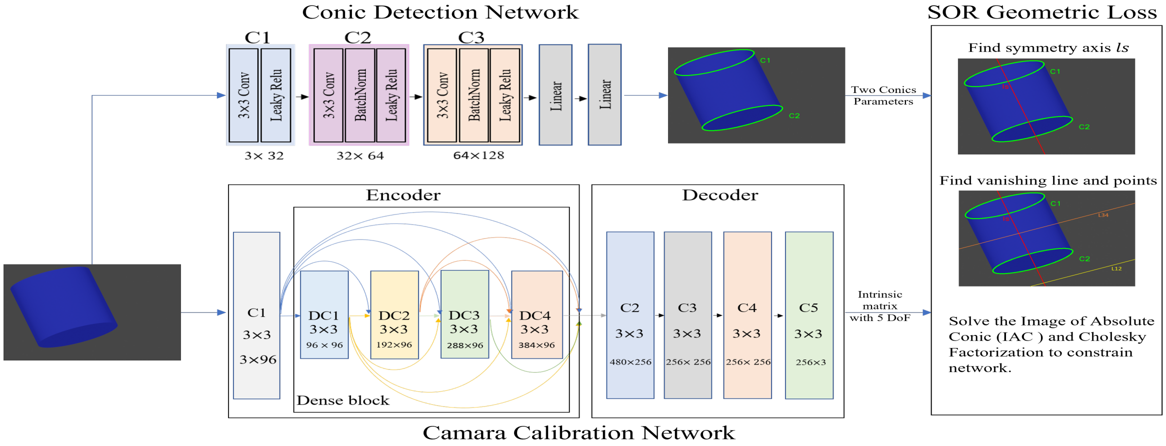

Fig. 1. Overview of the proposed camera calibration from Surfaces of Revolution (SOR) pipeline. For training, we apply an Encoder-Decoder network to minimize the geometric constraints from the SOR. A single image with conic information is input into the Encoder-Decoder network, which outputs a camera intrinsic matrix with 5 degrees of freedom (DoF). The intrinsic matrix is optimized with geometric constraints based on the symmetry properties of the SOR structure. The axis of rotation ls in the SOR and vanishing line and points will be calculated directly from the detected conics to establish the neural network constraints. During testing, only one image with conic shape is sufficient for our model to estimate the instrinsic matrix.

of occlusion, CNN-based ellipse detectors are being actively researched [18] to infer the geometric parameters of multiple elliptical objects in images. Based on the theory, a representation for an ellipse is 3 ×3 conic matrix with 6 degrees of freedom as shown:

$$Q\left(x,y\right) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \quad (1)$$

$$A_Q = \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix} \quad (2)$$

where the conic section satisfies a 2-degree polynomial equation which defines an ellipse. However, another representation for ellipses is to define an ellipse by its center, the lengths of its minor and major axes, and the angle of the major axis relative to the image coordinate frame. The ellipse parameters can be formulated as follows:

$$ellipse = \left(\left(x_{ctr}, y_{ctr}\right), \left(a, b\right), \theta\right) \quad (3)$$

where x_{ctr} and y_{ctr} represent the center coordinates of ellipse, a and b separately stand for major and minor axes and θ is the degree of rotation. Although the above two formulas represent different ellipse expressions, they can be converted to each other based on the following formulas:

$$\begin{aligned} &\frac{\left[\left(x - x_{ctr}\right) cos\theta + \left(y - y_{ctr}\right) sin\theta\right]^2}{a^2} \\ &+ \frac{\left[\left(x - x_{ctr}\right) sin\theta + \left(y - y_{ctr}\right) cos\theta\right]^2}{b^2} = 1 \end{aligned} \quad (4)$$

Equation 4 can be simplified into Equation 1 to obtain 6 coefficients. Similarly, the 5 parameters of the ellipse can also be obtained based on the 6 coefficients. Therefore, the above conic detection network learns 10 values representing the two ellipses of each cylinder image. In addition, this detection network introduces dropout and normalized labels by subtracting the mean and dividing by the standard deviation of all labels in the training set with L2 norm which is more smooth and accurate when outliers are present.

2.2. SOR Geometrical Constraints

2.2.1. Camera Elements For Calibration

The camera intrinsic parameters relate the image coordinates to the camera sensor and include the focal length f, skew s, 2D image scale factors m_x and m_y, and the principle point of the camera in the image (x_0, y_0). In the pinhole camera model, a 3D point $M = [X, Y, Z]^T$ in the world-coordinate-system (WCS) is projected onto an image from the perspective of the camera. An image point $m = [u, v]^T$ is expressed as $M' = \left[X', Y', Z'\right]^T$ in the camera coordinate system (CCS) whose origin is $C = [0, 0, 0]^T$. It is convenient at this point to express the points in homogeneous coordinates since the 3D point M is related to m up to an ambiguous scale factor. That is, a pencil of points along the ray from C to M are all the same point m in the image plane or in homogeneous coordinates $m = [u, v, 1]^T$, $M = [X, Y, Z, 1]^T$ and $M' = \left[X', Y', Z', 1\right]^T$. The relationship between M and m is by the transformation $sm = PM$, where s is an arbitrary scale factor and P is the camera transformation matrix. P consists of a hierarchy of coordinate transformations from the WCS to the CCS to the image plane and can be further broken down as $P = K\left[R|t\right]$, where K is the camera intrinsic matrix and $[R|t]$ is the camera extrinsic matrix.

The camera intrinsic matrix K relates the point in the camera coordinate system to the point on the image plane: $m = [u, v, 1]^T = K\left[X', Y', Z', 1\right]^T$. This transformation relies only on the parameters of the camera system and not on its location. The image coordinates and camera focal plane coordinates are related by a translation T and possibly a scaling S, and the focal plane coordinates relate to the camera coordinate system by a perspective transformation that depends solely on the camera's focal length f, which re-scales the image coordinates into pixels. Below K is expressed in terms of homogeneous coordinates:

$$\mathrm{K} = \underbrace{\begin{bmatrix} m_x & s & x_0 \\ 0 & m_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{T} = \begin{bmatrix} fm_x & fs & x_0 & 0 \\ 0 & fm_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5}$$

where fm_x is the horizontal scale factor, fm_y the vertical scale factor, fs the skew and (x_0, y_0) is the principal axis coordinate in pixels in the image plane. This provides 5 degrees of freedom (dof) for the camera intrinsics matrix. Often, assumptions are made to reduce the number of unknowns such as in traffic surveillance [19], where the object distance is far away and skew is assumed to be zero $(s = 0)$ and unit aspect ratios $m_x = m_y = 1$, reducing the matrix K to 3 dof; namely, f and the principal point (x_0, y_0).

2.3. Camera Calibration from Surfaces of Revolution

In the case of SOR, a homology is defined by a line of fixed points (the homology axis) and a fixed point (the vertex) that is not on the axis. In matrix representation, an homology is a 3x3 matrix W transforming points as $x_1 = Wx_3$. The transformation is defined below:

$$W = I + (\mu - 1)\frac{vl^T}{v^T l} \tag{6}$$

where I is the identity matrix, l is the axis, v is the vertex, and μ is a ratio of the homology's one distinct eigenvalue to its repeated one. Generally, the W matrix can be obtained by three points correspondences from two conics in an image of surfaces of revolution. However, when $\mu = -1$, the dofs are reduced from original 5 to 4. In addition, there is a corresponding homology called a harmonic.

Fig. 2 demonstrates two basic geometric properties of an imaged SOR. The four points x_1, x_2, x_3 and x_4 in Fig. 2 are solved by calculating the roots of the polynomial coefficients solved by the known conic. These four points represent two pairs of complex conjugate points, respectively x_1 and x_2, x_3 and x_4. The two pairs of complex conjugate points can separately calculate two lines l_{12} and l_{34} and the vanishing point can be determined through the cross product of these two lines. For determining the correct vanishing lines from the 4 point solutions, according to [16], if there is one visible conic from the image of the SOR, the vanishing line on the side of the visible conic will be chosen, which means if the bottom ellipse is visible, the bottom vanishing line will be chosen and vice versa. But if the top and bottom conics all are not visible, the vanishing line V_∞ between the two conics will be chosen. Therefore, after confirming the vanishing line l_∞, the complex conjugate points i and j corresponding to the vanishing line l_∞ are also determined. These points are so-called circular points because every conic intersects I_∞ at these points, and because these points lie on the image of the absolute conic (IAC), ω, their relationship can be formulated as follows:

$$\begin{aligned} i^T\omega i &= 0 \\ j^T\omega j &= 0 \end{aligned} \tag{7}$$

The circular points i and j are fixed entities that define orthogonal directions in Euclidean geometry, Other important fixed entities are the SOR axis of symmetry l_s, the vanishing points of the imaged cross-sections V_∞, and the vanishing line at infinity l_∞. These fixed entities can all be derived from visible segments of two imaged cross-sections.

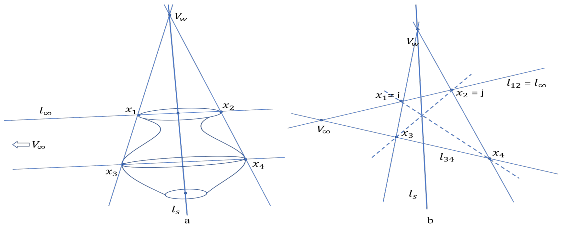

Fig. 2. The basic properties of the SOR. For the first property, pairs of corresponding points in the vertical direction (e.g. x_1 and x_3) satisfy W transformation and all lines generated by pairs of corresponding points meet at $V_w \in I_s$. For the second property, pairs of corresponding points in horizontal direction (e.g. x_1 and x_2) correspond under H and all lines generated by corresponding points in horizontal direction meet at $V_\infty \in I_\infty$. Fig. a is in the space view and Fig. b is a cross section in the vertical direction.

The imaged fixed entities discussed above can be used to express orthogonality of SORs in the scene by use of the (IAC), ω, where $\omega = K^{-T}K^{-1}$. The system of equations to solve are shown below:

$$\begin{cases} i^T\omega i = 0 \\ j^T\omega j = 0 \\ I_s = \omega V_\infty \end{cases} \tag{8}$$

This system provides four linear constraints on the IAC, but it has only three independent linear constraints (3 dofs), which is sufficient to calibrate a camera from a single image. The third constraint can be rewritten as $I_s \times \omega V_\infty = 0$ and transformed into a homogenous system and solved using singular value decomposition (SVD). In addition, other constraints can be applied such as zero skew and square pixels.

3. EXPERIMENT RESULTS

3.1. Dataset and setup

The cylinder datasets were generated by simulating a camera view at varying focal lengths and distance from a 3D cylin-

	Metrics	AAMED	Lu et al.	Our method
Cylinder dataset	Precision	0.331	0.581	**0.911**
	Recall	0.528	0.717	**0.867**
	F-score	0.406	0.642	**0.888**

Table 1. Conic detection results on our dataset compared with other state-of-the-art conic detection methods [20] [21].

der using the Blender software. The cylinder's position and rotation were varied in the virtual camera field-of-view. As a result, the top and bottom of the cylinder become 2D ellipse projections in the image. The ground truth focal length of the camera, the cylinder's elliptical projections, the corresponding elliptical points and 'hidden' label were recorded along with each generated image, where 'hidden' is a boolean value and True indicates that the camera cannot actually 'see' either complete ellipse, and therefore some of the ellipse points are 'hidden' by the cylinder object blocking the camera view.

During training, we first pre-train the conic detection module on our synthetic dataset with 2500 images which are resized to 120 × 160 together with adjusting the size of elliptical points and ellipse labels. We train our calibration model on 2500 synthetic images with the same 120 × 160 size on an NVIDIA Tesla P40 GPU with 24GB GPU memory. The configuration environment of our model is in PyTorch 1.7.1 with CUDA 10.1. For optimization, we apply Adam optimizer, where $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The learning rate is set to be $1e - 4$ without linear reduction.

3.2. Comparison with State-of-the-art Methods

Figure 3 compares the proposed method and other recent conic detection methods [20] [21] on the synthetic cylinder dataset. It can be observed that compared with the matrix-based [20] and line segments based-detection methods [21], our method is capable of estimating both conics on the upper surface and lower surface for all given samples, while [20] and [21] can only recover one conic shape which has a relatively large inclination. Our estimation results (marked in red) are rather close to the ground truth labels (marked in green), which demonstrates the superior accuracy and stability of our method.

Table 1 compares the precision, recall, and F-score of our proposed method with recent state-of-the-art methods [20] [21]. It can be observed that on the synthetic dataset, the proposed method can improve precision by 59.8% and 35.2%, respectively, for [20] and [21]. It is worth noting that although precision is the most significantly improved, recall and F-score are also largely enhanced. There are currently no existing methods that use at most two conics detection to achieve SOR-based camera calibration. At the same time, it is difficult for the existing conic detection methods to achieve at least two parallel conics detection to satisfy requirement of the SOR-based camera calibration. Therefore, we have not quantitatively compared with other methods. Instead, in Table 2, we depict the calibration results of our proposed network under both metrics using error distance RMSE and MRE. Our method performs 21.18% and 12.05% on the dataset.

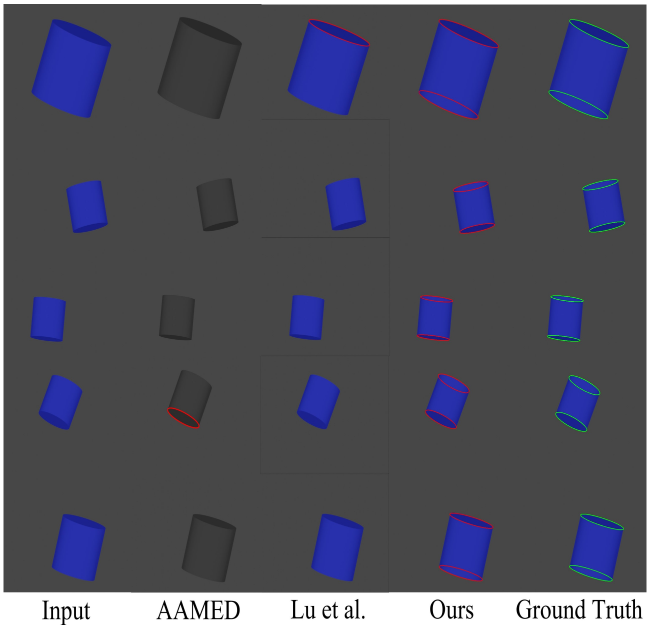

Fig. 3. Qualitative comparison of our method and other conic detection methods on the cylinder dataset.

	Cylinder dataset
RMSE	21.18
MRE	12.05

Table 2. Calibration results of focal length (RMSE and MRE) on the synthetic datasets.

4. CONCLUSION

In this paper, we investigated a camera calibration network using surfaces of revolution to realize an accurate calibration. We first presented a method for detecting the geometric primitive of a cylinder in camera images. We evaluated the accuracy of the model both qualitatively and quantitatively. The detection of the cylinder ellipses showed significantly improved results than extracting only single ellipses. Our designed calibration network, together with proposed geometric constraints based on the symmetry properties of the surfaces of revolution, our calibration framework is able to accurately learn the intrinsic parameter of the camera. To the best of our knowledge, this is the first attempt to design a deep neural network for camera calibration from the surfaces of revolution. Experiments show accurate calibration performance and an easier calibration method using only a single image.

5. ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Award No. 2126643.

6. REFERENCES

[1] C Cattaneo, G Mainetti, and R Sala, "The importance of camera calibration and distortion correction to obtain measurements with video surveillance systems," in *Journal of Physics: Conference Series*. IOP Publishing, 2015, vol. 658, p. 012009. 2

[2] Linshen Yao and Haibo Liu, "A flexible calibration approach for cameras with double-sided telecentric lenses," *International Journal of Advanced Robotic Systems*, vol. 13, no. 3, pp. 82, 2016. 2

[3] Ashutosh Saxena, Sung H Chung, and Andrew Y Ng, "3-d depth reconstruction from a single still image," *International journal of computer vision*, vol. 76, no. 1, pp. 53–69, 2008. 2

[4] Wolfgang Faig et al., "Calibration of close-range photogrammetric systems: Mathematical formulation.," 1975. 2

[5] Olivier Faugeras and OLIVIER AUTOR FAUGERAS, *Three-dimensional computer vision: a geometric viewpoint*, MIT press, 1993. 2

[6] Roger Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987. 2

[7] Zhengyou Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. 2

[8] Stephen J Maybank and Olivier D Faugeras, "A theory of self-calibration of a moving camera," *International journal of computer vision*, vol. 8, no. 2, pp. 123–151, 1992. 2

[9] Olivier D Faugeras, Q-T Luong, and Stephen J Maybank, "Camera self-calibration: Theory and experiments," in *European conference on computer vision*. Springer, 1992, pp. 321–334. 2

[10] Sylvain Bougnoux, "From projective to euclidean space under any practical situation, a criticism of self-calibration," in *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998, pp. 790–796. 2

[11] Bruno Caprile and Vincent Torre, "Using vanishing points for camera calibration," *International journal of computer vision*, vol. 4, no. 2, pp. 127–139, 1990. 2

[12] Erwan Guillou, Daniel Meneveaux, Eric Maisel, and Kadi Bouatouch, "Using vanishing points for camera calibration and coarse 3d reconstruction from a single image," *The Visual Computer*, vol. 16, no. 7, pp. 396–410, 2000. 2

[13] Richard I Hartley, "Self-calibration from multiple views with a rotating camera," in *European Conference on Computer Vision*. Springer, 1994, pp. 471–478. 2

[14] Gideon P Stein, "Accurate internal camera calibration using rotation, with analysis of sources of error," in *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 1995, pp. 230–236. 2

[15] Kwan-Yee K Wong, Paulo RS Mendonça, and Roberto Cipolla, "Reconstruction of surfaces of revolution from single uncalibrated views," *Image and Vision Computing*, vol. 22, no. 10, pp. 829–836, 2004. 2

[16] Federico Pernici, *Two Results in Computer Vision using Projective Geometry*, Ph.D. thesis, PhD thesis, University of Florence, Florence, Italy, 2005. 2, 4

[17] Carlo Colombo, Alberto Del Bimbo, and Federico Pernici, "Metric 3d reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 1, pp. 99–114, 2005. 2

[18] Wenbo Dong, Pravakar Roy, Cheng Peng, and Volkan Isler, "Ellipse r-cnn: Learning to infer elliptical object from clustering and occlusion," *IEEE Transactions on Image Processing*, vol. 30, pp. 2193–2206, 2021. 2, 3

[19] Zhaoxiang Zhang, Min Li, Kaigi Huang, and Tieniu Tan, "Practical camera auto-calibration based on object appearance and motion for traffic scene visual surveillance," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8. 4

[20] Cai Meng, Zhaoxi Li, Xiangzhi Bai, and Fugen Zhou, "Arc adjacency matrix-based fast ellipse detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 4406–4420, 2020. 5

[21] Changsheng Lu, Siyu Xia, Ming Shao, and Yun Fu, "Arc-support line segments revisited: An efficient high-quality ellipse detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 768–781, 2020. 5