2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Linear Auto-calibration of Pan-Tilt-Zoom Cameras With Rotation Center Offset

Yu Liu[1] and Hui Zhang[2]

***Abstract*— This paper addresses the linear auto-calibration problem of a pan-tilt-zoom (PTZ) camera. Unlike existing methods, we take full advantage of the offset of the camera center from the rotation center, which is usually non-negligible in bullet-type PTZ cameras. Without any prior assumption, we propose a linear method to recover all intrinsic parameters. First, we successively acquired at least four images using the zoom and rotation capabilities of the PTZ camera. Second, using the homography of two images at the same location but different scales, the principal point and zoom scalar can be linearly recovered. Finally, based on the unknown offset of the camera center and rotation center, we propose a linear method to solve the scale factor in the Kruppa equation and recover the remaining camera intrinsic parameters, namely focal lengths and skew. Synthetic and real experiments demonstrate the feasibility of our approach.**

I. INTRODUCTION

Robotic pan-tilt-zoom (PTZ) cameras are now ubiquitous and have spawned many applications such as security surveillance [1], traffic tracking [2], [3], [4], broadcast systems[5], [6], [7], distance learning, etc. Due to the frequent rotation and zooming of such cameras, the intrinsic parameters of the camera will change [1], affecting the performance of their applications. Although high-quality PTZ cameras can keep their intrinsic parameters unchanged, most PTZ cameras currently used in common scenarios are cheap and cannot meet such standards. In addition, since PTZ cameras are generally fixed in the scene, which makes the classical offline calibration methods with calibration objects no longer applicable, e.g., [8], [9], [10], [11], [12], [13]. Meanwhile, for bullet-type PTZ cameras, the camera center is not aligned with the rotation center [5], [1], which makes it no longer suitable for approximating its motion to pure rotation [14], [15]. Therefore, it is necessary to propose a real-time and accurate calibration method for common PTZ cameras.

Most of the bullet PTZ cameras on the market have an offset $\mathbf{t}$ between the camera projection center and the rotation center, which is the intersection of the pan and tilt axis [5], [16] (see Fig. 1). Traditionally, since the offset $\mathbf{t}$ is small

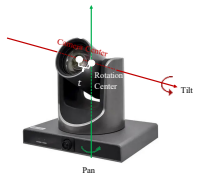

Fig. 1. **The bullet PTZ camera.** There is an offset $\mathbf{t}$ between the camera projection center and the rotation center, which is the intersection of the pan and tilt axes.

relative to the scene scale, the motion of a PTZ camera can be approximated as a rotation around its projection center [1], [2], [17], [18]. The camera intrinsics can then be calibrated by methods based on pure rotation [14], [19]. Some other existing PTZ camera calibration methods are based on restricted scenarios or require prior information. For example, the methods in [2], [20] require the existence of parallel lines in the scene, and then use the vanishing point obtained by their intersection to calibrate the camera. The methods in [21], [22] need to know the rotation angle. The method in [23] estimates the focal length from a single control point with known 3D coordinates. Additionally, many professional sports broadcasters have deployed visual tracking systems using PTZ cameras, such as StatsPerform and SportsLogiq. The calibration problem in such systems is equivalent to the homography estimation problem [5], [6], [7], so the intrinsic parameters are calibrated using known 3D points or plannar shapes on the playground. Other research methods [24], [1] exploit the zoom capabilities of PTZ cameras and propose calibrating cameras based on incremental zoom image sequences. However, these methods can only calibrate the focal length and involve many nonlinear steps to remove noise, making the calibration process time-consuming [24]. Wu [1] performs dynamic calibration by introducing a feature library containing at least 10 images, but the optimization process is inflexible. While the GA-PSO method [25] can calibrate the camera automatically, it has never been used for PTZ cameras and cannot obtain the zoom scalar.

Different from the above methods, we consider the offset $\mathbf{t}$ in the projection matrix and propose a novel two-step auto-calibration method for PTZ cameras. The method captures at least four images with the general operation of a PTZ camera, from which the zoom scalar and five camera intrinsics can be linearly solved simultaneously. First, we propose that the zoom scalar and the principal point can be obtained from the homography of image pairs with different zoom

*This work was supported by National Key R&D Program of China (2022YFE0201400), the National Natural Science Foundation of China (62076029), Guangdong Science and Technology Department (2022B1212010006, 2017A030313362), Guangdong Key Lab of AI and Multi-modal Data Processing (2020KSYS007), and internal funds of the United International College (R202012, R201802, R5201904, UICR0400025-21).

[1]Yu Liu is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, China, and also with the Department of Computer Science, United International College, BNU-HKBU, Zhuhai, China. yuliu@uic.edu.cn

[2]Hui Zhang (the corresponding author) is with the Department of Computer Science, United International College, Beijing Normal University-Hong Kong Baptist University, Zhuhai, China. amyzhang@uic.edu.cn

scalars under the same rigid transformation. Second, based on the obtained principal point, we propose a linear method to determine the scale parameter in the Kruppa equation under the same camera through the constraints between the homography and the fundamental matrix, and then further recover the remaining three intrinsic parameters. Synthetic and real experiments demonstrate the feasibility and accuracy of our method. We summarize our contributions as follows:

1) By considering the offset between the camera center and the rotation center, a novel two-step auto-calibration method for PTZ cameras is proposed;
2) Based on the constraints of the homography and the fundamental matrix, we propose a linear method to solve the scale parameter in the Kruppa equation;
3) The auto-calibration method considering the offest is more accurate than calibrating the bullet PTZ camera with pure rotation, and can achieve similar performance to the method with the calibration grid.

This paper is organized as follows. Section II introduces the basic concepts used in this paper. Section III describes the PTZ camera system used and proposes a method to linearly restore the zoom scalar and the principal point. Section IV presents a calibration method based on solving the scale parameter in the Kruppa equation. Synthetic and real experiments in Section V verify the feasibility and accuracy of our method. Section VI gives the conclusion.

II. Preliminary

This section briefly introduces the theoretical background of the camera model and the camera calibration of the pan-tilt and zoom cameras, respectively.

A. Epipolar geometry of a pan-tilt (PT) camera

Figure 2 shows a PT camera $\mathbf{c}_0$ rotated around the rotation center $\mathbf{O}_w$ to $\mathbf{c}_1$. $\mathbf{c}_0$ and $\mathbf{c}_1$ has projection matrices $\mathbf{P}_0$ and $\mathbf{P}_1$

$$\mathbf{P}_0 = \mathbf{K}[\mathbf{I}|\mathbf{t}] \quad and \quad \mathbf{P}_1 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}], \tag{1}$$

where $\mathbf{R}_1$ is the rotation matrices of $\mathbf{c}_1$, $\mathbf{t}$ is the non-zero offset vector, and $\mathbf{K}$ is the camera intrinsic matrix.

The epipolar geometry between these two views is described by the fundamental matrix $\mathbf{F}$. The baseline joining the two camera centers $\mathbf{c}_0$ and $\mathbf{c}_1$ intersects the two images I_0 and I_1 at the epipoles $\mathbf{e}$ and $\mathbf{e}'$, respectively [26]. $\mathbf{e}$ and $\mathbf{e}'$ are the image of $\mathbf{c}_1$ and $\mathbf{c}_0$ in I_0 and I_1, which can be obtained as the null space of $\mathbf{F}$ and $\mathbf{F}^T$, respectively. The absolute conic (AC) $\mathbf{\Omega}_\infty$ on the infinity plane $\boldsymbol{\pi}_\infty$ is projected into I_0 and I_1 as the image of absolute conic (IAC) $\boldsymbol{\omega}$ [26]. The plane defined by the two camera centers and any 3D point $\mathbf{X}$ on $\mathbf{\Omega}_\infty$ is known as the epipolar plane [26]. It intersects images I_0 and I_1 at a pair of corresponding epipolar lines $\mathbf{l}$ and $\mathbf{l}'$ [26]. $\mathbf{l}$ and $\mathbf{l}'$ are tangent to $\boldsymbol{\omega}$ at a pair of corresponding points $\mathbf{x}_0$ and $\mathbf{x}_1$, which are the projections of $\mathbf{X}$ in the two images, respectively.

The Kruppa equation is an algebraic representation of these correspondences and can be expressed as [26]

$$[\mathbf{e}']_\times \boldsymbol{\omega}^* [\mathbf{e}']_\times = \lambda \mathbf{F} \boldsymbol{\omega}^* \mathbf{F}^T, \tag{2}$$

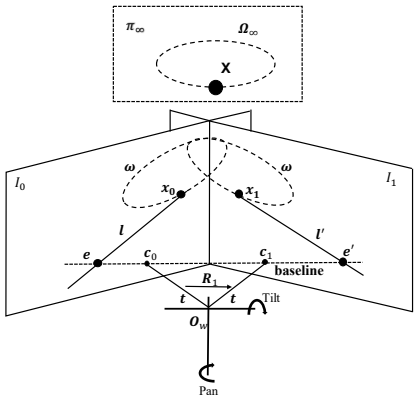

Fig. 2. **The epipolar geometry for a PT camera.** A PT camera $\mathbf{c}_0$ rotated around the rotation center $\mathbf{O}_w$ to $\mathbf{c}_1$. The projection matrices are $\mathbf{P}_0$ and $\mathbf{P}_1$, respectively. $\mathbf{R}_1$ is the rotation of $\mathbf{c}_1$, the non-zero $\mathbf{t}$ is the offset of the camera center to the rotation center. The baseline joining $\mathbf{c}_0$ and $\mathbf{c}_1$ intersects the two image I_0 and I_1 at the epipole $\mathbf{e}$ and $\mathbf{e}'$, respectively. Under the perspective projection, the absolute conic $\mathbf{\Omega}_\infty$ is projected as the image of the absolute conic (IAC) $\boldsymbol{\omega}$ in two views. And the corresponding epipolar lines $\mathbf{l}$ and $\mathbf{l}'$ are tangent to the IAC $\boldsymbol{\omega}$ at $\mathbf{x}_0$ and $\mathbf{x}_1$, which are the images of the 3D point $\mathbf{X}$ of AC.

where λ is a non-zero scale parameter, $[\mathbf{e}']_\times$ is the skew-symmetric matrix of the epipole $\mathbf{e}'$, and $\boldsymbol{\omega}^* = \mathbf{K}\mathbf{K}^\mathbf{T}$ is the dual image of the absolute conic (DIAC).

The fundamental matrix $\mathbf{F}$ in (2) can be easily obtained by the classical DLT normalized 8-points algorithm [26] or the 5-points method [27]. The epipole $\mathbf{e}'$ can be found as the *null-vector* of $\mathbf{F}^T$ [26]. Then once λ is determined, the camera intrinsic matrix $\mathbf{K}$ can be obtained by the Cholesky decomposition of $\boldsymbol{\omega}$ [28].

B. Pinhole camera and its zoom model

The intrinsic matrix $\mathbf{K}$ of a pinhole camera has the form [26]

$$\mathbf{K} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where f_x and f_y are the focal lengths on x- and y-axis, respectively. s is the skew, and (u_0, v_0) is the principal point.

Without loss of generality, for a pinhole camera with a zoom lens, the relationship between the focal lengths and the zoom scalar is a first-order approximation [26], [1], [29]. Therefore, the camera intrinsic matrix $\mathbf{K}_z$ of a pinhole PTZ camera with the zoom scalar $z \neq 1$ is

$$\mathbf{K}_z = \mathbf{K} diag\,(z, z, 1)\,. \tag{4}$$

III. Recovery of the Zoom Scalar and the Principal Point

Figure 3 shows the images captured in the process of PTZ camera calibration. A 3D point $\mathbf{X}$ is projected on the points $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ in four image I_0, I_1, I_2, I_3 through the projection matrices,

$$\begin{aligned} &\mathbf{P}_0 = \mathbf{K}[\mathbf{I}|\mathbf{t}], \mathbf{P}_1 = \mathbf{K}[\mathbf{R}_1|\mathbf{t}], \\ &\mathbf{P}_2 = \mathbf{K}[\mathbf{R}_2|\mathbf{t}], \mathbf{P}_3 = \mathbf{K}_z[\mathbf{R}_2|\mathbf{t}]. \end{aligned} \tag{5}$$

These indicates that we can first capture the image I_0. Using the rotation operation of the PTZ camera, the image I_1 and I_2 are then captured by different rotation matrices $\mathbf{R}_1$ and

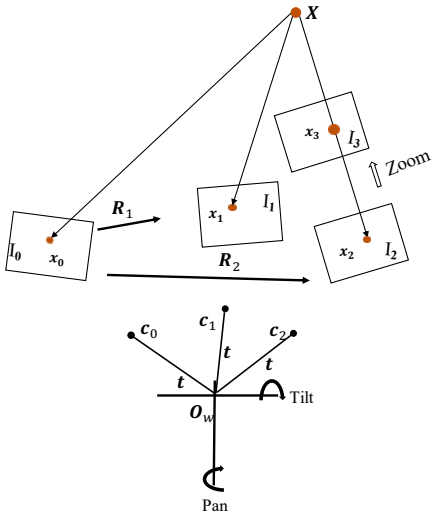

Fig. 3. **The images captured in the process of PTZ camera calibration.** A 3D point $\mathbf{X}$ is projected on the points $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ in four image I_0, I_1, I_2, I_3 through the projection matrices $\mathbf{P}_0$, $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, respectively. First, we can caputure the image I_0. Using the rotation operation of the PTZ camera, the image I_1 and I_2 are captured by different rotation matrices $\mathbf{R}_1$ and $\mathbf{R}_2$. The image I_3 is then captured just by zooming the focal length without any rotation.

$\mathbf{R}_2$. The image I_3 is finally captured just by zooming the focal length without any rotation. Due to the offset $\mathbf{t} \neq \mathbf{0}$, the epipole $\mathbf{e}' = \mathbf{P}_1 \left[-\mathbf{I}^T\mathbf{t}; 1\right] = \mathbf{K}\left[\mathbf{I} - \mathbf{R}_1\right]\mathbf{t}$ between $\mathbf{P}_0$ and $\mathbf{P}_1$ is a non-zero vector on the image I_1. Therefore, the fundamental matrix $\mathbf{F}$ between I_0 and I_1 exists and can be obtained by $\mathbf{F} = [\mathbf{e}']_\times \mathbf{H}$, where $\mathbf{H}$ is the homography induced by any world plane between these two images [26]. Moreover, the fundamental matrices between the images, I_0 and I_2, I_1 and I_2, also satisfy the relevant constraints.

For the image I_3 obtained by zooming in the focal length under the same rigid body transformation of the image I_2, the relative translation between I_2 and I_3 is $\mathbf{0}$, so only the homography can describe the correspondence between them [26], as

$$\begin{aligned}\mathbf{H}_z = \mathbf{K}_z\mathbf{K}^{-1} &= \mathbf{K} diag\left(z, z, 1\right)\mathbf{K}^{-1} \\ &= \begin{bmatrix} z & 0 & (1-z)\,u_0 \\ 0 & z & (1-z)\,v_0 \\ 0 & 0 & 1 \end{bmatrix}.\end{aligned} \tag{6}$$

It can be seen that $\mathbf{H}_z$ is an infinity homography between I_2 and I_3. Therefore, $\mathbf{H}_z$ can be obtained by the classical DLT 4-points method [26] with $\mathbf{x}_{3i} = \mathbf{H}_z\mathbf{x}_{2i}$, where $\mathbf{x}_{2i}$ and $\mathbf{x}_{3i}$ are the corresponding points on images I_2 and I_3, respectively. $i = 1, 2, ..., n$, and n is the number of the corresponding points.

Furthermore, since $\mathbf{H}_z$ is an upper triangular matrix with eigenvalues on the the diagonal elements, i.e., $\{z, z, 1\}$, the zoom scalar z can easily be obtained as the equal eigenvalues of $\mathbf{H}_z$. Back-substituting the unit eigenvalue into $\mathbf{H}_z$, we get

$$\mathbf{H}_z - \mathbf{I}_{3\times 3} = \begin{bmatrix} z-1 & 0 & (1-z)u_0 \\ 0 & z-1 & (1-z)v_0 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -u_0 \\ 0 & 1 & -v_0 \\ 0 & 0 & 0 \end{bmatrix},$$

and the corresponding eigenvector is the principal point $[u_0, v_0, 1]^T$.

IV. Full Camera Auto-calibration

In this section, knowing the principal point, we propose a linear method that uses images I_0, I_1, and I_2 to calibrate the remaining three camera intrinsic parameters. We will first recover a non-infinite homography $\mathbf{H}$, then the scale parameter λ in the Kruppa equation (2), and finally the focal length f_x, f_y and the skew s.

A. The real eigenvalue of the non-infinite homography $\boldsymbol{H}$

For two views I_0 and I_1 of a pan-tilt camera under the projection matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ in (5), given the fundamental matrix $\mathbf{F}$ and three arbitrary image point correspondences $\mathbf{x}_{0i}$, $\mathbf{x}_{1i}$ $(i = 1, 2, 3)$, any non-infinity homography $\mathbf{H}$ induced by the plane $\boldsymbol{\pi}$ of the 3D points can be obtained as ([26], Chapter 13)

$$\mathbf{H} \sim \mathbf{A} - \mathbf{e}'(\mathbf{Q}^{-1}\mathbf{b})^T, \tag{7}$$

where $\mathbf{A} = [\mathbf{e}']_\times \mathbf{F}$, and $\mathbf{b}$ is a 3-vector with components

$$b_i = \frac{(\mathbf{x}_{1i} \times (\mathbf{A}\mathbf{x}_{0i}))^T(\mathbf{x}_{1i} \times \mathbf{e}')}{(\mathbf{x}_{1i} \times \mathbf{e}')^T(\mathbf{x}_{1i} \times \mathbf{e}')},$$

and $\mathbf{Q}$ is a 3×3 matrix with rows $\mathbf{x}_{0i}^T$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{x}_{01} & \mathbf{x}_{02} & \mathbf{x}_{03} \end{bmatrix}_{3\times 3}^T.$$

In addition, $\mathbf{H}$ can also be represented with $\mathbf{H}_\infty$, i.e., the infinite homography induced by a plane at infinity $\boldsymbol{\pi}_\infty$, as

$$\begin{aligned}\mathbf{H} &\sim \mathbf{K}\left(\mathbf{R}_r - \frac{\mathbf{t}_r\mathbf{n}^T}{d}\right)\mathbf{K}^{-1} \\ &\sim \mathbf{K}\mathbf{R}_r\mathbf{K}^{-1} - \frac{1}{d}\mathbf{K}\mathbf{t}_r\mathbf{n}^T\mathbf{K}^{-1} \\ &\sim \mathbf{H}_\infty - \frac{1}{d}\mathbf{e}'\mathbf{n}^T\mathbf{K}^{-1} \\ &= \alpha\mathbf{H}_\infty + \mathbf{e}'\mathbf{y}^T,\end{aligned} \tag{8}$$

where $\mathbf{R}_r$ and $\mathbf{t}_r$ are the relative rotation and translation between the two views, $\mathbf{n}$ and d are the normal and depth of the plane $\boldsymbol{\pi}$, $\mathbf{y}$ is the imaged plane normal.

Since the homography $\mathbf{H}$ is a conjugate rotation, it has only one real eigenvalue d_1 [30], ([26] A7.1). The following proposition is proposed to obtain the coefficient α in (8).

Proposition 1: The real eigenvalue of $\mathbf{H}$ is the coefficient α in (8).

Proof: Since the infinite homography $\mathbf{H}_\infty$ has an unit eigenvalue and the corresponding eigenvector is the vanishing point of the rotation axis direction ([26], Chapter 19), we can have $\mathbf{H}_\infty\mathbf{v}_1 = \mathbf{v}_1$, where $\mathbf{v}_1$ is the real eigenvector of $\mathbf{H}$.

Since

$$\mathbf{F}^T\mathbf{H} = \mathbf{A}^T[\mathbf{e}']_\times\left(\mathbf{A} - \mathbf{e}'\left(\mathbf{Q}^{-1}\mathbf{b}\right)^T\right) = \mathbf{A}^T[\mathbf{e}']_\times\mathbf{A},$$

$\mathbf{F}^T\mathbf{H}$ is a skew-symmetric matrix, thus $\mathbf{H}$ is compatible and satisfies with [26] [31]

$$\mathbf{F}^T\mathbf{H} + \mathbf{H}^T\mathbf{F} = \mathbf{0}. \tag{9}$$

Similarly, we also have $\mathbf{F}^T\mathbf{H}_\infty + \mathbf{H}_\infty^T\mathbf{F} = \mathbf{0}$. We can then multiply the eigenvector $\mathbf{v}_1$ to both sides of (9) and get

$$\begin{aligned}\mathbf{F}^T\mathbf{H}\mathbf{v}_1 + \mathbf{H}^T\mathbf{F}\mathbf{v}_1 &= \mathbf{F}^T d_1\mathbf{v}_1 + \mathbf{H}^T\mathbf{F}\mathbf{v}_1\\ &= d_1\mathbf{F}^T\mathbf{v}_1 + \left(\alpha\mathbf{H}_\infty^T + \mathbf{y}\mathbf{e}'^T\right)\mathbf{F}\mathbf{v}_1\\ &= d_1\mathbf{F}^T\mathbf{v}_1 + \alpha\mathbf{H}_\infty^T\mathbf{F}\mathbf{v}_1 + \mathbf{0}_{3\times 1}\\ &= d_1\mathbf{F}^T\mathbf{v}_1 - \alpha\mathbf{F}^T\mathbf{H}_\infty\mathbf{v}_1\\ &= d_1\mathbf{F}^T\mathbf{v}_1 - \alpha\mathbf{F}^T\mathbf{v}_1\\ &= \mathbf{0}_{3\times 1}.\end{aligned}$$

Thus $\alpha = d_1$, which is the real eigenvalue of $\mathbf{H}$. ■

B. Recovery of coefficient in the Kruppa equation

The fundamental matrix $\mathbf{F}$ can be expressed by both the homography $\mathbf{H}$ and the infinite homography $\mathbf{H}_\infty$ [26], [32], as

$$\zeta\mathbf{F} = [\mathbf{e}']_\times\mathbf{H}, \tag{10}$$

and

$$\beta\mathbf{F} = [\mathbf{e}']_\times\mathbf{H}_\infty, \tag{11}$$

where ζ and β are two non-zero coefficients.

Since $\mathbf{H}_\infty = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}$, the DIAC $\boldsymbol{\omega}^*$ satisfies

$$\boldsymbol{\omega}^* = \mathbf{H}_\infty\boldsymbol{\omega}^*\mathbf{H}_\infty^T. \tag{12}$$

Substituting (12) into the left side of the Kruppa equation (2), we can get

$$[\mathbf{e}']_\times\mathbf{H}_\infty\boldsymbol{\omega}^*\mathbf{H}_\infty^T[\mathbf{e}']_\times = \lambda\mathbf{F}\boldsymbol{\omega}^*\mathbf{F}^T. \tag{13}$$

We substitute (11) into the left side of (13) and get

$$\beta\mathbf{F}\boldsymbol{\omega}^*\beta\mathbf{F}^T = \lambda\mathbf{F}\boldsymbol{\omega}^*\mathbf{F}^T \Rightarrow \lambda = \beta^2. \tag{14}$$

Thus, the λ can be obtained from β.

Additionally, substituting (8) into (10), we can also get

$$\begin{aligned}\zeta\mathbf{F} &= [\mathbf{e}']_\times\mathbf{H}\\ &= [\mathbf{e}']_\times\left(\alpha\mathbf{H}_\infty + \mathbf{e}'\mathbf{y}^T\right)\\ &= \alpha[\mathbf{e}']_\times\mathbf{H}_\infty + \mathbf{0}_{3\times 3}\\ &= \alpha\beta\mathbf{F}.\end{aligned}$$

Thus β can be recovered by

$$\beta = \frac{\zeta}{\alpha}, \tag{15}$$

where α is the real eigenvector of $\mathbf{H}$ (Proposition 1), ζ can be obtained by dividing the corresponding elements of $\mathbf{F}$ and $[\mathbf{e}']_\times\mathbf{H}$ from (10). Therefore, we can obtain β and then λ in the Kruppa equation using (14).

C. Recovery of the focal lengths and the skew

Since the principal point has already been recovered in Section III, the intrinsic matrix $\mathbf{K}$ can be rewritten as

$$\mathbf{K} = \begin{bmatrix}1 & 0 & u_0\\ 0 & 1 & v_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}f_x & s & 0\\ 0 & f_y & 0\\ 0 & 0 & 1\end{bmatrix},$$

so the DIAC $\boldsymbol{\omega}^*$ is

$$\begin{aligned}\boldsymbol{\omega}^* &= \mathbf{K}\mathbf{K}^T\\ &= \begin{bmatrix}1 & 0 & u_0\\ 0 & 1 & v_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}f_x^2+s^2 & sf_y & 0\\ sf_y & f_y^2 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & u_0\\ 0 & 1 & v_0\\ 0 & 0 & 1\end{bmatrix}^T\\ &= \begin{bmatrix}1 & 0 & u_0\\ 0 & 1 & v_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}w_1 & w_2 & 0\\ w_2 & w_3 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & u_0\\ 0 & 1 & v_0\\ 0 & 0 & 1\end{bmatrix}^T.\end{aligned}$$

Substituting this into (2), we can get a linear system from the image I_0 and I_1, as

$$\mathbf{M}\mathbf{w} = \mathbf{0}_{4\times 1}, \tag{16}$$

where $\mathbf{M}$ is a 4×4 coefficient matrix with $rank(\mathbf{M}) = 2$, $\mathbf{w} = [w_1, w_2, w_3, 1]^T$ is the unknown vector. Given at least an extra image I_2, we can then recover $\mathbf{w}$. Finally, the focal lengths f_x, f_y and skew s can be obtained as

$$f_y = \sqrt{w_3}, \quad s = \frac{w_2}{f_y}, \quad and \quad f_x = \sqrt{w_1 - s^2}. \tag{17}$$

In addition, for a zero-skew camera, the DIAC $\boldsymbol{\omega}^*$ has only two unknown parameters. The above linear system can be recovered from at least two views.

D. Algorithms and discussions

To summarize, the entire calibration process is listed in as follows.

Algorithm 1 PTZ Camera Auto-calibration

Input: Four images captured by a PTZ camera;

Output: Five camera intrinsic paramters and the zoom scalar;

1: The zoom scalar z and the principal point (u_0, v_0) can be firstly obtained from the eigenvectors and eigenvalues of the homography $\mathbf{H}_z$ obtained from images I_2 and I_3;
2: From the correspondence points between the image pairs I_0 and I_1, I_0 and I_2, the fundamental matrix $\mathbf{F}_i$ $(i = 1, 2)$ can be obtained and thus the epipoles;
3: The non-infinity homography $\mathbf{H}_i$ $(i = 1, 2)$ from any point triplets between images I_0 and I_1, image I_0 and I_2 can be found by (7);
4: The coefficient α_i can be obtained as the real eigenvalue of $\mathbf{H}_i$, thus the coefficient β with (15) and λ in the Kruppa equation with (14);
5: $\mathbf{w}$ in (16) can then be recovered linearly, thus the intrinsics, i.e., the focal lengths f_x, f_y and the skew s with (17).

Note that for a zero-skew camera, only one pair of views I_0 and I_1 is required to calibrate the focal lengths f_x, f_y. Further, that our method is suitable for calibrating bullet PTZ cameras. For the ball PTZ cameras, when the offset $\mathbf{t}$ is generally considered to be zero, its motion is pure rotation, so our method is not applicable.

V. Experiments

The synthetic and real experiments are carried out to evaluate the feasibility and accuracy of the proposed method.

A. *Synthetic experiments*

First, two cube grids of 3000 3D points (see Fig. 4) are projected into the images using four predefined projection matrices as (5). The camera intrinsic parameters of these four projection matrices are set as

$$\mathbf{K} = \begin{bmatrix} 1000 & 0.5 & 260 \\ 0 & 800 & 240 \\ 0 & 0 & 1 \end{bmatrix}, \tag{18}$$

and the zoom scalar z = 1.2. The *pan* and *tilt* angles are set to arbitrary values, and the *yaw* angle is 0. The offset $\mathbf{t}$ is set arbitrarily in the range $[-2, 2]$. To evaluate the robustness of our algorithm, zero-mean Gaussian noise of 0.0 to 3.0 pixels was added to each projected point with a noise step size of 0.1. For each noise level, 100 independent trials were conducted to evaluate the feasibility of the proposed method.

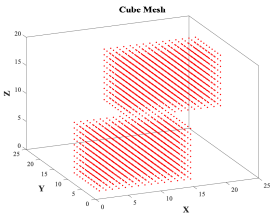

Fig. 4. The synthetic object is two cube grids of 3000 3D points.

We first use the homography between the last two images to recover the zoom scalar z and the principal point (u_0, v_0). Then we use the method in Section IV to estimate the remaining three intrinsic parameters f_x, f_y, and s. Fig. 5 shows the calibration results of the PTZ camera.

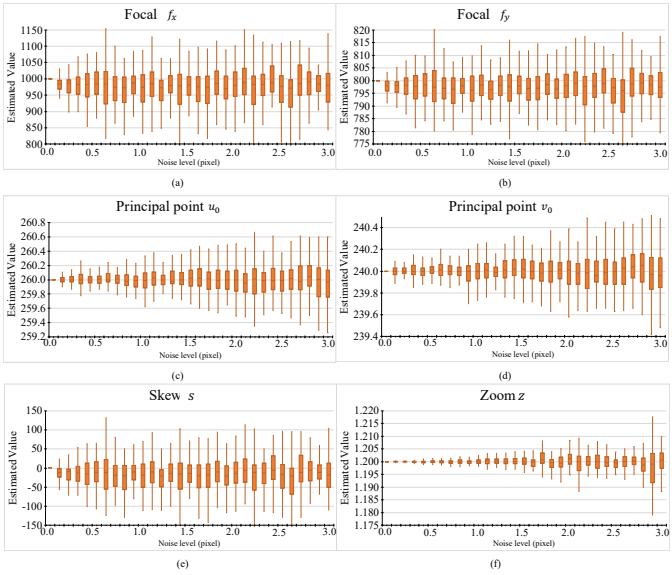

Fig. 5. Boxplots of (a &b) the estimated focal lengths, (c & d) the principal point, (e) the skew s, and (f) the zoom scalar z under different image noise levels.

It can be seen that our method is numerically stable even when the noise level reaches 3.0 pixels. The relative errors of all calibration parameters at noise level 1.0 are shown in Tab. I. Since the scale of the skew s is much smaller than that of the focal length, its direct estimation tends to be unstable. However, the error of skew angle θ in $s = -f_x \cot\theta$ [8] are less than 1.06 (1.18%) compared to the ground truth angle, 89.97°, which is also acceptable.

TABLE I

RELATIVE ERRORS OF CAMERA INTRINSIC PARAMETERS AT NOISE LEVEL 1.0.

Method	f_x	f_y	u_0	v_0	s	z
GT	1000	800	260	240	0.5	1.2
Ours (Rel.)	974.56 (2.54%)	796.41 (0.45%)	259.87 (0.05%)	239.92 (0.03%)	-18.54	1.1989 (0.092%)

In addition, for most cameras on the market, the skew $s = 0$ [26],[33],[15]. The resulting focal lengths f_x and f_y are shown in Fig.6. It can be seen that with the estimates

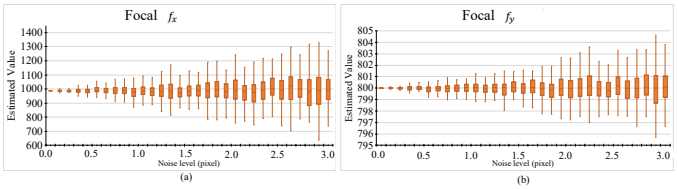

Fig. 6. Boxplots of the estimated focal lengths, (a) f_x and (b) f_y with assumption that skew s = 0 under different image noise levels.

on f_x, f_y are improved when s = 0. Their estimates and the corresponding relative errors at the noise level 1.0 are f_x = 983.20 (1.68%) and f_y = 800.05 (0.01%), both are better than the results in Tab. I, respectively.

Furthermore, to demonstrate the calibration accuracy under different external parameters, i.e., pan, tilt angles, and distance ratios (the distance from the object to the camera over the magnitute of the offset $\mathbf{t}$) (see Fig. 7), we

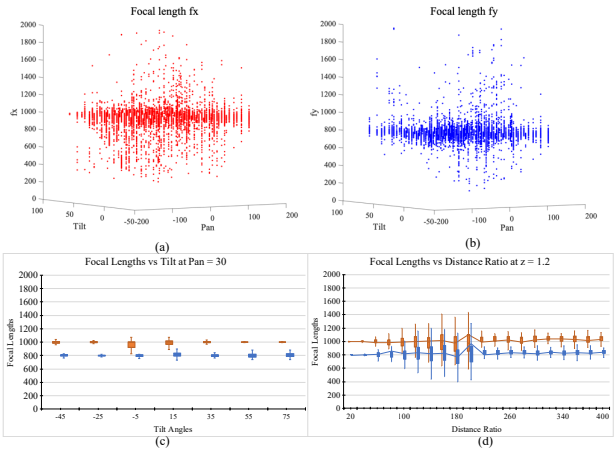

Fig. 7. Boxplots of recovered focal lengths under the zooming scalar $z = 1.2$ at the noise level of 1.0. (a) f_x, (b) f_y under different pan and tilte angels. (c) At the pan angle 30°, boxplots of focal lengths under different tilte angels. (d) Boxplots of focal lengths under different distance ratios.

performed 100 independent trials under the zooming scalar $z = 1.2$ at the noise level of 1.0. The *pan* angles are from $[-170°, 170°]$, the *tilt* angles are from $[-45°, 75°]$ (with a step size of 20°), and the distance ratio is from $[20, 400]$ (with

a step size of 20). It can be seen from Fig. 7.(a)-(b) that the recovered focal lengths are with greater errors when the the pan and tilt angle are close to zero. Specifically, in order to have accurate results, it is better to make sure that span $\in [-170°, -30°]\&[30°, 170°]$. Under the pan angle $30°$, it is better to make sure that tilt $\in [-45°, -10°]\&[15°, 75°]$. Meanwhile, we also show the recovered focal lengths under different distance ratios (Fig. 7(d)). It can be seen that when the ratio distance is larger than 200, the medians of the recovered focal lengths are more diverted from the ground truth value as 1049.07 (4.91%).

B. Real scenes

In the real experiments, we use images captured by a HIKVISION U102D PTZ camera to validate the proposed method. We performed 10 independent trials on different images and used the median value as the final results. We use Zhang's [8] results as ground truth for camera intrinsics. For the zoom scalar, we also use Zhang's [8] method to recover the focal lengths of two different scales, and then use their ratios as ground truth. We provide a comparison between the results of our methods and those of pure rotation [14], [15]. Furthermore, since our method uses the Kruppa equation for camera auto-calibration, we also provide a result comparing with GA-PSO [25], which is based on optimizing the the Kruppa equation through iteration.

Tab. II shows the calibration results of three methods. It

TABLE II

RECOVERY OF THE CAMERA INTRINSIC PARAMETERS AND ZOOM SCALAR WITH THREE METHODS.

Method	f_x	f_y	s	u_0	v_0	z
Camera with non-zero skew (from 4 images):						
Zhang [8]	1106.03	1108.37	-0.86	982.38	542.12	1.5556
Ours	**1131.89**	**1115.45**	87.66	**978.48**	549.13	1.5657
(Rel.)	**(2.34%)**	**(0.64%)**		**(0.40%)**	(1.29%)	(0.64%)
Rotation [14]	1276.88	1144.79	274.25	997.95	466.28	**1.5476**
(Rel.)	(15.45%)	(3.29%)		(1.58%)	(13.99%)	**(0.51%)**
GA-PSO [25]	1343.58	1388	**62.10**	986.93	546.31	-
(Rel.)	(21.48%)	(25.23%)		(0.46%)	(0.77%)	
Zero-skew camera (from 3 images):						
Zhang [8]	1109.95	1112.23	–	985.22	541.58	1.5632
Ours	**1119.47**	**1126.00**	–	**983.40**	**547.45**	1.5689
(Rel.)	**(0.86%)**	**(1.24%)**		**(0.18%)**	**(1.08%)**	(0.36%)
Rotation [14]	1302.28	1366.57	–	980.53	529.27	**1.5588**
(Rel.)	(17.33%)	(22.87%)		(0.47%)	(2.27%)	**(0.28%)**

can be seen that the relative error of the zoom scalar z of all methods is less than 0.65%, which is consistent with the synthetic results, indicating that our method has high accuracy. For camera intrinsics, the results of our method are more accurate than those for pure rotational motion, although the skew s is not very accurate. This is because the value of s is very small (-0.86) compared to the other values. Furthermore, the results of our method are also better than GA-PSO. That's because GA-PSO is based on genetic optimization and is greatly affected by the initial value. From Tab.II, it can be seen that when we consider $s = 0$ [26], [33], a common assumption, our results still outperform those of pure rotation. In summary, our method can accurately recover the parameters of the PTZ camera, and the results also show that the offset between the camera center and the rotation center needs to be considered for calibrating the bullet PTZ camera.

C. 3D reconstruction and re-projection errors

To verify the applicability of our method, we perform 3D reconstruction using calibrated camera intrinsics. Given the Indoor (the camera is fixed) and Toy (the camera is moving) image sets of 21 images, respectively, the camera external parameters and the 3D models are recovered by the SFM algorithm [34] (see Fig. 8). In addition, Tab. III

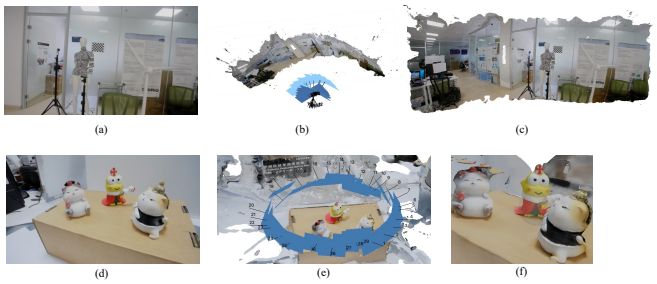

Fig. 8. **The results of 3D reconstruction.** (a & d) One view of the Indoor and Toy image sets used to calibrate the cameras. (b & e) The views with calibrated camera extrinsics are shown in blue planes. (c & f) The reconstructed 3D models.

compares the reprojection error of our method with that of Zhang's method with a calibration board [8], which is regarded as the baseline. The results show that, in practice, the reprojection error of our method is close to that of Zhang, which means that the accuracy of our auto-calibration method is comparable to that of the calibration method using the calibration plane.

TABLE III

THE REPROJECTION ERRORS OF ZHANG'S METHOD AND OURS.

Method	Toy	Indoor
Zhang [8]	0.66	0.53
Ours	0.63	0.55

VI. CONCLUSIONS

By exploiting the offset between the camera center and the rotation center, a practical technique for PTZ camera auto-calibration is proposed in this paper. The use of the offset makes the calibration process easier and does not require the use of any precisely manufactured devices such as calibration grids, thus the method can be seamlessly incorporated into PTZ camera based applications. We propose a linear method to solve the zoom scalar and principal point from homography of two views from camera zooming. We then propose a new method to linearly solve the coefficient in the Kruppa equation and further recover the remaining three camera intrinsic parameters. As demonstrated by experimental results on synthetic and real data, it is necessary to account for the offset in the projection matrix, With our method, the focal lengths are estimated with high accuracy as well as the zoom scalar, with an error less than 2.4% for realistic levels of noise. However, the proposed method cannot be applied to calibrate the ball PTZ cameras as there is no such offset.

REFERENCES

[1] Z. Wu and R. J. Radke, "Keeping a pan-tilt-zoom camera calibrated," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1994–2007, 2013.

[2] K. Song and J.-C. Tai, "Dynamic calibration of pan-tilt-zoom cameras for traffic monitoring," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, pp. 1091–1103, 2006.

[3] L. Álvarez, L. Gómez, P. Henríquez, and J. Sánchez, "Real-time camera motion tracking in planar view scenarios," *Journal of Real-Time Image Processing*, vol. 11, pp. 287–299, 2013.

[4] G. Chen, P.-L. St-Charles, W. Bouachir, G.-A. Bilodeau, and R. Bergevin, "Reproducible evaluation of pan-tilt-zoom tracking," *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 2055–2059, 2015.

[5] J. Chen, F. Zhu, and J. Little, "A two-point method for ptz camera calibration in sports," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 287–295, 2018.

[6] N. Homayounfar, S. Fidler, and R. Urtasun, "Sports field localization via deep structured models," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4012–4020, 2017.

[7] L. Sha, J. Hobbs, P. Felsen, X. Wei, P. Lucey, and S. Ganguly, "End-to-end camera calibration for broadcast videos," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13 624–13 633, 2020.

[8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1330–1334, 2000.

[9] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE J. Robotics Autom.*, vol. 3, pp. 323–344, 1987.

[10] Z. Zhang, "Camera calibration with one-dimensional objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 892–899, 2004.

[11] H. Zhang, G. Zhang, and K.-Y. K. Wong, "Camera calibration with spheres: linear approaches," *IEEE International Conference on Image Processing 2005*, vol. 2, pp. II–1150, 2005.

[12] M. Wilczkowiak, E. Boyer, and P. F. Sturm, "Camera calibration and 3d reconstruction from single images using parallelepipeds," *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, pp. 142–148 vol.1, 2001.

[13] H. Huang, H. Zhang, and Y. ming Cheung, "The common self-polar triangle of concentric circles and its application to camera calibration," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4065–4072, 2015.

[14] L. de Agapito, E. Hayman, and R. I. Hartley, "Linear self-calibration of a rotating and zooming camera," *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1, pp. 15–21 Vol. 1, 1999.

[15] Y. Seo and K. S. Hong, "About the self-calibration of a rotating and zooming camera: Theory and practice," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, pp. 183–189 vol.1, 1999.

[16] G. A. Thomas, "Real-time camera tracking using sports pitch markings," *Journal of Real-Time Image Processing*, vol. 2, pp. 117–132, 2007.

[17] J. Lu, J. Chen, and J. Little, "Pan-tilt-zoom slam for sports videos," in *BMVC*, 2019.

[18] F. Rameau, A. Habed, C. Demonceaux, D. Sidibé, and D. Fofi, "Self-calibration of a ptz camera using new lmi constraints," in *Asian Conference on Computer Vision*, 2012.

[19] Y. Seo and K. S. Hong, "Theory and practice on the self-calibration of a rotating and zooming camera from two views," 2001.

[20] B. He and Y. Li, "Camera calibration with lens distortion and from vanishing points," *Optical Engineering*, vol. 48, p. 013603, 2009.

[21] I. N. Junejo and H. Foroosh, "Practical ptz camera calibration using givens rotations," *2008 15th IEEE International Conference on Image Processing*, pp. 1936–1939, 2008.

[22] R. Wang, R. Huang, and J. Yang, "Facilitating ptz camera auto-calibration to be noise resilient with two images," *IEEE Access*, vol. 7, pp. 155 612–155 624, 2019.

[23] Y. Li, J. Zhang, W. Hu, and J. Tian, "Method for pan-tilt camera calibration using single control point." *Journal of the Optical Society of America. A, Optics, image science, and vision*, vol. 32 1, pp. 156–63, 2015.

[24] S. N. Sinha and M. Pollefeys, "Pan-tilt-zoom camera calibration and high-resolution mosaic generation," *Comput. Vis. Image Underst.*, vol. 103, pp. 170–183, 2006.

[25] J. Li, Y. min Yang, and G. Fu, "Camera self-calibration method based on ga-pso algorithm," *2011 IEEE International Conference on Cloud Computing and Intelligence Systems*, pp. 149–152, 2011.

[26] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, / 2nd ed ed. Cambridge University Press, 2003.

[27] D. Baráth, "Five-point fundamental matrix estimation for uncalibrated cameras," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 235–243, 2018.

[28] J. E. Gentle, "Numerical linear algebra for applications in statistics," 1998.

[29] P. Saponaro and C. Kambhamettu, "Towards auto-calibration of smart phones using orientation sensors," *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 20–26, 2013.

[30] G. Jiang, L. Quan, and H.-T. Tsui, "Circular motion geometry using minimal data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 721–731, 2004.

[31] Q.-T. Luong and O. D. Faugeras, "Self-calibration of a moving camera from point correspondences and fundamental matrices," *International Journal of Computer Vision*, vol. 22, pp. 261–289, 1997.

[32] G. Wang, Q. M. J. Wu, and W. Zhang, "Kruppa equation based camera calibration from homography induced by remote plane," *Pattern Recognit. Lett.*, vol. 29, pp. 2137–2144, 2008.

[33] T. Fetzer, G. Reis, and D. Stricker, "Stable intrinsic auto-calibration from fundamental matrices of devices with uncorrelated camera parameters," *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 221–230, 2020.

[34] Y. Ma, S. Soatto, J. Koseck, and S. S. Sastry, "An invitation to 3-d vision: From images to geometric models," 2003.