A Global Image Orientation Method of the Self-Rotating Pan-Tilt-Zoom Camera for Photogrammetric Applications

Teng Xiao[1], Qi Hu[1], Junhua Kang[2], Qi Zhang[1], Zhiwei Ye[1*], Fei Deng[3]

[1] School of Computer Science, Hubei University of Technology, 430068 Wuhan, China – (xiao, huqi1081, zhangqi_hbut)@hbut.edu.cn, weizhiye121@163.com

[2] College of Geological Engineering and Geomatics, Chang'an University, 710054 Xi'an, China – junhua.kang@chd.edu.cn

[3] School of Geodesy and Geomatics, Wuhan University, 430079 Wuhan, China – fdeng@sgg.whu.edu.cn

Commission II

KEYWORDS: Global Image Orientation, Structure from Motion (SfM), Pan-tilt-zoom (PTZ) Camera, Pure Rotation.

ABSTRACT:

Pan-tilt-zoom (PTZ) cameras are widely used in surveillance systems due to their wide field of view and high resolutions. However, the lack of accurate orientation information limits their full utilization in photogrammetry. Therefore, for photogrammetric applications, the primary task for PTZ cameras is to achieve image orientation. Cameras mounted on gimbals can only self-rotate around the base, resulting in acquired images that are nearly purely rotated. Current conventional structure from motion (SfM) pipelines assume pixel parallax exists between matching correspondences and generate object points through 3D triangulation. Applying these methods to estimate the interior and exterior orientation parameters of images from a self-rotating PTZ camera is challenging. To address this issue, this paper employs the concept of global SfM and proposes an improved global image orientation method for the pure rotation motion of PTZ cameras. Initially, a subset of image pairs is selected for internal orientation, and the internal orientation parameters of the images are estimated. Subsequently, global external orientation is performed on all images to estimate their external orientation parameters. Finally, bundle adjustment of the collinearity equation without object points optimizes both the internal and external orientation parameters. Experiments on synthetic and real-scene datasets demonstrate the practicality and accuracy of this method. For synthetic datasets, the estimated focal length of our method deviates from the true value by within 1 pixel, and the mean location error of the principal points is 0.93 pixels. For real-scene datasets, the mean reprojection error of the checkpoints of our method is 2.72 pixels, with a maximum of 4.66 pixels. In contrast, Agisoft Metashape's mean reprojection error is 4.73 pixels, with a maximum reaching 8.06 pixels. This shows that our method can accurately determine the image orientation parameters of PTZ cameras and achieve higher accuracy compared to the popular commercial software Agisoft Metashape.

1. INTRODUCTION

Pan-tilt-zoom (PTZ) cameras (Sinha, 2021) have some degrees of pan, tilt, and zoom control and enable a flexible way for images acquisition. They are widely used in surveillance systems with the advantages of a wide field of view and high resolutions. Regarding the other applications of PTZ cameras, some works have studied panoramic stitching (Yong et al., 2019), object recognition and tracking (Jinlong et al., 2024; Nebeluk et al., 2023; Yun et al., 2021) and intelligent transportation (Haghighat and Sharma, 2023; Zhang et al., 2022). However, PTZ cameras are not fully utilized in the field of photogrammetry and geoinformatics. One reason is that these captured images often lack of accurate orientation information. Therefore, for photogrammetric applications, the first task for PTZ cameras is to achieve image orientation, i.e., to be able to automatically calibrate the interior and exterior orientation parameters (Liu and Zhang, 2023; Wu and Radke, 2012; Zhang et al., 2020).

Some related works on estimating image orientation parameters are first reviewed. They can be divided into two categories according to whether reference object points are used. By placing a plane calibration plate indoor, Wu and Radke (2012) established a functional relationship between the interior orientation parameters and the zoom ratio of lens of PTZ cameras. In additional, similar to place recognition, it roughly estimated the exterior orientation parameters by matching the pre-generated feature database with the PTZ camera image. For special scenes such as the sports field, Chen et al. (2018) employed the structured point-line features to calculate the image orientation parameters of the PTZ camera. For road scenes, Wang et al. (2020), Song et al. (2021) and Esin et al. (2024) used lane boundary lines and vehicle outline to calculate the vanishing point and then used it to estimate the image orientation parameters. These methods are often limited to the reference points in specific structured scenes.

Correspondingly, the methods that do not use reference object points are more flexible. In this category, it is typically to employ the so-called structure from motion (SfM) technique (Schonberger and Frahm, 2016). For a set of multi-view images, SfM can automatically estimate the interior and exterior orientation parameters at the same time based on the multi-view geometry. As a result, SfM has become one of the most popular methods for achieving image orientation parameters. However, conventional SfM pipelines assume pixel parallax exists between matching correspondences and then generate object points by 3D triangulation. Subsequently, a self-calibrating bundle adjustment is employed and the interior and exterior parameters of image orientation can be accurately obtained. As for PTZ cameras, they are mounted on the gimbal and can only self-rotate around the base. It results in the acquired images being with nearly pure rotation. And for these images, it is not feasible to generate object points. Therefore, conventional SfM methods are not applicable to image orientation tasks of PTZ cameras.

To solve this problem, this paper considers the idea of

*Corresponding author, weizhiye121@163.com.

rotation averaging (Xiao et al., 2021; Zhang et al., 2023) and propose an improved global image orientation pipeline for estimating interior and exterior parameters for PTZ cameras. Compared with the popular commercial software Agisoft Metashape, our method can obtain higher accuracy on the orientation results.

2. RELATED WORKS

Our work is related to the image orientation methods, typically SfM. It can be divided into two categories: incremental and global SfM. Incremental SfM starts with an initial reconstruction of a subset of images, e.g., a pair or a triplet, and then sequentially adds further images to the block with repeated bundle adjustment. Global SfM, on the other hand, deals with all available images simultaneously, i.e. relative orientations of all overlapping image pair are taken as input to compute global poses, typically in a two-step procedure, first carrying out global rotation averaging for global rotation matrices, followed by global translation averaging for global translation matrices.

Incremental SfM: Schönberger and Frahm (2016) proposed a very popular incremental SFM framework COLMAP, which added some robust strategies to the conventional pipeline. For example, in selecting the next best view it considers the distribution of matching points and gives priority to candidate images with more uniform distribution. In order to achieve robust and efficient triangulation, the RANSAC strategy is added to the DLT algorithm, and the intersection angle and object point depth are verified, and local bundle adjustment is performed after adding this image. In addition, in scene expansion, global bundle adjustment is performed after adding a certain number of images. After adding images using the incremental SfM method, Wang and Heipke (2020) first calculated the global rotation of the candidate image based on the relative rotation between the new image and the oriented image using the single rotation average (Hartley et al., 2013), and then used the linear equation to solve its global translation, thereby improving the efficiency of absolute orientation of the new image.

In recent years, some research works have used the concept of view-graph representing the matched image pairs to improve the time efficiency of incremental SfM (Cui et al., 2021; Gong et al., 2023) to meet the requirements of large-scale image 3D reconstruction. Cui et al. (2021) proposed a new method for constructing an sparse view-graph, and used an incremental local reconstruction strategy to save computational memory, thereby improving the efficiency and robustness of SfM when processing large data. Gong et al. (2023) proposed a coarse-to-fine subgraph extraction method from the original view-graph, which on the one hand eliminates redundant and erroneous image pairs, and on the other hand takes into account the distribution of subgraph vertices, thereby greatly improving the computational efficiency of the subsequent incremental SfM.

Rotation averaging: Hartley et al. (2013) reviewed rotation averaging in detail with different forms of rotation deviation measurement. Chatterjee and Govindu (2013) further emphasized the efficiency and robustness of rotation and translation. They first used Lie groups to represent rotation and L1 norm to relatively calculate initial rotation values, and then optimized them through iterative reweighted least squares method and implemented a loss function similar to Huber. The negative impact of gross errors was resisted by dynamically adjusting the weights of the residuals. Reich et al. (2017) improved the method of Chatterjee and Govindu (2013) through convex relaxation semidefinite procedures to obtain more robust global rotation initial values. Wilson et al. (2016) pointed out that local convexity analysis is almost impossible to achieve in L2 norm, and pointed out that it is mainly affected by the relative rotation noise and the network structure of image association. Wilson and Bindel (2020) extended the work of Wilson et al. (2016) on the non-convexity of rotational averaging, studied the spatial distribution of local minima, and derived theoretical boundaries for different distributions, pointing out that local convexity can be used to represent the intrinsic measure of the difficulty of rotational averaging. Chatterjee and Govindu (2017) tested different robust estimation methods and suggested using the $L_{1/2}$ norm when performing rotational translation to reduce the negative impact of gross errors in relative orientation. Gao et al. (2021) introduced the incremental SfM robust mechanism into rotation averaging to improve its robustness, and proposed a corresponding incremental rotation averaging method. They first selected image triplets to construct the camera global rotation as seeds, and then incrementally added the best candidate images. During the expansion process, they were optimized by local rotation averaging and global rotation averaging until the global rotation of all images was successfully solved. Zhang et al. (2023) revisited the problem of rotation averaging by leveraging the uncertainties inherent in two-view epipolar geometry and investigating robust loss functions. They demonstrated that incorporating the covariance matrix of uncertainties directly into the optimization process can significantly enhance the quality of estimated global rotation matrices.

3. METHOD

This paper proposed an improved global image orientation pipeline for PTZ cameras, and the workflow is given in Figure 1. For a set of input images with pure rotation of PTZ camera, SIFT algorithm is typically used for feature extraction and matching, and the infinite homography matrices of image pairs are calculated. The subsequent process can be divided into three steps: the estimation of the interior parameters, the estimation of the exterior parameters, and the final global optimization using bundle adjustment without object points.

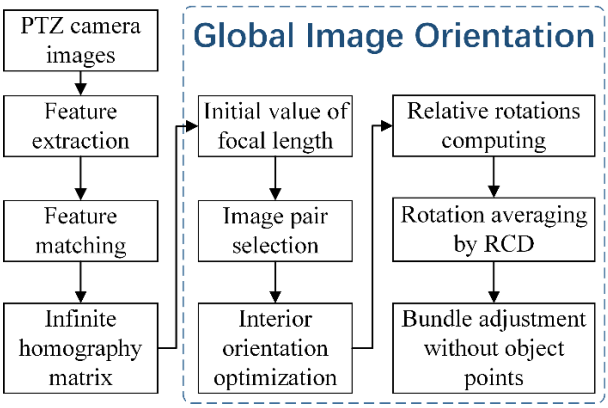

Figure 1. The workflow of the proposed method.

3.1 Estimation of the interior parameters

The first task for the estimation of interior orientation parameters is to obtain an initial focal length value. Specifically, we refer to the multi-view geometry to obtain initial focal length value by using all the infinite homography matrices. For each matched image pair (i, j), its homography H_{ij} is known and has been normalized, such that $\det(H_{ij}) = 1$. And the equation (1) shows the relationship between the image of the absolute conic ω and the homography H_{ij}.

$$\omega = (H_{ij})^{-T} \omega (H_{ij})^{-1} \quad (1)$$

Then it can be rewritten as

$$Ac = 0 \tag{2}$$

Where A is a $6m \times 6$ matrix, here m means the number of matched image pairs, and c is arranged as a 6-vector with the elements of the conic ω. A least-squares solution of the equation (2) can be obtained by SVD decomposition. ω is carried out by Cholesky decomposition as $\omega = UU^T$, and thence we use the equation (3) to obtain the camera matrix.

$$K = U^{-T} \tag{3}$$

The principal point is set to the center of the image, and the initial value of the distortion parameter is set to zero. Through these processes, the initial values of the interior orientation parameters can be obtained. However, these values are very far from the true value, especially the focal length. If the focal length deviation is very large, it will reduce the accuracy of the subsequent relative rotation calculation. Therefore, it is critical to obtain a more accurate initial value of the focal length.

To do so, we designed the interior orientation optimization (see Figure 1). Our idea to achieve this is to select a set of image pairs with good configuration, which helps to find a focal length value close to ground truth. According to the number of the matching correspondences, we select the top k matched image pairs $\mathcal{E}_{select}$ with the largest number and have the correspondingly set of homography matrices $\{H_{ij}\}_{select}$. Then, the K matrix can be refined by minimizing

$$\sum_{(i,j)\in\mathcal{E}_{select}} d\left(x_j, KR_{ij}K^{-1}x_i\right)^2$$

over K and R_{ij}, where x_i, x_j are the locations of the correspondences between image pair (i,j). And the initial estimates of R_{ij} for the minimization is obtained by $R_{ij} = K^{-1}H_{ij}K$.

3.2 Estimation of the exterior parameters

In this stage, the matrix of the selected image pair and the initial values of the interior orientation parameters mentioned above are taken as input, and the relative rotations are calculated. Then, the global rotation matrix of all images, i.e., the exterior orientation elements of each image, is calculated by rotation averaging (Dümbgen et al., 2024; Zhang et al., 2023). Specifically, several relative rotations R_{ij} are given with different coordinate frames, and n global rotations $\{R_i | i \in (1 \cdots n)\}$ are computed to satisfy the compatibility constraint $R_{ij} = R_j R_i^{-1}$. The index (i,j) is in the set of the selected image pairs S. In the presence of noise, the rotation averaging problem is expressed as seeking

$$\underset{R_1 \cdots R_n}{\operatorname{argmin}} \sum_{(i,j)\in S} d\left(R_{ij}, R_j R_i^{-1}\right)^2 \tag{4}$$

Where $d(\cdot,\cdot)^2$ measures the difference of two rotation matrices with $L2$ norm. In our case, we employ rotation coordinate descent (RCD, Parra et al., 2021) algorithm for equation (4) to achieve enhanced computational efficiency.

3.3 Bundle adjustment without object points

Similar with conventional SfM methods, the last step is to use bundle adjustment to optimize the interior and exterior orientation parameters. But in this case, the pure rotational motion of the camera makes it impossible to generate the 3D object points and collinear equation is not suitable for this problem. A special idea is to use bundle adjustment without object points, such as structureless bundle adjustment (Cefalu et al., 2016) and pointless bundle adjustment (Rupnik and Pierrot-Deseilligny, 2023).

The initial values of K and the global rotations $\{R_i | i \in (1 \cdots n)\}$ have been obtained. Then the correspondences between the selected image pairs $\mathcal{E}_{select}$ are used to refine K and $\{R_i\}$ by minimizing the equation (5).

$$K, R_i = \operatorname{argmin} \sum_{(i,j)\in\mathcal{E}_{select}} d\left(x_j, KR_j R_i^T K^{-1} x_i\right)^2 \tag{5}$$

where x_i, x_j are the locations of the correspondences between image pair (i,j).

4. RESULTS

To evaluate our proposed method for PTZ cameras, experiments on four synthetic and one real-scene dataset were conducted. For comparisons, we also used the commercial software Agisoft Metashape to orientate these PTZ camera images.

4.1 Measurement of rotation accuracy evaluation

In order to measure the difference between the estimated rotation matrix and the ground true value, a distance function between two rotation matrices is defined:

$$\delta(R_1, R_2) = \arccos\left(\frac{(trace(R_1 R_2^T) - 1)}{2}\right) \tag{6}$$

Here $\delta(R_1, R_2)$ is the angle value in radians, which will be further converted to angle values for a more intuitive effect. In the absence of control points for absolute orientation, the estimated global rotation matrices $\{R_1, \cdots R_n\}$ is not in the same coordinate system as the ground truth values $\{R_1^{gt}, \cdots R_n^{gt}\}$. In this case, we need to find a rotation matrix $\bar{R}$ to align the two coordinate systems first and it should minimize the difference between the transformed rotation matrix $\{R_i \bar{R}\}$ and the ground truths $\{R_i^{gt}\}$. It will have the equation (7)

$$\bar{R} = \underset{\bar{R}}{\operatorname{argmin}} \sum_{i=1}^{n} d\left(\bar{R}, R_i^T R_i^{gt}\right) \tag{7}$$

After transformation by $\bar{R}$, the equation (6) is used to evaluate the accuracy of the rotation matrix of each image, and then the minimum, median and maximum values can be statistically analyzed.

4.2 Synthetic datasets

Four synthetic datasets (see Table 1) were generated by simulating the PTZ camera shooting process with specific orientation parameters in scenes that have been pre-built into 3D models. We put the prepared 3D model (mesh model with texture) into the 3D rendering software, and set different interior and exterior orientation parameters to render image datasets of different perspectives. Because of the size of the display screen, the image resolution is 1920×1080.

As for the interior orientation parameters, we simulated the different focal lengths, and the principal point is the center of the image. It should also be noted that these rendered images are distortion-free. As for the exterior orientation parameters, we simulated a series of P, T values using fixed angle intervals, and then converted them into different rotation matrices and input them to the rendering software. Figure 2 expresses the entire scene by stitching all the images into a panoramic image. It can be seen that the interior and exterior orientation parameters of the synthetic images are known and these synthetic datasets can be used to evaluate the accuracy of the estimated intrinsic parameters, especially focal lengths, and the estimated extrinsic parameters, e.g., global rotation matrices.

Datasets	No.	Size /pixels	Focal length /pixels	Sample images
Theater	156	1920×1080	800	
Playground	96	1920×1080	1800	
City1	54	1920×1080	2000	
City2	240	1920×1080	3000	

Table 1. Data description of the synthetic datasets. 'No.' means the number of images, 'Size' represents the resolution of images and 'Focal length' here indicates the ground truth values when generating the rendered images. Sample images from different perspectives by setting different P and T values.

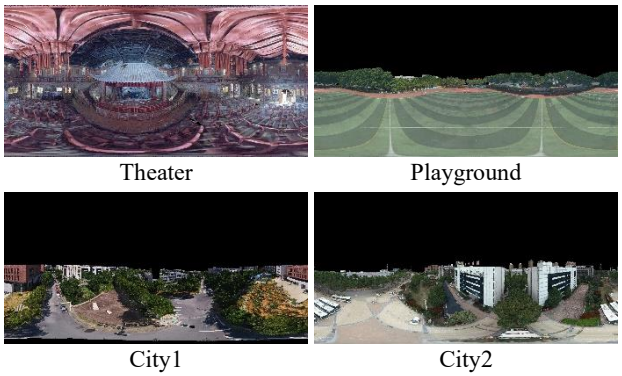

Figure 2. Four synthetic datasets displayed in panoramic images.

These Synthetic datasets were carried out by our method and the Agisoft Metashape and the results of orientation parameters are numerically analyzed, as is shown in Table 2. The error of interior orientation parameters is the difference between the estimated focal length, principal point coordinates and the ground truth ones. The error of exterior orientation parameters is the difference between the estimated rotation matrices and the ground truth ones, measured by angles.

As for the interior orientation parameters, expect the *Theater* dataset, our obtained the comparable accuracy and the closer values to the ground truth focal length. As for the exterior orientation parameters, our method achieved the better accuracy for all datasets. It can be seen that our method can obtain the more accurate orientation parameters than Agisoft Metashape. According to the numerical analysis in Table 3, it can be seen that the focal length directly calculated using only the homography matrices has a large deviation from the true value. However, after the interior orientation optimization using the selected image pair, the focal length is very close to the true value.

To further analyze our method, we will evaluate the accuracy of the results with initial values of the intrinsic parameters before bundle adjustment (see Table 3). Special attention should be paid to the estimation and optimization process of focal length. Due to the zooming of the PTZ camera lens, it is difficult to empirically give the initial value of the focal length of the image obtained. This shows that the optimization step we adopted is very helpful in finding a more accurate initial focal length value, which has a very positive effect on the subsequent relative orientation and bundle adjustment.

Dataset	Method	Error of IO /(pixels)			Error of EO /(degrees×10^{-1})		
		df_{final}	dpx_{final}	dpy_{final}	min	median	max
Theater	Ours	0.49	0.76	0.96	**0.90**	**0.91**	**0.93**
	AM	**0.18**	**0.31**	**0.41**	1.02	1.06	1.13
Playground	Ours	**0.87**	**0.59**	1.24	**0.88**	**1.04**	**1.25**
	AM	0.88	1.05	**0.75**	0.99	1.09	1.29
City1	Ours	**0.41**	**0.92**	0.81	**0.81**	**0.85**	**0.91**
	AM	0.90	1.09	**0.25**	0.91	1.05	1.19
City2	Ours	**0.80**	3.32	0.88	**0.71**	**0.96**	**1.22**
	AM	1.48	**2.33**	**0.71**	1.29	1.77	1.51

Table 2. Comparison of accuracy of orientation parameters. 'AM' means Agisoft Metashape, 'IO' means interior orientation and 'EO' means exterior orientation. df_{final}, dpx_{final}, dpy_{final} represent the difference between the final focal length and principal point coordinates after bundle adjustment and the ground truth ones.

Datasets	df_{homo}	df_{opti}	dpx_{opti}	dpy_{opti}
Theater	3650.78	0.53	0.81	1.00
Playground	8,999.02	1.31	0.82	1.29
City1	598.92	1.83	0.89	0.77
City2	9,692.2	7.64	5.65	4.01

Table 3. Accuracy evaluation of initial values of internal parameters before bundle adjustment. df_{homo} means the difference between the average of the focal lengths directly calculated from the homography matrices of all matched image pairs and the ground truth one (as shown in Table 1). df_{opti}, dpx_{opti}, dpy_{opti} are the differences after the interior orientation optimization form the selected image pairs.

4.3 Real scene dataset

The real-scene dataset was collected by a Hikvision PTZ camera installed on a high tower in the city. During the rotation of the camera, the zoom ratio was fixed with a step size of 30 degrees in the horizontal direction and 14 degrees in the vertical direction. One set of 360-degree panoramic images with 66 video frames and [0,71] degrees in the vertical direction was collected. Table 4 shows the internal orientation parameter results obtained by our method and AM software processing real scene data, which are numerically consistent. In addition, the panoramic image spliced according to our estimated internal and external parameters is shown in Figure 3, and there is no obvious misalignment in the panoramic image.

method	f/(pixels)	px/(pixels)	py/(pixels)	k_1	k_2
Ours	2101.130	949.675	572.746	-0.182	0.147
AM	2103.640	958.798	569.203	-0.189	0.154

Table 4 Results of the interior orientation parameters of our method and Agisoft Metashape.

In order to further evaluate the accuracy of the orientation results, the panoramic image is absolutely oriented with the help of control points, the pose in world coordinates is restored, and then the reprojection error of the control points and check points is evaluated. The points are punctured on the panoramic image to obtain the correspondence between the pixel coordinates of the control points and the world coordinates, and the PnP method is used to absolutely orient the panoramic image. Finally, the world coordinates of the control points and check points are projected onto the panoramic image according to the absolute orientation results, and the reprojection error is calculated.

And also, we assigned 25 control points and 5 check points in this city scene. The point distribution is shown in Figure 3. Among them, the control points were used in combined bundle

adjustment for absolute orientation results, and the reprojection errors of check points were used to evaluate the accuracy of our and Agisoft Metashape's results, as is shown in Figure 4. It can be seen that the reprojection errors of the checkpoints obtained by our method are all lower than those of Agisoft Metashape, which means that our method estimates more accurate orientation parameters.

Figure 3. One real-scene datasets displayed in the form of panoramic images

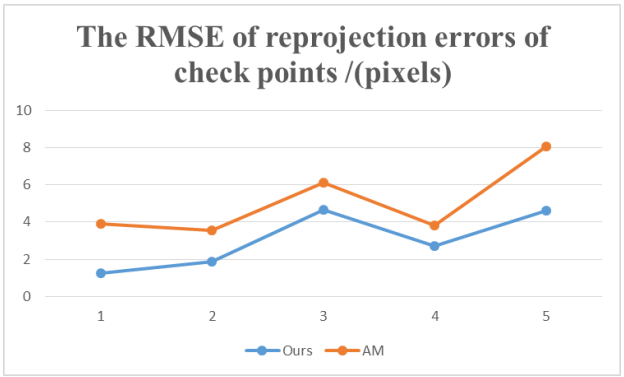

Figure 4. Accuracy analysis of orientation results, and 'AM' means Agisoft Metashape.

5. DISCUSSION AND CONCLUSION

Experiments on synthetic and real-scene datasets demonstrate the practicality and accuracy of our method. For the synthetic datasets, the estimated focal length of our method deviates from the true value within 1 pixel, and the mean location error of the principal points is 0.93 pixels. For the real-scene dataset, the mean reprojection error of the checkpoints of our method is 2.72 pixels, with the maximum being 4.66 pixels. For the results of Agisoft Metashape, in contrast, its mean reprojection error is 4.73 pixels, with the maximum up to 8.06 pixels.

It shows that our method can accurately determine the image orientation parameters of the PTZ camera. Compared with the popular commercial software Agisoft Metashape, our method can obtain higher accuracy on the orientation parameters. Our method enhances the application potential of PTZ cameras in the field of photogrammetry.

REFERENCES

Cefalu, A., Haala, N., & Fritsch, D. (2016). Structureless bundle adjustment with self-calibration using accumulated constraints. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3, pp. 3-9.

Chatterjee, A., & Govindu, V. M. (2013). Efficient and robust large-scale rotation averaging. In Proceedings of the IEEE international conference on computer vision, pp. 521-528.

Chatterjee, A., & Govindu, V. M. (2017). Robust relative rotation averaging. IEEE transactions on pattern analysis and machine intelligence, 40(4), pp. 958-972.

Chen, J., Zhu, F., & Little, J. J. (2018, March). A two-point method for PTZ camera calibration in sports. In 2018 IEEE winter conference on applications of computer vision (WACV), pp. 287-295.

Cui, H., Shi, T., Zhang, J., Xu, P., Meng, Y., & Shen, S. (2021). View-graph construction framework for robust and efficient structure-from-motion. Pattern Recognition, 114, 107712.

Dümbgen, F., Holmes, C., Agro, B., & Barfoot, T. D. (2024). Toward globally optimal state estimation using automatically tightened semidefinite relaxations. IEEE Transactions on Robotics, pp. 4338-4358.

Esin, Y. E., Özdil, Ö., Bilge, Y. C., & Öztürk, Ş. (2024, May). Precise On-Site Calibration of PTZ Cameras. In 2024 32nd Signal Processing and Communications Applications Conference (SIU), pp. 1-4.

Gao, X., Zhu, L., Xie, Z., Liu, H., & Shen, S. (2021). Incremental rotation averaging. International Journal of Computer Vision, 129, pp. 1202-1216.

Gong, Y., Zhou, P., Liu, Y., Dong, H., Li, L., & Yao, J. (2023). View-graph key-subset extraction for efficient and robust structure from motion. The Photogrammetric Record, 38(183), pp. 252-296.

Haghighat, A., & Sharma, A. (2023). A computer vision-based deep learning model to detect wrong-way driving using pan–tilt–zoom traffic cameras. Computer-Aided Civil and Infrastructure Engineering, 38(1), pp. 119-132.

Hartley, R., Trumpf, J., Dai, Y., & Li, H. (2013). Rotation averaging. International journal of computer vision, 103, pp. 267-305.

Jinlong, E., Han, F., He, L., Xu, W., Li, Z., Chai, Y., & Liu, Y. (2024). WiseCam: A Systematic Approach to Intelligent Pan-Tilt Cameras for Moving Object Tracking. IEEE Transactions on Mobile Computing, pp. 1-14.

Liu, Y., & Zhang, H. (2023). Linear Auto-calibration of Pan-Tilt-Zoom Cameras With Rotation Center Offset. In 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 11461-11467.

Nebeluk, R., Zarzycki, K., Seredyński, D., Chaber, P., Figat, M., Domański, P. D., & Zieliński, C. (2023). Predictive tracking of an object by a pan–tilt camera of a robot. Nonlinear Dynamics, 111(9), pp. 8383-8395.

Parra, A., Chng, S. F., Chin, T. J., Eriksson, A., & Reid, I. (2021). Rotation coordinate descent for fast globally optimal rotation averaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4298-4307.

Reich, M., Yang, M. Y., & Heipke, C. (2017). Global robust image rotation from combined weighted averaging. ISPRS journal of photogrammetry and remote sensing, 127, pp. 89-101.

Rupnik, E., & Pierrot-Deseilligny, M. (2023). Pointless Global Bundle Adjustment With Relative Motions Hessians. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6517-6525.

Schonberger, J. L., & Frahm, J. M. (2016). Structure-from-motion revisited. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4104-4113.

Sinha, S. N. (2021). Pan-Tilt-Zoom (PTZ) Camera. In Computer Vision: A Reference Guide. Cham: Springer International Publishing, pp. 941-947.

Song, J., Song, H., & Wang, S. (2021). PTZ camera calibration based on improved DLT transformation model and vanishing Point constraints. Optik, 225, 165875.

Wang, W., Zhang, Z., Tang, X., & Song, H. (2020). Dynamic Self-Calibration Algorithm for PTZ Camera in Traffic Scene. In Information Technology and Intelligent Transportation Systems, IOS Press, pp. 176-185.

Wang, X., & Heipke, C. (2020). An improved method of refining relative orientation in global structure from motion with a focus on repetitive structure and very short baselines. Photogrammetric Engineering & Remote Sensing, 86(5), pp. 299-315.

Wilson, K., & Bindel, D. (2020). On the distribution of minima in intrinsic-metric rotation averaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6031-6039.

Wilson, K., Bindel, D., & Snavely, N. (2016, September). When is rotations averaging hard?. In European Conference on Computer Vision, Cham: Springer International Publishing, pp. 255-270.

Wu, Z., & Radke, R. J. (2012). Keeping a pan-tilt-zoom camera calibrated. IEEE transactions on pattern analysis and machine intelligence, 35(8), pp. 1994-2007.

Xiao, T., Wang, X., Deng, F., & Heipke, C. (2021). Sequential Cycle Consistency Inference for Eliminating Incorrect Relative Orientations in Structure from Motion. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 89(3), pp. 233-249.

Yong, H., Huang, J., Xiang, W., Hua, X., Zhang, L. (2019). Panoramic background image generation for PTZ cameras. IEEE Transactions on Image Processing, 28(7), pp. 3162-3176.

Yun, K., Kim, H., Bae, K., & Park, J. (2021, January). Unsupervised moving object detection through background models for ptz camera. In 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, pp. 3201-3208.

Zhang, C., Rameau, F., Kim, J., Argaw, D. M., Bazin, J. C., & Kweon, I. S. (2020). Deepptz: Deep self-calibration for ptz cameras. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1041-1049.

Zhang, G., Larsson, V., & Barath, D. (2023). Revisiting rotation averaging: Uncertainties and robust losses. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 17215-17224.

Zhang, X., Feng, Y., Angeloudis, P., Demiris, Y. (2022). Monocular visual traffic surveillance: A review. IEEE Transactions on Intelligent Transportation Systems, 23(9), pp. 14148-14165.