Real-Time Pan-Tilt-Zoom Camera Calibration in Virtual Scene

Liangliang Cai
State Key Laboratory of Virtual Reality Technology and Systems
Beihang University Beijing, China
caill0429@buaa.edu.cn

Zhong Zhou*
State Key Laboratory of Virtual Reality Technology and Systems
Beihang University Beijing, China
zz@buaa.edu.cn

***Abstract*—Pan-tilt-zoom (PTZ) cameras are pervasive in Modern Society. However, research on the mixed reality system of PTZ Cameras is scarce due to the flexibility of PTZ cameras relying on visual information without sensors feedback and the complexity of the real world, such as lighting changes, occlusion, and lack of texture. To solve this problem, we propose a motion estimation model for PTZ cameras obtained from real PTZ camera motion observation. We show how the motion estimation model real-time estimates the pose of the camera in moving. Next, we propose a state-of-art PTZ calibration algorithm based on key-rays collection. We first extract SIFT features from the scene of the PTZ camera to build a key-rays collection. Then we obtain a group of the parameters of the current image against the collection using the two-point method and find best the parameters by the idea of RANSAC. In different real scenes, we evaluate the motion estimation model and calibration method on different cameras, demonstrating the accuracy of the motion estimation model and the effectiveness of the calibration.**

***Index Terms*—mixed reality, PTZ camera, camera calibration, PTZ model, key-rays collection**

I. Introduction

Nowadays, the large-scale mixed reality system play a critical role in surveillance, security, criminal investigation, sports analysis and epidemic prevention. Most of previous researches focused on the calibration of a fixed or handheld camera [1], [2]. It is difficult for the user to fully comprehend the movement of dispersed objects in different screens. Tracing the target from camera to another is challenging because the spatial relation between different camera views is not easy to perceive. Besides, the quality of photos and 3D models will influence the result of the camera calibration. The Pan-Tilt-Zoom (PTZ) camera is a better choice, which changes its view to observe a wider perspective and to easy tracking targets by users. Moreover, PTZ cameras have become increasingly popular in many scenes, such as hotels, schools, universities, companies, sport games, etc. Many companies like SPORTLOGiQ [3] try to integrate PTZ cameras with 3D scenes to extract players statistics.

Most previous approaches of PTZ cameras calibration in mixed reality systems focus on sports analysis [4]. In Sports analysis systems, the camera of the new image is calibrated by matching keypoints (e.g. SIFT [5]) from the new image against reference images. On the one hand, many real scenes is more complex than sport fields, and these calibration methods cannot achieve good results. More robust PTZ calibration method for any scene needs to be proposed. On the other hand, most of the ways mainly solve the camera calibration of single image, but solving the real-time calibration of the PTZ camera is intractable. Although camera tracking in mixed reality systems of mobile handhold cameras [2], [6] can solve this problem, it depends on gyroscopes, acceleration sensors, global position system (GPS) or other sensor data, which are not found on general PTZ cameras.

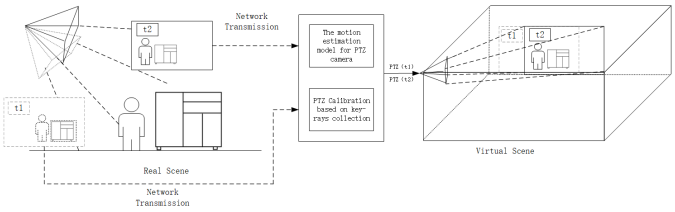

Fig. 1. Real-time PTZ Camera Calibration in virtual scene. Our Method sets up a bridge between PTZ cameras in real scene and PTZ cameras in virtual scene. t_1 and t_2 represent the last moment and the current moment. The PTZ camera in real scene translates the current image to our method, and our method obtain the pose and send it to the PTZ camera in virtual scene. The PTZ camera in virtual scene use the pose to instead of the last pose.

In this paper, we present a new framework for solving the PTZ camera calibration in mixed reality systems. Instead of conventional methods, we propose a motion estimation model for the PTZ camera. The model calibrates the camera of images from the PTZ camera in moving. Also, we present a novel PTZ camera calibration based on key-rays collection, which use the two-point method [7] and the idea of RANSAC [8]. First, we extract SIFT features from images in the real scene, and then we store the features with spherical coordinates, which are treated as rays, into the key-rays collection. Third, we estimate a group of initial guesses for the camera of the frame from the video using the two-point method and find the best estimate by the idea of RANSAC. This method requires no information from the virtual scene and directly estimates all parameters of the model without homography decompose. Fig. 1 illustrates the idea of our method. The camera in the real

National Key R&D Program of China under Grant No.2018YFB2100601 and Natural Science Foundation of China under Grant No.61872024.

Zhong Zhou is corresponding author.

world streams video to the PTZ camera real-time calibration module. The module is responsible for the real-time camera calibration and translates videos and camera poses to virtual scenes, such as half-sphere models, 3d models, and point clouds. In addition to this, our method requires a preliminary stage for acquisition of the key-rays collection and the camera motion estimation model. In summary, our paper has three main contributions:

- An effective motion estimation model for PTZ cameras is presented. This model establish the functional relationship between time and PTZ camera parameters during rotation and zooming.
- A new PTZ camera calibration algorithm based on key-rays collection is developed to calibrate a fixed location PTZ camera, which use the two-point method and the idea of RANSAC.
- A mixed reality framework of the PTZ camera is proposed to achieve the projection of moving images from the PTZ camera to virtual scenes.

II. Related Work

A. PTZ camera calibration

PTZ camera calibration algorithms estimate the parameters of the camera model by finding correspondences between points in images (or points from images and points from 3D models). Most of them focus on sports analysis. Homayounfar et al. [9] formulate this problem as a branch and bound inference in a Markov random field, which is fully automatic and depends only on a single image from the broadcast video; Liu et al. [10] present a novel homography computational algorithm which can improve the homography computational accuracy and reduce the processing time; Chen et al. [11] use a siamese network to learn compact deep features and use a novel two-GAN model to detect field marking in real images. Citraro et al. [12] propose a novel framework that use a fully-convolutional deep network to combine localization and robust identification of keypoints in the image. Sha et al. [13] design an end-to-end deep framework for single moving camera calibration. Those methods depend on remarkable lines and intersections in courts, but they would fail when they are applied to other scenes.

Few works attempt to find such a relationship between control points and the parameters of the PTZ camera. Wu et al. [14] present a dynamic calibration algorithm based on matching the current image against a stored feature collection created at the time the PTZ camera is mounted. This method directly obtains the PTZ camera's intrinsics and extrinsics by feature point pairs. But this method is suitable for small angle and focal length offset. Chen et al. [10] propose a two-point method which requires only two point correspondences to calibrate the PTZ camera, and they also propose a fast random forest method to predict pan-tilt angles without image-to-image feature matching. The random forest need control point pairs for training, and the more point pairs trained, the better the result predicted. Chen et al. [15] have further extended the Two-point method into long sequences in the next year, and improve offline random forest algorithm to online random forest algorithm to meet real-time needs; An et al. [16] propose a novel two-point calibration method (TPCM) that can estimate the focal length and 3-DoF rotation matrix with only two control points from one image, which considers the effect of the radial distortion.

B. Fixed camera and handhold camera calibration in mixed reality systems

Mixed reality truly reflects the real scene through images captured from cameras and reduces the workload of 3D scene modelling and rendering. Sawhney et al. [17] first present a video flashlight system that illuminates a static 3D graphics model with live video textures from stationary and moving cameras. Chen et al. [18] propose a novel visualization framework for surveillance systems, which project the large-scale display area using the fixed projector and project a fovea area with the high-resolution image of a selected camera. Zhou et al. [19] propose a novel virtual-real video fusion system based on a video Model which uses the single-image modelling technology. Pece et al. [6] present a PanoInserts system that uses a combination of marker- and image-based tracking to position the video inserts within the panorama. Tompkin et al. [20] create a video-collections+context interface by embedding videos into a panorama. In this paper, they build a spatio-temporal index and tools for fast exploration of space and time of the video collection. Young et al. [21] present a system that provides immersive telepresence and remote collaboration on mobile and wearable devices. They build a live spherical panoramic representation of a user's environment that can be viewed in real-time by a remote user who can independently choose the viewing direction.

III. A Motion Estimation Model for PTZ Camera

The PTZ camera model used in most researches is usually the common pinhole camera model with rotation matrixes, representing the static state of a PTZ camera. This model could not describe changes of PTZ camera parameters in moving. Our propose a novel motion estimation model that is continuous as a function of time. We can accurately obtain the model with a series of simple initialization steps.

In this section, we first introduce the pinhole camera model with rotation matrixes, treated as the static PTZ camera model. Then we describe the motion estimation model we proposed in detail.

A. Pinhole camera model with rotation matrixes

Most researchers [7], [10] use the pinhole model to describe a PTZ camera

$$P = \underbrace{KQ_\phi Q_\theta}_{\text{PTZ}} \underbrace{S[I \mid -\mathbf{C}]}_{\text{prior}} \tag{1}$$

Where C is the camera's position. The group of $Q_\phi Q_\theta S$ represents rotations from world to camera coordinates. The S indicates the initial PTZ camera position and orientation.

Then the camera can reach new orientation by Q_θ and Q_ϕ. K is the intrinsic matrix. We Assume square pixels and principal point at the image center (u, v), the focal length f is the only unknown variable in K.

P can be separated into two parts. The right part $S[I|-C]$ is time invariant for a fixed PTZ camera. Various PTZ configurations [22] can estimate this part by images. The left part $KQ_\phi Q_\theta$ is time variant, which needs a motion estimation model to characterize.

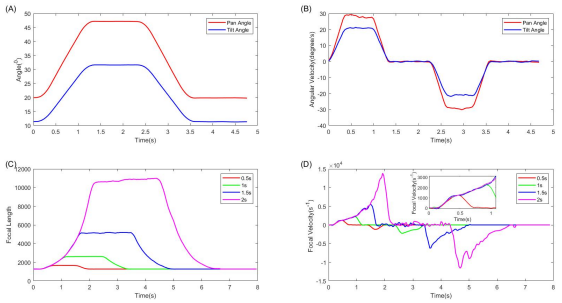

Fig. 2. The average relationships of the pan θ, the tilt ϕ, and the focal length f with time. (A): pan/tilt angle as a function of time; (B): pan/tilt angle velocity as a function of time; (C) focal length as a function of time; (D) focal length velocity as a function of time.

B. Motion estimation model of PTZ camera

In order to require the motion estimation model, we conduct a simplified experiment. We let a PTZ camera be directed to purely pan (or tilt) for 1s and stop, wait for 1s, and then return to original position. And the results of the pan (or tilt) angle change are shown in fig. 2(A). For the zoom of a PTZ camera, we control the camera be directed to zoom purely for 0.5s, 1s, 1.5s, return to the original position after waiting for the same time. The result of the focal length change is displayed in fig. 2(C). To get the accurate result of each frame, A series of images are acquired before and after each motion. And cameras of these images can be calibrated by the camera calibration method mentioned in Sec. 5. The average relationships of the pan θ, the tilt ϕ, and the focal length f with time are illustrated in Fig. 2.

As evident from the graph of the first row, purely rotating the PTZ camera progress through three phases of acceleration, linearity and deceleration. In the second row, there is a one-to-one correspondence between focal length and time in purely zooming of the camera. Therefore, we devise a novel PTZ motion estimation model

$$\begin{aligned}\{\theta, \phi, f\} &= h\left(t \mid v_p, v_t, v_f, t_{p1}, t_{t1}\right) \\ &= \begin{cases} h_1\left(t \mid v_p, t_{p1}\right) \\ h_2\left(t \mid v_t, t_{t1}\right) \\ h_3\left(t \mid v_f\right) \end{cases}\end{aligned} \tag{2}$$

Where

$$h_1(t|v_p, t_{p1}) = \begin{cases} \frac{1}{2}v_p t^2, & (t < t_{p1}) \\ v_p t - \frac{1}{2}v_p t_{p1}, & (t_{p1} < t < t_{p2}) \\ \frac{1}{2}v_p(t + t_{p2} - t_{p1}) & \\ +\frac{1}{2}v_p(1-t)(t - t_{p1}), & (t > t_{p2}) \end{cases} \tag{3}$$

$$h_2(t|v_t, t_{t1}) = \begin{cases} \frac{1}{2}v_t t^2, & (t < t_{t1}) \\ v_t t - \frac{1}{2}v_t t_{t1}, & (t_{t1} < t < t_{t2}) \\ \frac{1}{2}v_t(t + t_{t2} - t_{t1}) & \\ +\frac{1}{2}v_t(1-t)(t - t_{t1}), & (t > t_{t2}) \end{cases} \tag{4}$$

$$h_3(t|v_f) = \frac{1}{2}v_f t^k + \sqrt[k]{2v_f(f_1 - f_0)}t + f_1 \tag{5}$$

v_p, v_t, v_f are constants of the motion estimation model and are different for each PTZ camera. t_1 and t_2 represent acceleration time and the sum of acceleration time and constant speed time determined by users. f_0 and f_1 describe focal length in $Zoom = 1$ and before zooming respectively.

IV. PTZ Calibaration Algorithm Based on Key-rays Collection

The motion estimation model can real-time calibrate the moving PTZ camera. However, it's unreasonable to expect results of the model to be accurate. We desire a dynamic correction method that ensures the camera calibration accurately. Our approach is to build a key-rays collection of the camera's environment, and calibrate online images using two-point methods.

A. Building key-rays collection of the scene

[15] proposes the idea using rays as landmarks for tracking and mapping, and we extend this idea as key-rays collection of the camera's scene.

We first build a full PTZ panoramic image, which consists of four stages. Stage 1 generates background images from the set of single view images [23]. Stage 2 calculates the camera parameters to determine the 3D coordinate for each single view image [23]. Meanwhile, three parameters (θ, ϕ, f) of each single view image are estimated. Stage 3 transforms the background images from all views into the panoramic image using spherical projection. Stage 4 fuses and blends the overlapping regions among background images from different perspectives on panorama [24].

The image location $p = (x, y)$ is projected by a ray $r = (\theta_p, \phi_p)$, so the label is given by:

$$\mathrm{r} = \left[\theta + \arctan\frac{x-u}{f}, \phi + \arctan\frac{y-v}{f}\right]^T \tag{6}$$

Where (θ, ϕ, f) represents the pan, the tilt and focal length of each single view image. (u, v) is the image center.

We extract all the SIFT features from each background image and transform image coordinates of features to rays. Then features from different images that highly overlap in both descriptor and ray position are merged, so that each feature appears only once in the collection. Finally, we store all

features in a collection as the key-rays collection of the scene. Figure 3 shows the pipeline of building key-rays collection of the scene.

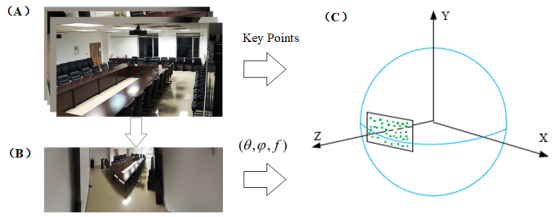

Fig. 3. Pipeline of building the key-rays collection. (A): a set of background images that are used to stitch the panorama and provide SIFT features; (B): the panorama image that provide the (θ, ϕ, f) of background images; (C) the correspondence between key-points and key-rays. We translate the image coordinates of features to the sphere coordinates.

B. PTZ calibration using two-point method

We capture a current image from the real-time video stream and get the $(\hat{\theta}_0, \hat{\phi}_0)$ of the current image by the motion estimation model. And we obtain the key-rays near the $(\hat{\theta}_0, \hat{\phi}_0)$ from the key-rays collection as a sub-collection L. We extract the set of SIFT features, called as S, acquired from the current image. And we try to match each feature in the S to the L, resulting in a set of putative matches M: $\{(x_i, y_i) \leftrightarrow (\theta_i, \phi_i), i = 1, \ldots, N\}$. The feature matches are computed using Brush-Force matching between SIFT descriptors.

We can obtain a extract guess for (θ'_0, ϕ'_0, f') by removing the outliers with three-parameter RANSAC [8], which computes (θ'_0, ϕ'_0, f') for randomly selected feature pairs with two-point method and evaluates the fit with

$$\theta'_0 + \arctan\frac{x_i - u}{f'_0} - \theta_i + \phi'_0 + \arctan\frac{y_i - v}{f'_0} - \phi_i < \varepsilon \quad (7)$$

This can effectively remove the mismatched feature pairs.

V. Experiments

We conducted experiments to evaluate the proposed model and algorithm. And all cameras used are Hikvision PTZ cameras. Our approach is implemented with C++ on an Inter ® core ™ i7-975H CPU, NVIDIA GeForce GTX 2070M graphics card, 16GB memory Windows system.

A. Motion estimation model error

We obtain moving images and motion estimation models of different cameras using the method mentioned in Sec. 3. And the ground truth camera parameters are manually calibrated. We first obtain background images of the camera's scene and the parameters using the method in Sec. 4. Then we obtain at least two correspondences between images and background images from human annotation, and transform background images' position to rays position. Third, we get cameras parameters using the two-point method as the ground truth. Table I shows errors between the motion estimation model and the ground truth of different cameras.

In Table I, we present the mean value standard errors of motion estimation models from different scenes. The rotational

TABLE I
The Motion Model Errors

PTZ Camera Type	Model Errors		
	Pan$(^\circ)$	***Tilt***$(^\circ)$	***Focal length***
DS-2DC42231W-D	0.00 ± 0.35	-0.14 ± 0.32	0.52 ± 167.85
DS-2DE7172-A	-0.30 ± 0.47	0.08 ± 0.33	-0.05 ± 67.54
DS-2DC5220IW-A	0.28 ± 0.35	0.07 ± 0.22	$0.52 \pm 0.162.45$

error of pan angle is about 0.6° and the rotate error of tilt angle is about 0.3°. Considering the view of the camera is generally 50° and the average rotation angle in rotating is 30°, the pan error and the tilt error are 2% and 1% respectively, which can be ignored in rotating highly. The focal length error of model is about 150 and the range of the focal length tested is $[1200, 12000]$. The range of error percent is $[1\%, 10\%]$ and the percent decreases with focal length. The average velocity of focal length is $5000/s$, since the focal length error can be ignore in zooming. Table I demonstrates the accuracy of the motion estimation model.

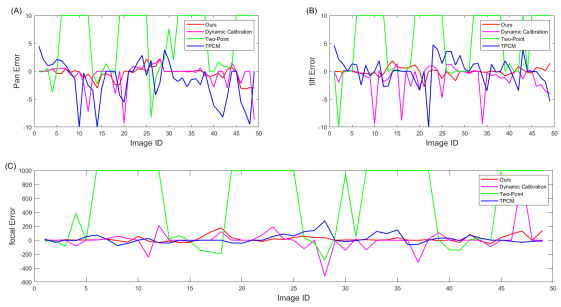

Fig. 4. Results of several PTZ calibration methods. (A) The error in the pan angle θ. (B) The error in The Tilt angle ϕ. (C) The error in the focal length f.

B. Results on PTZ calibration based on key-rays collection

We test our approach at the scene indoor1 in Table II. We collect 50 images from different values of pan, tilt and zoom as the test dataset. The reference images are from the background dataset in Sec. 4.

Baselines: We compare our approach against several PTZ camera calibration methods which can directly compute the parameters of the PTZ camera by feature point pairs: the dynamic calibration [14], two-point method [7], and the TPCM method [16]. Because these three methods depend on accurate feature correspondence, we first remove the obviously wrong mismatched point pairs with RANSAC.

Fig. 4 shows the errors between those methods and the ground-truth established by manual checking. The reported error for each estimated parameter is computed as $|param_{est} - param_{ground-truth}|$. From the results, we see that our PTZ calibration method outperforms for all the parameters and has

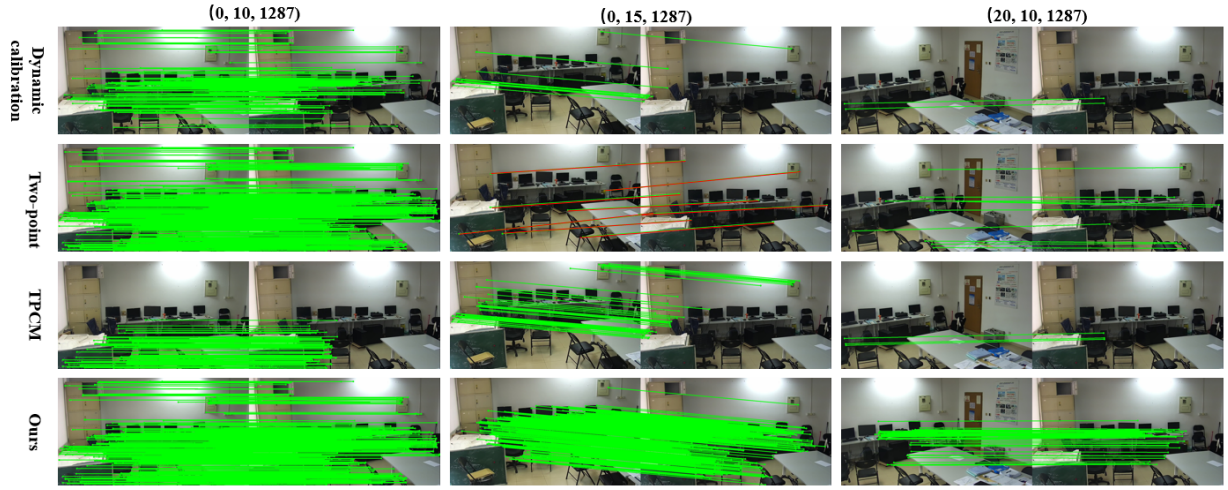

Fig. 5. Qualitative results of several PTZ calibration methods. (0,10,1287) means that the pan angle of test image is 0 degree, the tilt angle of test image is 10 degree, and the focal length of test image is 1287.

perpendicular errors. The two-point method has a large error, which is because the random forest depends on images trained. The more images trained, the better the predict result of the random forest. We only use the background dataset to train the random forest, so the predict result has a high error. Because the TPCM method is implemented via many additions and multiplications which enlarge the errors, leading to results of the TPCM have great fluctuations.

We also provide qualitative comparison on these 50 images with their background images, as shown in Fig. 5. We use poses estimated by four methods to remove the mismatched pairs. In pose (0,10,1287), the test image shift is small relative to the background image and four methods obtain relatively good performance. In pose (0,15,1287) and (20,10,1287), the test image shift is larger.From the results, our method always outperformed the other methods. The two-point method get a wrong result in pose (0,15,1287), which show that the predict result from the random forest is wrong. This conclusion is consistent with the result of quantification above.

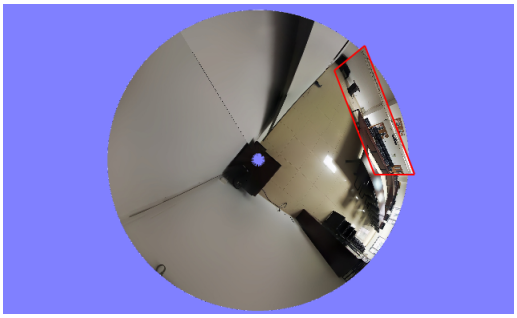

Fig. 6. The half-sphere model and the video projection. Red box indicates the video projection and the circle is the view that users watch the model from above. The blue part indicates the virtual space.

TABLE II
THE INFORMATION OF REAL SCENES

PTZ Camera Type	**Scene**	**Scene Size** ($l \times w \times h$)
DS-2DC4223IW-D	Indoor1	$8m \times 6m \times 3m$
DS-2DE7172-A	Indoor2	$12m \times 8m \times 3m$
DS-2DC5220IW-A	Outdoor	$50m \times 50m \times 50m$

C. Experimental validation of real scenes

For the experimental validation of the proposed method, we set up three real scenes with a PTZ camera mounted (see Table II). The first scene Indoor1 is a $8m \times 6m \times 3m$ laboratory with a DS-2DC4223IW-D PTZ camera, which has chairs and tables. The second scene Indoor2 is a $12m \times 8m \times 3m$ meeting room with a DS-2DE7172-A PTZ camera, which has less features. The third scene Outdoor is an $50m \times 50m \times 50m$ outdoor scene which has pedestrians and vehicles, mounting a DS-2DC4223IW-D PTZ camera. In different virtual scene models which is a half-sphere model built by panorama images mentioned in Section 4, we test our PTZ calibration method and the motion estimation model at different zoom values. The half-sphere model is shown in Fig. 6. All results are shown in Fig. 7.

In Fig. 7, we present the result of the proposed method in three real scenes. We observe that our approach still remains well on less texture region from the first row. In Outdoor, more features mean more distractors. However, our method obtains a good outcome which demonstrates the applicability of our method both indoors and outdoors. In the third row, we try to prove the robustness for movable objects such as chairs and tables.

VI. CONCLUSION

We proposed a motion estimation model for PTZ camera and PTZ calibration method based on key-rays collection.

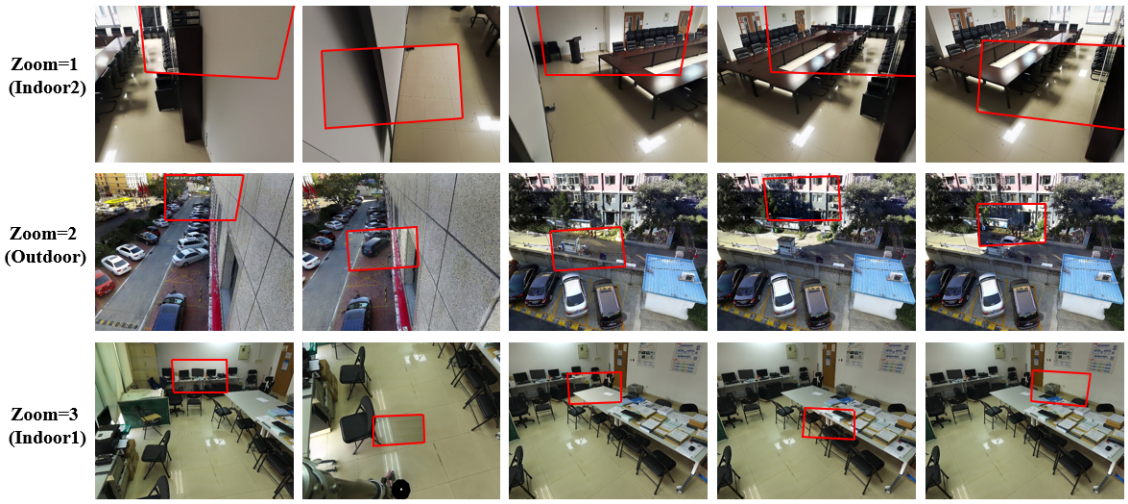

Fig. 7. The result of the proposed PTZ calibration method in real scenes. Red box indicates the real-time video stream from the PTZ camera in real scene. The part outside the red box is from the part view of the half-sphere model.

Firstly, we observed the moving result of the PTZ camera and designed a novel motion estimation model. Furthermore, we presented a robust PTZ calibration based on key-rays collection, using the two-point method. Experiments in different cameras and scenes show that the PTZ calibration method is effective and accurate. The motion estimation model can real-time estimate the poses of a PTZ camera in moving.

However, there are several limitations of this work. We have not considered the lens distortion which carries some level of impact when calibrating with accurate 3D models. Also, we suppose that the principal point coincides with the projection center and zooming center. However, this may not be the truth for cameras in zooming. Finally, while we assumed that the optical center of the PTZ camera is fixed, in practice it will change when the camera rotate. Those factors need to be taken into account if we obtain a more accurate result.

REFERENCES

[1] S. Chen, C. Lee, C. Lin, I. Chen, Y. Chen, S. Shih, and Y. Hung, "2D and 3D Visualization with Dual-Resolution for Surveilance." IEEE Computer Society Conference on Computer Vision & Pattern Recognition Workshops IEEE, 2012:23-30.

[2] J. Young, T. Langlotz, S. Mills, and H. Regenbrecht. "Mobileportation: Nomadic Telepresence for Mobile Devices." Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies 4.2(2020):1-16.

[3] Sportlogiq. Http://sportlogiq.com/. september 11, 2020.

[4] R. A. Sharma, B. Bhat, V. Gandhi, and C. Jawahar. "Automated top view registration of broadcast football videos." IEEE, 2018.

[5] D. G. Lowe. "Distinctive image features from scale-invariant keypoints." International Journal of Computer Vision 60.2(2004):91-110.

[6] F. Pece, et al. "PanoInserts: Mobile Spatial Teleconferencing." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013.

[7] J. Chen, F. Zhu, J. L. James. "a two-point Method for PTZ Camera Calibration in Sports." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.

[8] M.A. Fischler, and C. Robert. "Random Sample Consensus: A Paradigm for Model Fitting with Apphcatlons to Image Analysis and Automated Cartography." Communications of the ACM 24.6 (1981): 381-395.

[9] N. Homayounfar, F. Sanja, and U. Raquel. "Sports Field Localization via Deep Structured Models." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[10] S. Liu, J. Chen. C. Chang, Y. A. "A New Accurate and Fast Homography Computation Algorithm for Sports and Traffic Video Analysis." IEEE Transactions on Circuits and Systems for Video Technology 28. 10(2017):2993-3006.

[11] Chen J, Little JJ. "Sports camera calibration via synthetic data." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.

[12] Citraro, Leonardo, et al. "Real-time camera pose estimation for sports fields." Machine Vision and Applications 31.3 (2020): 1-13.

[13] Sha, Long, et al. "End-to-end camera calibration for broadcast videos." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[14] Z. Wu and R. J. Radke. "Keeping a pan-tilt-zoom camera calibrated." IEEE transactions on pattern analysis and mechine intelligence 35.8(2012): 1994-2007.

[15] J. Lu, J. Chen, and J. L. James. "Pan-tilt-zoom SLAM for Sports Videos." arXiv preprint arXiv:1907.08816(2019).

[16] An, Pei, et al. "Two-point calibration method for a zoom camera with an approximate focal-invariant radial distortion model." JOSA A 38.4 (2021): 504-514.

[17] H. S. Sawhney, et al. "Video Fashlights: real time rendering of multiple videos for immersive model visualization." Proceedings of the 13th Eurographics workshop on Rendering, 2002, pp. 157-168.

[18] S. Chen, et al. "2D and 3D Visualization with Dual-Resolution for Surveillance." 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2012.

[19] Y. Zhou, M. Meng, W. Wu, and Z. Zhou. "Virtual-reality video fusion system based on video model." Journal of System Simulation 30.07 (2018): 133-140.

[20] J. Tompkin, F. Pece, R. Shah, S. Izadi, J. Kautz, and C. Theobalt. "Video collections in panoramic contexts." Proceedings of the 26th annual ACM symposium on User interface software and technology. 2013.

[21] J. Young, T. Langlotz, M. Cook, and H. Regenbrecht. "Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices." IEEE transactions on visualization and computer graphics 25.5 (2019):1908-1918.

[22] G. Thomas. "Real-time camera tracking using sports pitch markings." Journal of Real-Time Image Processing, 2.2-3(2007):117–132.

[23] H. Yong, et al. "Panoramic Background Image Generation for PTZ Cameras." IEEE Transactions on Image Processing 28.7 (2019):3162-3176.

[24] M. Brown, and D. G. Lowe. "Automatic Panoramic Image Stitching using Invariant Features." International journal of computer vision 74.1 (2007): 59-73.