Visual Expansion and Real-time Calibration for Pan-tilt-zoom Cameras Assisted by Panoramic Models

Liangliang Cai[1] and Zhong Zhou[1]

[1]State Key Laboratory of Virtual Reality Technology and Systems

August 25, 2024

Abstract

Pan-tilt-zoom (PTZ) cameras, which dynamically adjust their field of view (FOV), are pervasive in large-scale scenes, such as train stations, squares, and airports. In real scenarios, PTZ cameras are required to quickly make decisions informed about where to direct its focus through contextual cues from the surrounding environment. To achieve this goal, some researches project camera videos into three-dimensional (3D) models or panoramas and allow operators to perceive spatial relationships. However, these works face several challenges in terms of real-time processing, localization accuracy, and realistic reference. To address this problem, we propose a visual expansion and real-time calibration for PTZ cameras assisted by panoramic models. We attempt to meet the demand for real-time processing with a motion estimation model for a PTZ camera, to improve calibration performance of PTZ images with only two feature point pairs, and to provide a realistic environmental context through a panoramic model. We verify our methods on both public and our self-built test scene. It can be seen from the experimental results that our method can exhibit impressive accuracy and efficiency.

Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Visual Expansion and Real-time Calibration for Pan-tilt-zoom Cameras Assisted by Panoramic Models

Liangliang Cai[1] | **Zhong Zhou**[1,2]

[1]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China

[2]Department of Mathematics and Basic Research, Zhongguancun Laboratory, Beijing, China

Correspondence

Zhong Zhou, XueYuan Road No.37, HaiDian District, Beijing, 100191, China.
Email: zz@buaa.edu.cn

Funding Information

This research was supported by the Science and Technology Project of Hainan Provincial Department of Transportation, Grant/Award Number: HNJTT- KXC-2024-3-22-02; National Natural Science Foundation of China, Grant/Award Number: 62272018.

Abstract

Pan-tilt-zoom (PTZ) cameras, which dynamically adjust their field of view (FOV), are pervasive in large-scale scenes, such as train stations, squares, and airports. In real scenarios, PTZ cameras are required to quickly make decisions informed about where to direct its focus through contextual cues from the surrounding environment. To achieve this goal, some researches project camera videos into three-dimensional (3D) models or panoramas and allow operators to perceive spatial relationships. However, these works face several challenges in terms of real-time processing, localization accuracy, and realistic reference. To address this problem, we propose a visual expansion and real-time calibration for PTZ cameras assisted by panoramic models. We attempt to meet the demand for real-time processing with a motion estimation model for a PTZ camera, to improve calibration performance of PTZ images with only two feature point pairs, and to provide a realistic environmental context through a panoramic model. We verify our methods on both public and our self-built test scene. It can be seen from the experimental results that our method can exhibit impressive accuracy and efficiency.

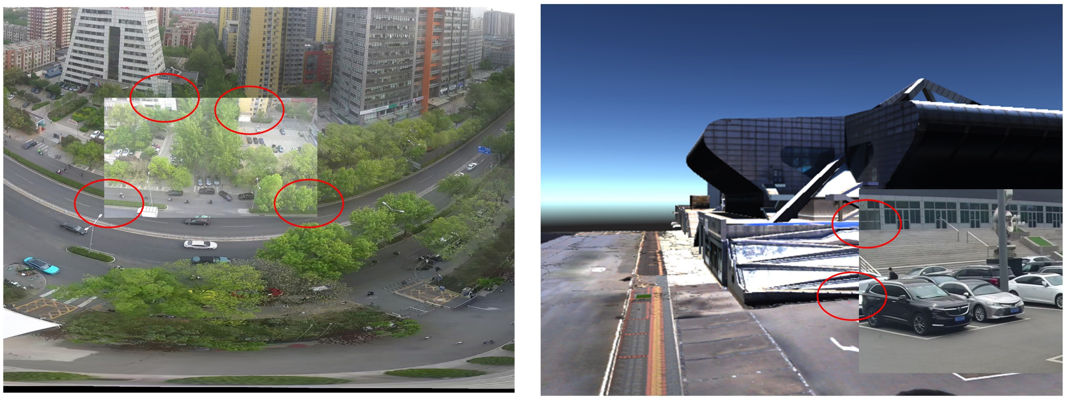
KEYWORDS

augmented virtual environment, PTZ camera, camera calibration, panorama, key-ray collection

1 | INTRODUCTION

Pan-tilt-zoom (PTZ) cameras offer the capability for extensive surveillance of the surrounding environment through flexible control, and have been broadly implemented in various scenes, such as schools, universities, companies, stadiums, and so on. Due to the wide array of emergencies that occur in real-world scenarios, the response time of PTZ cameras is essential. To achieve this goal, numerous studies about augmented virtual environment technologies [1,2,3,4] have introduced background models as environmental cues to guide the operation of PTZ cameras, such as three-dimensional (3D) models or panoramas. These methods are real-time image projection technologies in a virtual environment for painting realistic-looking textures on reference models. However, the majority of studies suffers from misalignments between images as well as backgrounds and lack of realistic reference, as shown in Fig. 1.

To solve these problems, many researches use tilt photography models and improve camera calibration algorithms. The camera calibration is the process that estimates the interior and exterior parameters of the camera and determines the orientation and position of the camera relative to the reference model [5,6,7]. At present, camera calibration methods are categorized into pattern-based calibration [8,9,10], infrastructure-based calibration [11,12,13], and self-calibration [14,15,16,17,18,19]. The pattern-based methods leverage the regularity and symmetry of the patterns to facilitate precise measurements and accurate camera parameter estimation. Due to the requirement of using specific calibration patterns, they are difficult to be installed in outdoor scenes. The infrastructure-based approaches typically involve building a controlled environment with known reference points or structures that the cameras can use to calibrate their settings. But the construction and installation of the infrastructure often demand high costs of labor. The self-calibration methods estimate camera intrinsic and extrinsic parameters by generating a multitude of point correspondences



(a) Camera calibration assisted by the panoramic image

(b) Camera calibration assisted by the 3D model

FIGURE 1 Misalignments and lack of realistic reference suffered in camera calibrations. The areas enclosed in red indicate that embedded images do not align properly with the background. The distortion of the panorama and the unreal textures of the 3D model cause the lack of reality of the background model.

between adjacent images. However, the accuracy of these methods can be affected by illumination conditions, moving objects, feature density, and so on.

In this paper, we present a visual expansion and real-time calibration for PTZ cameras assisted by panoramic models, integrating the advantages of both the infrastructure-based approach and the self-calibration method. The method also considers the geometric property of the PTZ camera, shown in Fig. 2. We first collect scene images from the PTZ camera and remove the moving objects to acquire pure background images. Then we stitch these images utilizing the image mosaic method, generating the panorama of the current scene. In order to improve the quality of image mosaic, we strengthen image mosaic by optimizing the two phases of feature matching and parameter estimation. We eventually map the panorama into a half-sphere model to construct a panoramic model. Also, we propose a real-time PTZ camera calibration algorithm. The algorithm mainly comprises the motion estimation model and the camera calibration algorithm based on key-ray collection. The motion estimation model of a PTZ camera is derived from the parameter variation in the camera's motion, and the model can rapidly estimate the camera's pose while the PTZ camera is in motion. The camera calibration algorithm based on key-ray collection can realize high calibration preformance of the PTZ camera in panoramic models. In order to further promote the robustness of PTZ camera calibration, we propose the two-ray method for PTZ camera calibration. It only takes two pairs of matched points between adjacent images to estimate the pose of the PTZ camera, which will play a significant role in extreme weather. This is an important improvement over previous methods.

In contrast to traditional PTZ camera calibrations and augmented virtual environment technologies, we introduce the panoramic model to provide a realistic environmental context with low costs of labor. The panoramic model assists the operator in determining the direction and zooming of the camera. In addition, we concentrate on the PTZ camera calibration algorithm only using two feature pairs. Compared to other PTZ camera calibration methods, it implements fewer feature pairs and achieves the higher calibration accuracy. Also, the motion estimation model for the PTZ camera meets the demand for real-time processing with less resource usage. Our contribution can be summarized as:

- A two-ray method for PTZ camera calibration is developed from 2D-3D correspondences to 2D-2D correspondences. The method can improve pose accuracy of a PTZ camera relying on two adjacent PTZ images.
- A panoramic model construction based on PTZ camera model is presented to construct a 3D scene background model. We optimize the two stages of feature matching and parameter estimation for image stitching via the PTZ camera model. It can calculate more accurate parameters of current camera and improve the accuracy of image stitching.
- A state-of-art and real-time PTZ camera calibration algorithm is proposed. The method is mainly composed of the motion estimation model and the camera calibration algorithm based on key-ray collection. It can register the camera frames into the panorama model in real time.
- A complete framework of a visual expansion and real-time calibration for PTZ cameras assisted by panoramic models is designed to reduce the burden of understanding the spatial relation and to manipulate the PTZ camera for operators.

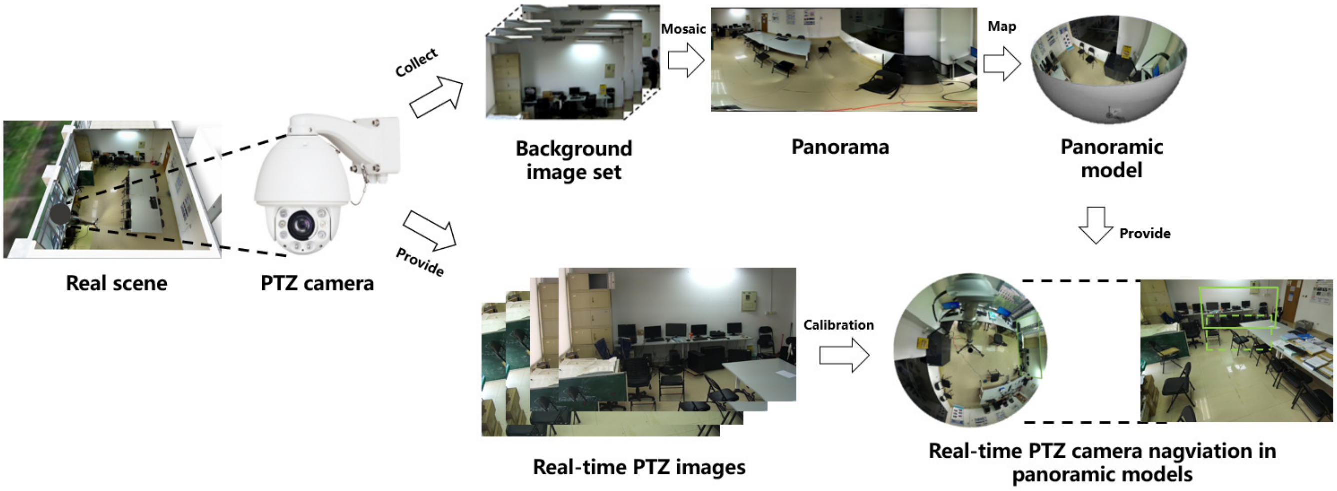


FIGURE 2 Visual expansion and real-time calibration for PTZ cameras assisted by panoramic models. In real scenes, PTZ cameras capture background images for background image sets, which are used to stitch pamoramas. Then we construct panorama models using panorama images generated. Finally we register PTZ camera frames into panorama models by the PTZ camera calibration in real time.

The remainder of this paper is organized as follows. Section 2 reviews some related works and Section 3 presents the two-ray method for PTZ camera calibration. Section 4 introduces the panoramic model construction. Section 5 presents our method for PTZ camera caliration in detail in real time. Section 6 shows our results of methods proposed and Section 7 concludes the paper.

2 | RELATED WORK

2.1 | Camera calibration

The pattern-based calibration, the infrastructure-based calibration, and the self-calibration make up the fixed/portable camera calibration categories. The pattern-based calibration estimates camera parameters making use of unique calibration patterns, such as checkerboards. The pattern-based method is generally utilized to estimate internal parameters of cameras and depends on specific calibration patterns that are difficult to be installed in outdoor scenes. Chen et al.[7] established constraint equations by correlating image matched points before and after preknown camera motions. The preknown camera motions are special patterns. The infrastructure-based calibration uses point-cloud models or tilt photography models as infrastructures, establishing a correlation between images and infrastructures to estimate camera parameters. Campbell et al.[11] proposed a robust and globally-optimal inlier set maximisation approach that jointly estimates the optimal camera pose, taking into account the identification of cross-modality correspondences between 2D image points and a 3D point-set. The infrastructure-based method can precisely estimate both intrinsic and extrinsic parameters of cameras, but the construction of infrastructures requires high costs of labor. In some scenarios, such as soccer fields, basketball courts, or hockey arenas, researchers commonly supplant sport field models simplified, decreasing labor costs, to point clouds or tilt photography models[20,21,22,23,24,25]. The self-calibration method generates a large number of point correspondences between adjacent images for the purpose of estimating the internal and external camera parameters. The theory of camera self-calibration, which requires only point matches from image sequences, was first proposed by Faugeras[26]. Song De Ma[14] achieved camera calibration based on a sequence of specifically designed motion by the active vision system. Q.-T. Luong and O.D. Faugeras[15] recovered the internal orientation of the uncalibrated moving camera using point correspondences between three images and the fundamental matrices computed from these point correspondences. Vasconcelos et al.[16] proposed an automatic camera calibration to stuff an uncalibrated node into a network of calibrated cameras using only pairwise points. Liu et al.[17] present a novel homography computational algorithm that can increase the precision of homography computation and decrease processing time. Considering the effect of radial distortion, An et al.[18] proposed a novel two-point calibration method (TPCM) to estimate the focal length and 3-DoF rotation matrix with only two control points from one image. In such restrictive cases with limited contiguous regions between images, these self-calibration approaches offer substantial advantages over the other two methods. However, most of they must fulfill the requirements that there are four feature pairs in adjacent images. Otherwise, these self-calibration methods would be unworkable. Additionally, the precision of the self-calibration method also falls under the influence of illumination, occlusion, and lack of texture.

Different from the fixed/handheld camera calibration, the PTZ camera calibration has its own characteristics. In general, the optical center and geometric center of a PTZ camera are supposed to be known. Thus, the location of one single view image depends on its pan, tilt and zooming coordinates, which often can be obtained from the camera. Considering these characteristics of PTZ cameras, a number of researchers calibrate PTZ cameras utilizing particular geometries between adjacent PTZ images. Wu et al.[27] present a dynamic calibration algorithm based on aligning the current image against a collection of offline-stored features. This method directly estimates the intrinsics and extrinsics of the PTZ camera. However, this technique is only appropriate for tiny angle and focal length offsets. Chen et al.[28] proposed a two-point method requiring only two point correspondences to calibrate a PTZ camera and a rapid random forest method to predict pan-tilt angles without matching image-to-image features.Further, they presented an online SLAM system based on the two-point method and PT random forest that can track PTZ cameras in highly dynamic sporting activities[29]. Chen et al.[30] employ a siamese network to excavate compact deep features and use a novel two-GAN model to detect field marking in real images. These techniques have substantial room for development in terms of precision and robustness.

2.2 | Image mosaicing

Extensive researches have been conducted on Image Mosaic techniques, which stitch multiple images into an image to gain a broader visual perception of a large-scale scene. M. Brown and D. G. Lowe[31] first proposed a comprehensive framework for mosaicing images using invariant local features, such as SIFT (scale-invariant feature transform), to extract feature matches between images. Meanwhile, bundle adjustment, which is a photogrammetric technique to combine multiple images of the same scene into an accurate 3D reconstruction, has also been applied to stitch images. The idea behind multi-band blending is to blend low-frequency bands over a large spatial range and high-frequency bands over a short range to compensate for exposure differences and misalignments. Gao et al.[32] described a method for constructing a seamless stitching image of a panoramic scene with two predominate planes: a distant back plane and a ground plane that sweeps out from the location of the camera. Zaragoza et al.[33] seamlessly bridges image regions that were inconsistent using moving direct linear transformation (Moving DLT). The method was intended to be globally projective while permitting local non-projective deviations to account for violations of the presumed imaging conditions. Li et al.[34] proposed a parallax-tolerant method for image stitching based on robust elastic warping that simultaneously achieved precise alignment and efficient processing. Li et al.[35] proposed a local-adaptive image alignment method based on triangular facet approximation, which directly manipulated the corresponding data in the camera coordinates, thereby enhancing performance of imaging models of cameras. Yong et al.[36] proposed a fast multi-band blending method to improve the efficiency in panoramic image fusion and mosaicing. The approach proposed exhibited impressive efficiency in PTZ panorama generation as well as panoramic image mosaicing. The majority of research on image mosaic focuses primarily on public datasets, but very few works have been reported on panoramic image generation for PTZ cameras.

2.3 | The augmented virtual environment technology

The augmented virtual environment (AVE) technology is a real-time image projection technology in a virtual environment for painting realistic-looking textures on 3D models. Sawhney et al.[1] first present a video flashlight system that illuminates a static 3D graphics model with live video textures captured by stationary and moving cameras. Chen et al.[37] propose a novel visualization framework for surveillance systems that projects the large-scale display area with a fixed low-resolution camera and the fovea region with a second high-resolution camera. Pece et al.[38] demonstrate a PanoInserts system that inserts video within the panorama using a combination of marker- and image-based tracking. Tompkin et al.[2] create a video-collections+context interface by embedding videos into a panorama. In this work, a spatio-temporal index and instruments for rapid exploration of the video collection's space and time are developed. Zhou et al.[39] propose a novel virtual-real video fusion system based on video Models that exploits the single-image modelling technology. Young et al.[4] introduce 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, and the environment can be viewed by a remote user who can independently choose the viewing direction.

When these methods enhance the virtual environment with images, we observe that these images also express more content through the virtual environment.

3 | TWO-RAY METHOD FOR PTZ CAMERA CALIBRATION

Inspired by Chen's work[28], we extend the two-point method from 2D-3D correspondences to 2D-2D correspondences. This modified method is called as the two-ray method. We first establish the transformation between the rays and the image pixels, converting the pixels on the image into rays. Then, according to the rays' connection in the overlapping region of adjacent images, we calculate the pose of the PTZ camera.

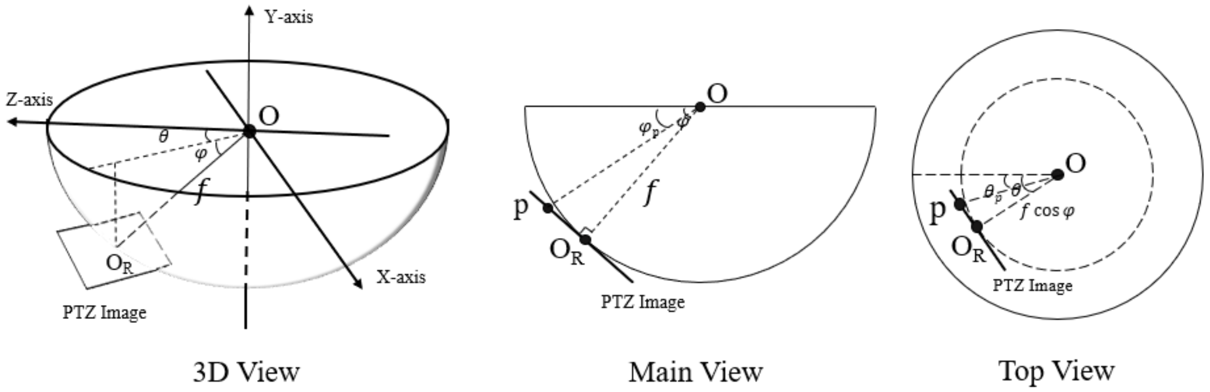


FIGURE 3 The geometric relationship between the PTZ image and the PTZ camera. The main view is obtained from the direction along the negative half axis of the X-axis. And the top view is obtained from the direction along the negative half-axis of the Z-axis.

It is common knowledge that any image from a PTZ camera satisfies the geometric property shown in the 3D view of Fig. 3. The **O** is the optical center of the PTZ camera, and (θ, φ, f) denotes the orientation and focal length of the camera. $\mathbf{O}_R(x_{OR}, y_{OR})$ (the image coordinate system) is the center of the PTZ image. The PTZ image is tangent to a sphere of radius f. For any point $\mathbf{p}(x_p, y_p)$ (image coordinate system) on the image, the orientation of the corresponding ray $\mathbf{r}_p$ is (θ_p, φ_p). According to the main view and the top view of Fig. 3, we can obtain:

$$\theta_p = arctan(\frac{x_p - x_{OR}}{fcos(\varphi)}) + \theta \tag{1}$$

$$\varphi_p = arctan(\frac{y_p - y_{OR}}{f}) + \varphi \tag{2}$$

We further analyze the spatial property between adjacent PTZ images, as shown in Fig. 4. Both the reference image and the target image are from the PTZ camera, and they are adjacent and have overlapping regions. We assume that the $(\theta_r, \varphi_r, f_r)$ of the reference image is known, and the focal length f_t of the target image is not consistent with that of the reference image. We randomly choose two points $\mathbf{p}_1$ and $\mathbf{p}_2$ from the overlapping region. The Angle between ray $\mathbf{r}_{p1}(\theta_{p1}, \varphi_{p1})$ and ray $\mathbf{r}_{p2}(\theta_{p2}, \varphi_{p2})$ is α. There exists:

$$\begin{aligned} cos(\alpha) &= cos(\theta_{p1})cos(\varphi_{p1})cos(\theta_{p2})cos(\varphi_{p2}) + sin(\theta_{p1})cos(\varphi_{p1})sin(\theta_{p2})cos(\varphi_{p2}) + sin(\varphi_{p1})sin(\varphi_{p2}) \\ &= \frac{(K^{-1}\mathbf{p}_1)^T(K^{-1}\mathbf{p}_2)}{\sqrt{(K^{-1}\mathbf{p}_1)^T(K^{-1}\mathbf{p}_1)}\sqrt{(K^{-1}\mathbf{p}_2)^T(K^{-1}\mathbf{p}_2)}} \end{aligned} \tag{3}$$

where the K is the internal matrix of the PTZ camera. In the reference image, the image coordinates of $\mathbf{p}_1$ and $\mathbf{p}_2$ are represented as $(x_{r1}, y_{r1}, 1)^T$ and $(x_{r2}, y_{r2}, 1)^T$, respectively. We can calculate $\mathbf{r}_{p1}$ and $\mathbf{r}_{p2}$ using the formula 1 and 2. In the target image, the image coordinates of $\mathbf{p}_1$ and $\mathbf{p}_2$ are denoted as $(x_{t1}, y_{t1}, 1)^T$ and $(x_{t2}, y_{t2}, 1)^T$, respectively. We assume PTZ images are undistorted and their sizes are known. So the focal length f_t is the only unknown element in K_t of the target image. We can calculate the f_t by Eq.3.

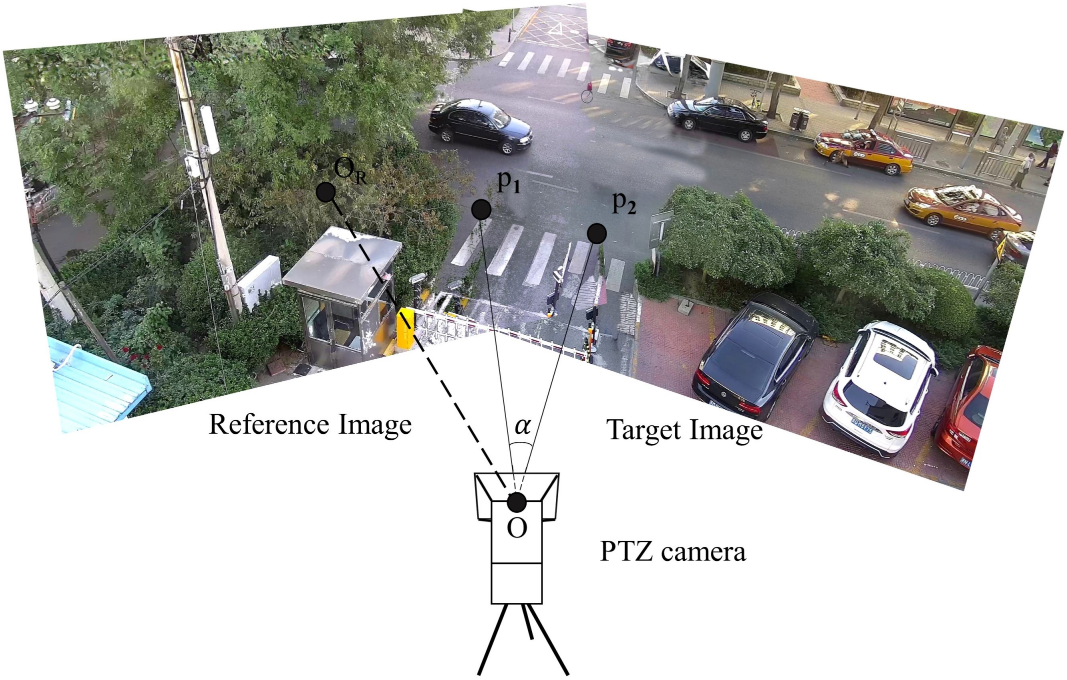


FIGURE 4 The relationship between PTZ adjacent images.

We attempt to find the association between the (θ, φ) of $\mathbf{r}_p$ and $\mathbf{p}$. Any point $\mathbf{p}$ in the target image satisfies:

$$K^{-1}\mathbf{p} = \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \mathbf{R}_\varphi \mathbf{R}_\theta \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4}$$

where $[XYZ]^T$ is the coordinates of 3D point corresponding to $\mathbf{p}$. We have:

$$\frac{X}{\sqrt{X^2 + Y^2 + Z^2}} = \sin\theta_p \cos\varphi_p \tag{5}$$

$$\frac{Y}{\sqrt{X^2 + Y^2 + Z^2}} = \sin\varphi_p \tag{6}$$

$$\frac{Z}{\sqrt{X^2 + Y^2 + Z^2}} = \cos\theta_p \cos\varphi_p \tag{7}$$

We separate $\sin\theta$ and $\cos\theta$ from the formula 4, which causes:

$$\begin{bmatrix} A_\varphi & B_\varphi \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{8}$$

where

$$A_\varphi = \begin{bmatrix} -XU\cos\varphi - Z & X - ZU\cos\varphi \\ X\sin\varphi - XV\cos\varphi & Z\sin\varphi - ZV\cos\varphi \end{bmatrix}$$

$$B_\varphi = \begin{bmatrix} YU\sin\varphi \\ Y\cos\varphi + YV\sin\varphi \end{bmatrix}$$

.

From the formula 8, we have:

$$\cos\theta = \frac{-YZU + XY(\cos\varphi + V\sin\varphi)}{\det A_{\varphi}} \tag{9}$$

$$\sin\theta = \frac{XYU + YZ(\cos\varphi + V\sin\varphi)}{\det A_{\varphi}} \tag{10}$$

where $\det A_{\varphi} = (X^2 + Z^2)(V\cos\varphi - \sin\varphi)$. Because $\sin^2\theta + \cos^2\theta = 1$, we have a quadratic equation of $\tan\theta$ by

$$a\tan^2\varphi + b\tan^2\varphi + c = 0 \tag{11}$$

where

$$\begin{aligned} a &= \sin^2\varphi_p(U^2 + V^2) - \cos^2\varphi_p \\ b &= 2V \\ c &= \sin^2\varphi_p(1 + U^2) - \cos^2\varphi_p V^2 \end{aligned}$$

From the formula 11, we can calculate φ up to 2 solutions. We can eliminate one of them by setting the valid range to $(0°, 90°)$. The θ can be calculated from the formula 9 and 10. Consequently, we can estimate the (θ, φ, f) of a specific PTZ image using only two rays.

4 | PANORAMIC MODEL CONSTRUCTION BASED ON PTZ CAMERA MODEL

The principle of panoramic model construction is to stitch multiple images into a panorama. The primary process of image mosaicing includes feature extraction and matching, parameter estimation, projection transformation, optimal seam, and image fusion. Image Mosaicing frequently creates the problem of ghosts and discontinuity. Our analysis shows that inaccurate parameter estimation is an essential cause of ghosts and discontinuity. Therefore, we first optimize the feature matching and parameter estimation stages to improve the performance of parameter estimation. Then we stitch the images using the parameter estimation method with modifications and generate the panoramic model.

Most of researches on feature matching rely on the distance between feature descriptors, which leads to numerous false matches. Some studies such as the work of Ma et al.[40] utilize the local neighborhood structures of those potential true matches to heighten the accuracy of feature matching. No prior information is available for these analyses. We make an effort to enhance the accuracy of feature matching by using preset information. The transformations between two PTZ images can be considered as a unique homography matrix, since all PTZ images share a common optical center. The homography matrix between the image I_i and the image I_j is

$$H_{ij} = K_i R_i R_j^{\mathrm{T}} K_j^{-1} \tag{12}$$

where $K_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $R_i = R_{\varphi i} R_{\theta i}$. The value of R_i is equal to the formula 4. Some PTZ cameras provide the camera parameters of the current view, such as the pan angle θ, the tilt angle φ and the focal length f, but these values are often imprecise caused by mechanical drifts. We utilize the preset information to get the homography matrix $\hat{H}_{ij}$. We have

$$p_{jk} = \underset{p_{jm} \in N(\hat{H}_{ij} p_{ik})}{\arg\min} \left\| D_{p_{jm}} - D_{p_{ik}} \right\|_2 \tag{13}$$

where p_{ik}denotes the k-th feature point in the image I_i. $N(\hat{H}_{ij} p_{ik})$ is the set of feature points in the neighborhood of the projection of p_{ik} onto the image I_j. The radius of the neighborhood is set to 60 pixels. D_p denotes the descriptor of the feature point p. $\| * \|_2$ is the euclidean distance.

4.0.0.1 | *Parameter estimation:*

Currently, several methods of parameter estimation first roughly estimate parameters and then optimize these parameters using the bundle adjustment(BA) method. The BA method can achieve excellent results in 2D-3D correspondence circumstances. Due to the lack of a stable reference, the BA method in the 2D-2D correspondence circumstance mostly results in an immense drift between the estimation and the reality. Moreover, the BA method adjusts by the gradient and ignores the geometric property among the PTZ images. We propose an innovative and more accurate parameter estimation method for PTZ camera images.

We select an image from the background image set $S = \{I_1, I_2, ..., I_n\}$ as the reference image and assume that the preset parameters of the reference image are authentic. Then we divide S into the calibrated set S_c and the uncalibrated set $S_u = S - S_c$. In the initial state, only the reference image is in the S_c. We select an image I_i from the S_c. $S_u(I_i)$ denotes the adjacent image set of I_i in the S_u. We choose an image I_j from $S_u(I_i)$, extract SIFT features[41] from image I_i and I_j, and perform the feature matching using the formula 13. We receive the matched-point pair set $M = \{(\mathbf{p}_{ik}, \mathbf{p}_{jk})\}_{k=1}^{N}$. According to Sec. 3, $\mathbf{p}_{ik}$ can be converted to $\mathbf{r}_{ik}$ by the formula 1 and 2. We get the set of rays and point $M_r = \{(\mathbf{r}_{ik}, \mathbf{p}_{jk})\}_{k=1}^{N}$. We can estimate the f_j of image I_j by

$$f_j = \frac{\sum_{k_1=1}^{N} \sum_{k_2=k_1+1}^{N} f(r_{ik_1}, r_{ik_2}, p_{jk_1}, p_{jk_2})}{C_N^2} \tag{14}$$

where $f(*)$ is the focal length estimated by the formula 3, C_N^2 denotes the combination of two pairs. We can convert M_r to $\phi = \{(\theta_{jk}, \varphi_{jk})\}_{k=1}^{N}$ by the formula 9 and 11. We can estimate the θ_j and φ_j of image I_j by

$$(\theta_j, \varphi_j) = \underset{(\theta_{jk}, \varphi_{jk}) \in \Phi}{\arg\min} \sum_{m=1}^{M} \| H_{ij}(\theta_{jk}, \varphi_{jk}, f_j) p_{im} - p_{jm} \|_2 \tag{15}$$

We estimate the $(\theta_j, \varphi_j, f_j)$ of image I_j by the formula 14 and 15. Then we add I_j into S_c, and iterate the above steps util S_u is empty.

4.0.0.2 | *panoramic model construction:*

The initial orientation of the PTZ camera is $\theta = 0°$ and $\varphi = 10°$. We set φ to 10°, 30°, 50°, and 70° , respectively. We collect 36 images as the background image set with a horizontal interval of 40°. We utilize the image mosaic method improved to stitch the background images into the panorama. We project the panorama into the inner surface of the hemispherical model, which generates the panoramic model. The panoramic model is shown in the figure 11 (b).

5 | REAL-TIME PTZ CAMERA CALIBRATION ALGORITHM

We need to update PTZ camera poses frequently to achieve PTZ camera real-time calibration. On account of the image blur and illumination change caused by the camera motion, the feature-based method may fail. We determine that the camera motion is traceable and propose a novel real-time calibration algorithm combining a motion estimation model and a ray-based method. We first collect motion trajectories of horizontal movement, vertical movement, and zoom movement. Then we formulate the motion estimation model with time. The model can estimate the camera poses in real time when the PTZ camera is in motion. Then we compute the precise camera pose using the PTZ camera calibration based on key-ray collection when the PTZ camera is at the static. More details show as below.

5.1 | A Motion Estimation Model for PTZ Camera

PTZ camera models in most researches are usually the standard pinhole camera model with rotation matrixes, representing the static state of a PTZ camera. The models could not describe variations of PTZ camera parameters in moving. We propose a novel motion estimation model that is continuous as a function of time. We can quickly formulate the model with a series of simple initialization steps. This section describes the motion estimation model we proposed in detail.

In order to get the motion estimation model, we conducted 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 the original position. The results of the pan (or tilt) angle variation are shown in fig. 5(A)(or (B)). For the zoom of a PTZ camera, we control the camera to be directed to zoom purely for 0.5s, 1s, 1.5s, and return to the original position after waiting for the same time. The result of the focal length variation is portrayed in fig. 5(C). To get the accurate value of each frame, a series of images are collected before and after each motion. Parameters of these images have been calibrated by the parameter estimation method mentioned in sec. 4. The average tendencies of the pan θ, the tilt φ, and the focal length f with time are illustrated in Fig. 5.

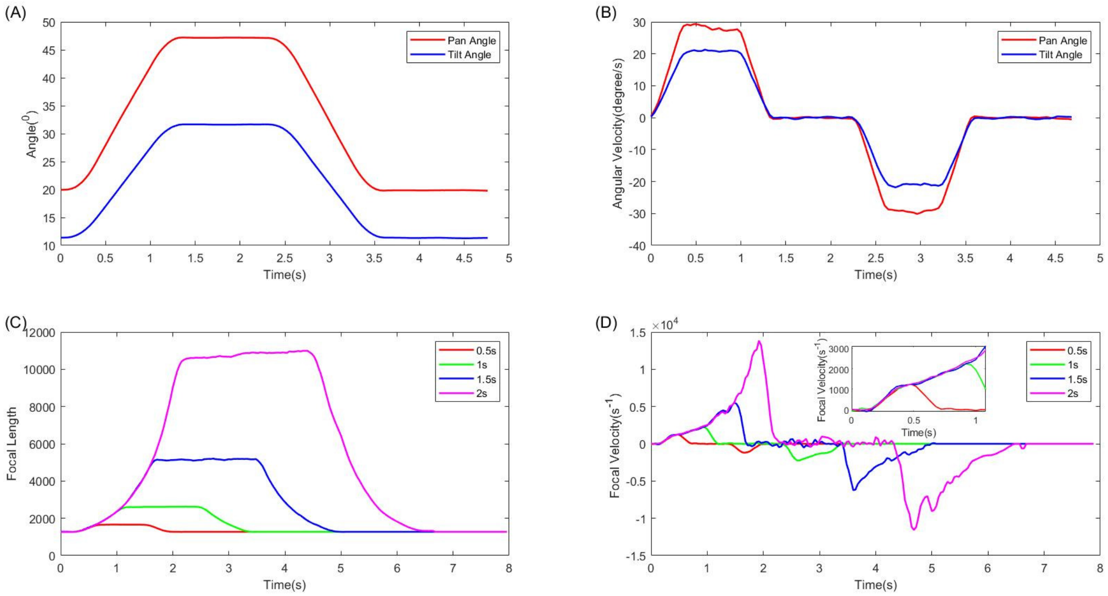


FIGURE 5 The average relationships of the pan θ, the tilt φ, and the focal length f with time. (A): pan/tilt angle as a function of time; (B): pan/tilt angle velocity as a function of time; (C) focal length as a function of time; (D) focal length velocity as a function of time.

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:

$$\begin{aligned}\{\theta,\varphi,f\} &= h\left(t \mid v_p, v_t, v_f, t_{p1}, t_{t1}\right)\\ &= \begin{cases} h_1\left(t \mid v_p, t_{p1}\right)\\ h_2\left(t \mid v_t, t_{t1}\right)\\ h_3\left(t \mid v_f\right)\end{cases}\end{aligned} \tag{16}$$

Where

$$h_1(t|v_p, t_{p1}) = \begin{cases} \frac{1}{2}v_p t^2, & (t < t_{p1})\\ v_p t - \frac{1}{2}v_p t_{p1}, & (t_{p1} < t < t_{p2})\\ \frac{1}{2}v_p(t + t_{p2} - t_{p1}) & \\ +\frac{1}{2}v_p(1-t)(t-t_{p1}), & (t > t_{p2}) \end{cases} \tag{17}$$

$$h_2(t|v_t, t_{t1}) = \begin{cases} \frac{1}{2}v_t t^2, & (t < t_{t1})\\ v_t t - \frac{1}{2}v_t t_{t1}, & (t_{t1} < t < t_{t2})\\ \frac{1}{2}v_t(t + t_{t2} - t_{t1}) & \\ +\frac{1}{2}v_t(1-t)(t-t_{t1}), & (t > t_{t2}) \end{cases} \tag{18}$$

$$h_3(t|v_f) = \frac{1}{2}v_f t^k + \sqrt[k]{2v_f(f_1 - f_0)}\,t + f_1 \tag{19}$$

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

5.2 | PTZ Calibaration Based on Key-ray Collection

The motion estimation model can real-time predict the moving PTZ camera. However, it is unreasonable to expect results of the model to be precise. We desire a dynamic correction method that ensures the camera calibration accurately. Our thought is to build a key-ray collection of the current scene and calibrate online images using the two-ray method.

The image location $p = (x, y)$ is projected by a ray $r = (\theta_p, \varphi_p)$, which is given by the formula 1 and 2. We extract all the SIFT features from each background image and transform image coordinates of features into rays. Then features from different images that highly overlap in both descriptors are merged, causing that each feature appears only once in the collection. Finally, we store all features (including rays and descriptors) in a collection as the key-ray collection of the scene. Figure 6 shows the pipeline of building the key-ray collection of the scene.

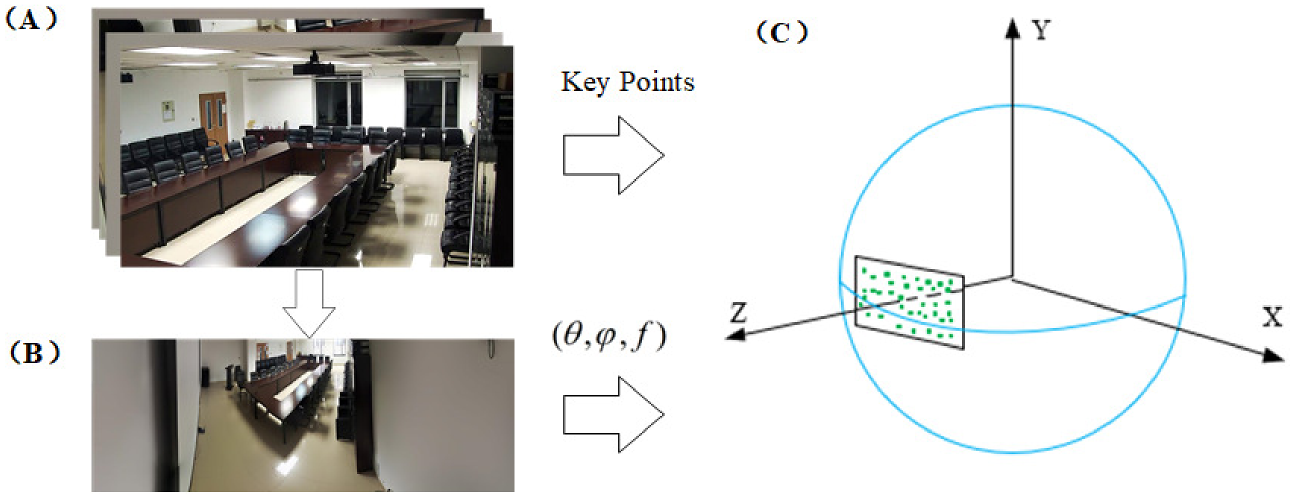


FIGURE 6 Pipeline of building the key-ray collection. (A): a set of background images that are used to stitch the panorama and provide SIFT features; (B): the panorama image that provide the (θ, φ, f) of background images; (C) the correspondence between key-points and key-rays. We translate the image coordinates of features to the rays.

We capture a current image from the real-time PTZ images and get the $(\hat{\theta}, \hat{\varphi})$ of the current image by the motion estimation model. We fetch out key rays around the pose of $(\hat{\theta}, \hat{\varphi})$ from the key-ray collection as a sub-collection M'_r. We extract the set of SIFT features, called as M', acquired from the current image. And we try to match each feature in the M' to the M'_r, resulting in a set of putative matches $M_r : \{p_k, r_k\}_{k=1}^{N}$, where p_k is from M' and r_k is from M'_r. The feature matches are computed using Brush-Force matching between SIFT descriptors.

We can estimate a precise guess for (θ, φ, f) of the current image using the idea of RANSAC[42]. We first select two feature pairs from M_r and use the two-ray method to calculate $(\hat{\theta}, \hat{\varphi}, \hat{f})$, which is a guess of (θ, φ, f). The inlier set P of the guess is

$$P_{inlier} = \{p_k | (\hat{\theta} + \arctan\frac{x_k - u}{\hat{f}\cos\hat{\theta}} - \theta_k)^2 + (\hat{\varphi} + \arctan\frac{y_k - v}{\hat{f}} - \varphi_k)^2 \leq \varepsilon\}_{k=1}^{N} \tag{20}$$

where $(x_k, y_k)^T$ is the coordinates ofp_k, $(\theta_k, \varphi_k)^T$ is the coordinates ofr_k, and (u, v) is the image center of the current image. The more accurate the guess, the greater the number of inliers the guess corresponds to. We repeat the above steps N_s times (we set $N_s = 1000$) and select a guess with the largest number of inliers as the estimation of the image parameters.

6 | EXPERIMENTS

We conducted experiments to evaluate the proposed model and algorithm using a public dataset and several self-built scenes. In self-built scenes, All cameras used are Hikvision PTZ cameras. Our approach is implemented with C++ on an Intel R core ^{TM} i7-975H CPU, NVIDIA GeForce GTX 2070M graphics card, 16GB memory Windows system.

6.1 | Experimental Setup

To the best of our knowledge, few works have been reported on datasets of PTZ camera images. Yong et al.[36] built a PTZ image dataset, which makes the in-depth investigation of PTZ panorama generation possible. Besides, we preform the data augmentation with weather conditions on this dataset. We also built four real scenes, with PTZ cameras, for experiments.

PTZ image dataset: This dataset possesses four groups of images from different natural traffic scenes, such as intersections or overpasses. Scene names of each group of images are 803, 878, 8425, and 8505, respectively. Each image group consists of 52 images. Please refer to the reference for the details of the dataset.

Data augmentation with weather conditions: We develop a weather augmentation algorithm inspired by the work of Kang et al[43]. To simulate extreme weather, the RGB (red, green, blue) image must be converted to the HLS (hue, saturation, lightness) format, and the lightness value would be adjusted based on different weather conditions.

Brightness and darkness: By adjusting lightness values of images, we are able to simulate sunny days and early evenings.

Rainy: We use randomly generated gray lines (corresponding to RGB values 160,166,166) to represent raindrops, and size and quantity of lines represent size and quantity of rain. We employ the fisheye effect to distort local regions of images, indicating that raindrops are stuck to camera lens. We simulated three types of rain: light rain, moderate rain and heavy rain.

Snowy: We use randomly generated white dots to depict snowflakes, with the size and number of dots representing snowflake size and quantity. This method expresses snow on the floor via extracting pixels from the specified lightness value in an image. We simulated three levels of snowfall: light snow, moderate snow and heavy snow.

Foggy: Fog can change brightness and sharpness of images and make the image blurred. We simulated fog, primarily mist and fog, with Gaussian blur and altered the lightness to make the images appear more realistic.

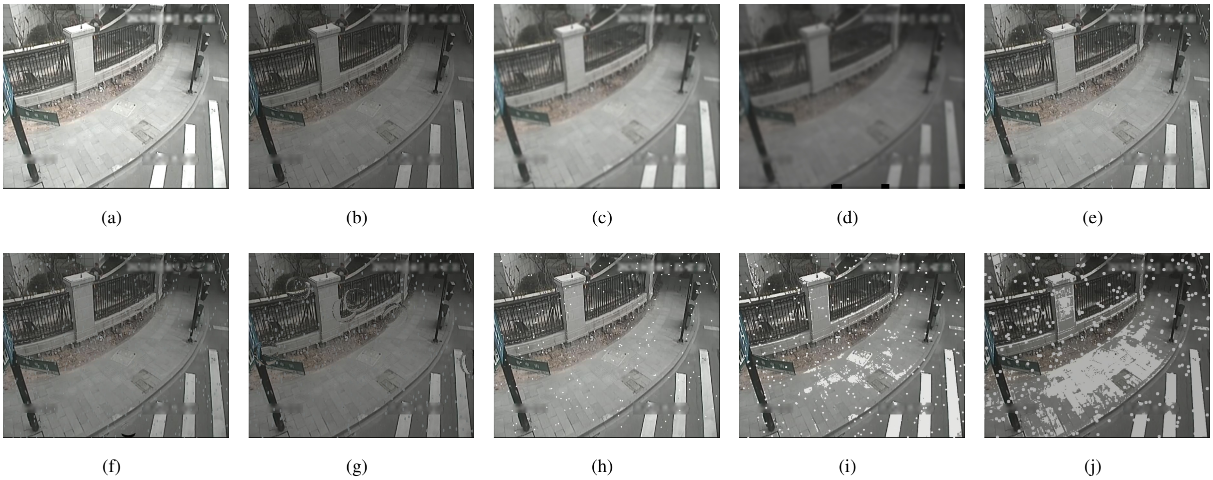


FIGURE 7 Visualization of weather augmentation: (a) brightness; (b) darkness; (c) mist; (d) fog; (e) light rain; (f) moderate rain; (g) heavy rain; (h) light snow; (i) moderate snow; (j) heavy snow.

We applied the above weather augmentation algorithm to the PTZ image dataset and obtained 10 datasets with different weather, illustrated in Fig. 7.

Real test scenes: We set up for real test scenes with a PTZ camera mounted (see Table 1). The first scene $Indoor_1$ is a $8m \times 6m \times 3m$ laboratory with a DS-2DC4223IW-D PTZ camera, which has chairs and tables. The second scene $Indoor_2$ is a $12m \times 8m \times 3m$ meeting room with a DS-2DE7172-A PTZ camera, which has few features. $Outdoor_1$ and $Outdoor_2$ are $50m \times 50m \times 50m$ outdoor scenes respectively, which have pedestrians and vehicles, mounting DS-2DC4223IW-D PTZ cameras.

TABLE 1 The Information of Real Test Scenes

Scene Name	PTZ Camera Type	Scene Size ($l \times w \times h$)	Scene Characteristics
$Indoor_1$	DS-2DC4223IW-D	$8m \times 6m \times 3m$	a laboratory with chairs, tables, and computers
$Indoor_2$	DS-2DE7172-A	$12m \times 8m \times 3m$	a meeting room with few feature points
$Outdoor_1$	DS-2DC5220IW-A	$50m \times 50m \times 50m$	a traffic scene with a PTZ camera at a lower height
$Outdoor_2$	DS-2DC5220IW-A	$50m \times 50m \times 50m$	a traffic scene with a PTZ camera at a higher height

6.2 | Comparison with Panoramic Model Construction

We would validate our parameters estimation algorithm on the PTZ image dataset. We use the parameter estimation from Yong's paper, called as the BAY method, as a benchmark. The method only optimized the three parameters of pan angle, tilt angle, and focal length. The BAY method tries to find the smallest residual possible. The idea may produce abnormal results. So we make a modification to the BAY method so that the first input image does not participate in optimization. We call the BAY method modified as the BAYs method. The essence of the BAYs method is to optimize other images with the first image as the reference. We use the BAYs method as the other benchmark. We performed qualitative and quantitative comparisons on the dataset, and results are shown in Tab. 2 and Fig. 8.

For quantitative comparison, we first take the first input image as the reference image. Assuming that the (θ, φ) are known, we only estimate the focal length of the reference image by the bundle adjustment method. We use the focal length estimated as the focal length of the reference image. Then in each scene, we select three consecutive images from different tilt angles and perform parameter estimation by our method, BAY method, and BAYs method. Due to the lack of ground truth, we calculated parameter estimation offsets using Euclidean distance between results parameter estimation and the preset parameters. Since there are only minor perturbations in image collection stage, correct parameters should be very close to preset values. So large offset indicates mistakes in the estimation. Meanwhile, we utilize results of parameter estimation to construct a homography matrix and calculate the reprojection error of each feature pair between adjacent images. The score of the reprojection error impacts the precision of parameter estimation. The experimental results are shown in Tab. 2. Tab. 2 reveals that although the BAY method has the smallest reprojection errors in many conditions, its offset is significantly larger than those of the other two methods. One reason lies in the lack of reference in the optimization stage, while the other comes from the target of the smallest reprojection error. The BAYs method overcomes this deficiency, but its accuracy is inferior to ours. Our method achieves excellent performance on both parameter estimation offset and reprojection error, which indicates that our method outperforms the other two methods in parameter estimation.

We project some of the parameter estimation results in Tab. 2 into a spherical surface, and results are shown in Fig. 8. Compared with the other two methods, our method can eliminate ghosts caused by inaccurate preset parameters (areas enclosed by red wireframes in the figure 8). Besides, the result of the BAY method in third column differs from that of the preset value in image size. This evidences that the image size of result from the BAY method changed due to the lack of reference in the optimization stage. The conclusion is consistent with that from in Tab. 2.

We further test the effect of extreme weather on our approach. We select two images with sequence number 18 and 19 under different weather conditions and scenes, and use our method, BAY, BAYs to estimate their parameters respectively, and calculate the reprojection error between them. The experimental results are listed in the Tab. 3.

As can be shown in Tab. 3, all three methods exhibit outstanding performance in weather of brightness, darkness, light rain, and light snow. This demonstrates that the three algorithms are capable of removing interference caused by tiny weather change. However, as the severity of weather increases, such as moderate rain and mist weather, the BAY method and the BAYs method start yielding wrong estimation in the 878 scene. In the weather of moderate snow, heavy rain, heavy snow, and heavy fog, the method of BAY and BAYs almost fails, either producing a completely wrong estimation (e.g. the 8505 scene in moderate snow and the 878 scene in fog), or failing to estimate the parameters (e.g. the 878 scene in heavy rain and the 8425 scene in heavy snow). Our method can still maintain good robustness in these weather conditions. Even if other methods fail, estimation results

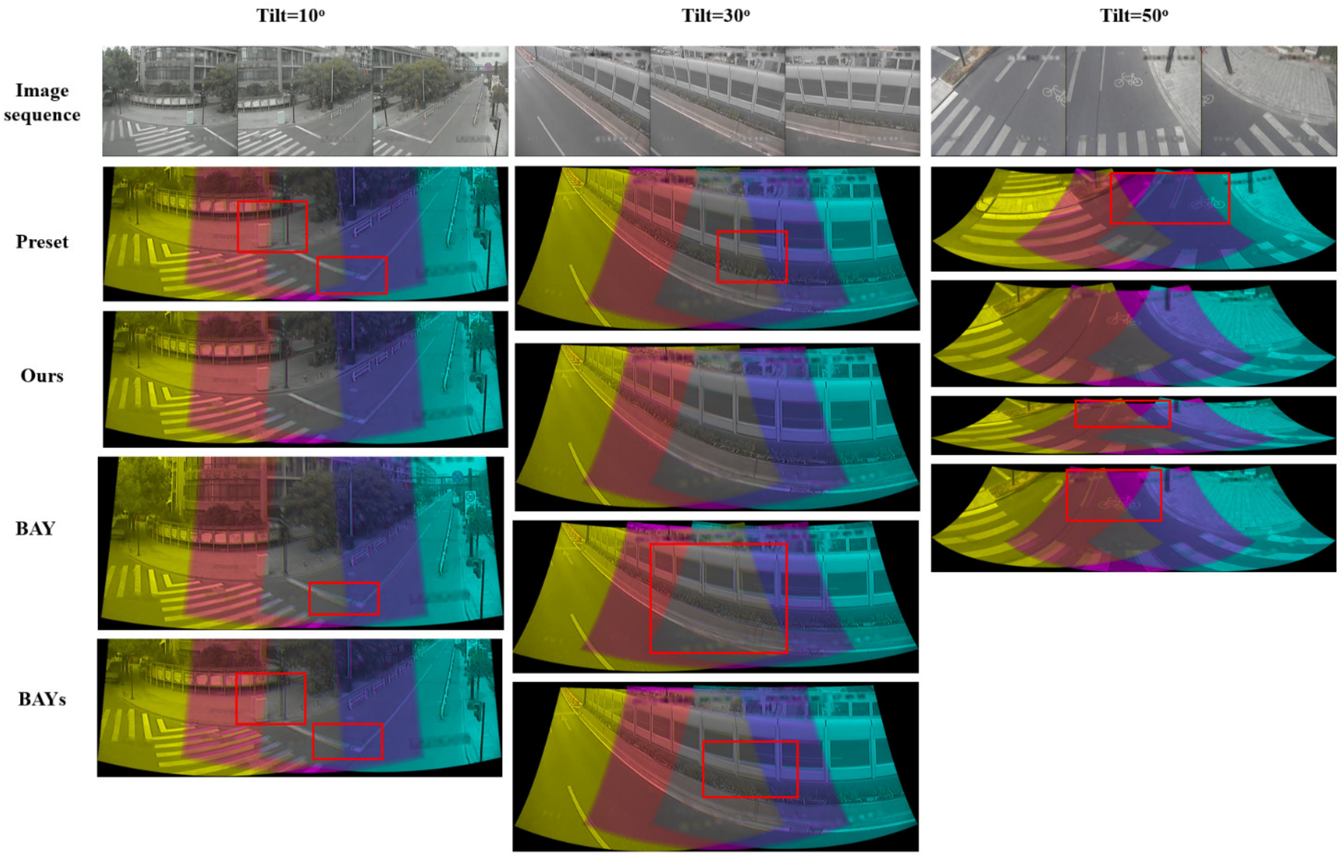


FIGURE 8 Qualitative comparison of parameter estimations in Yong's dataset. The image sequence of the first column is (1,2,3) of the 803 scene, that of the second column is (27,28,29) of the 8505 scene, and that of the second column is (40,41,42) of the 878 scene. The yellow indicates the image one, the green indicates the image three, and other colors indicate overlapping areas between adjacent images. Areas enclosed in red indicate obvious misalignment.

of our method increase only several pixel offsets over the normal case (e.g. the 878 scene in heavy rain and the 8425 scene in heavy snow).

6.3 | Performance Analysis of PTZ Camera Real-time Calibration Algorithm

6.3.0.1 | *Performance of motion estimation model:*

We first get background images of the camera's scene and the parameters using the method in Sec. 4. Then we collect scene PTZ images and formulate motion estimation models of different cameras using the method mentioned in Sec. 5. We establish no less than two correspondences between PTZ images and background images with human annotation, transforming background images' position to rays position. Finally, we estimate camera parameters using the two-ray method as the ground truth. Table 4 shows errors between the motion estimation model and the ground truth of different cameras.

In Table 4, we present the means and standard errors of motion estimation models from different scenes. The mean rotational error of the pan angle is about $0.6°$ and the mean rotational error of the 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 quickly. The focal length error of the 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 percentage decreases with focal length. The average velocity of focal length is $5000/s$, since the focal length error can be ignored in zooming. Table 4 demonstrates the accuracy of the motion estimation model.

TABLE 2 Quantitative comparison of parameter estimations in the PTZ image dataset. **BOLD**/BLUE indicates state-of-the-art/second-best performance.

Scene name	Image number			Method	parameter estimation offset			The mean reprojection error of single matching pair	
	Img 1	Img 2	Img 3		Img 1	Img 2	Img 3	Img 1 and 2	Img 2 and 3
803	1	2	3	OurMethod	**8.72**	**8.76**	**8.75**	7.82	6.06
				BAY	267.83	267.78	267.83	**3.53**	**2.58**
				BAYs	12.24	12.25	12.26	9.63	6.10
	27	28	29	OurMethod	**10.41**	**10.43**	**10.46**	10.43	5.52
				BAY	293.72	293.70	293.73	**7.28**	**6.76**
				BAYs	13.01	13.01	13.05	10.64	7.66
	41	42	43	OurMethod	7.56	7.87	7.87	**9.45**	**7.01**
				BAY	1572.43	1572.41	1572.44	29.17	17.13
				BAYs	**0.00**	**1.60**	**2.70**	35.74	14.98
878	1	2	3	OurMethod	4.65	4.71	4.72	5.65	4.85
				BAY	293.15	293.09	293.13	**4.74**	**2.07**
				BAYs	**1.46**	**1.60**	**1.67**	6.14	5.03
	27	28	29	OurMethod	3.59	3.94	4.08	6.65	**4.54**
				BAY	221.88	221.86	221.87	**9.85**	**4.54**
				BAYs	**0.76**	**1.29**	**1.56**	17.79	7.00
	41	42	43	OurMethod	**8.92**	**9.31**	**9.46**	6.59	**6.93**
				BAY	881.46	881.58	881.77	**5.87**	11.46
				BAYs	37.58	37.58	37.58	21.86	8.12
8425	1	2	3	OurMethod	**8.17**	**8.19**	**8.18**	4.54	4.12
				BAY	242.38	242.34	242.40	**1.27**	**1.45**
				BAYs	9.47	9.47	9.48	6.29	4.58
	27	28	29	OurMethod	**1.42**	2.04	1.81	**10.19**	5.80
				BAY	385.76	385.73	385.76	13.70	**2.77**
				BAYs	1.52	**1.70**	**1.67**	10.25	8.42
	41	42	43	OurMethod	**9.61**	**9.62**	**9.71**	16.37	22.32
				BAY	163.46	163.53	163.69	**3.14**	9.28
				BAYs	22.95	23.01	22.97	4.59	**8.57**
8505	1	2	3	OurMethod	5.79	5.81	5.80	4.72	4.65
				BAY	230.45	230.41	230.46	**2.79**	**2.68**
				BAYs	**1.54**	**1.55**	**1.57**	4.85	5.16
	27	28	29	OurMethod	4.16	4.21	4.24	**5.74**	**4.71**
				BAY	357.70	357.66	357.69	8.23	5.98
				BAYs	**1.12**	**1.30**	**1.56**	11.12	6.17
	41	42	43	OurMethod	**8.91**	**9.21**	**9.22**	**10.82**	**8.63**
				BAY	1559.39	1559.34	1559.35	16.50	11.65
				BAYs	33.92	33.94	33.93	25.66	11.48

6.3.0.2 | *Performance of PTZ calibration:*

We test our approach at the scene *indoor*$_1$ in Table 1. In the scene *indoor*$_1$, we collect 50 images from pan, tilt, and zoom values. The reference images are from the background dataset in Sec. 4. We compare our approach against several PTZ camera calibration

TABLE 3 Reprojection errors of parameter estimation in extreme weather. **BOLD** indicates the best performance, RED indicates the enormous error, and '-' indicates estimation failure.

Scene	Method	Origination	Brightness	Darkness	Light Rain	Mod. Rain	Heavy Rain	Light Snow	Mod. Snow	Heavy Snow	Mist	Fog
803	Ours	7.29	8.25	10.90	7.19	24.46	**5.63**	8.01	**6.52**	**12.17**	9.52	**5.53**
	BAY	**4.59**	**7.78**	**4.48**	**6.84**	**5.50**	36.63	**5.48**	132.23	76.24	**6.30**	8.19
	BAYs	12.11	9.20	14.95	11.00	6.25	36.34	9.97	147.40	123.95	8.80	8.28
878	Ours	14.72	8.28	14.91	13.80	**15.95**	**15.76**	**13.50**	**21.64**	16.99	**19.39**	31.80
	BAY	**2.39**	**3.05**	**2.35**	**2.64**	92.06	-	18.06	99.17	**0.07**	284.42	264.43
	BAYs	10.14	10.69	11.04	9.95	107.59	-	28.54	97.56	0.07	284.49	231.46
8425	Ours	4.73	4.74	5.05	**4.68**	6.35	5.75	4.98	**3.40**	**12.31**	4.73	4.25
	BAY	**2.79**	**3.43**	**2.93**	5.45	**4.45**	**2.61**	**3.53**	26.27	-	**2.25**	**3.27**
	BAYs	6.33	6.24	7.78	8.64	3.81	7.03	8.19	41.53	-	6.80	2.72
8505	Ours	**6.08**	**6.43**	**5.97**	**5.99**	**6.14**	3.75	**6.58**	**12.09**	**17.44**	**6.34**	**21.50**
	BAY	6.98	6.47	8.69	6.69	8.13	**0.14**	7.24	238.04	197.91	9.05	275.75
	BAYs	11.30	10.14	11.97	10.95	7.13	0.13	10.99	242.10	189.99	9.01	272.19

methods which can directly compute the parameters of the PTZ camera by feature point pairs: the dynamic calibration[27], two-point method[28], and the TPCM method[18]. Because these three methods depend on accurate feature correspondence, we first remove the evidently wrong mismatched pairs with RANSAC.

Fig. 9 shows the biases between those methods and the ground truth established by manual determination. The reported bias for each parameter is computed by $||param_{est} - param_{ground-truth}||_2$. From the results, we analyze that our PTZ calibration method outperforms for all the parameters and the estimation biases are smaller than other methods. The two-point method has a large bias because the random forest depends on trained images. The more images trained, the better the predicted result of the random forest. We only use the background dataset to train the random forest, so the predicted result has a conspicuous error. Because the TPCM method is implemented via multiple additions and multiplications which enlarge the errors, resulting in amazing fluctuations the TPCM produce.

We also collected images of the outdoor scene $outdoor_2$ at 2:00 p.m. and 6:00 p.m. to evaluate the effects of illumination on camera calibration. We chose three groups of images: images with few substantial change in illumination (pan = 0°, $tilt$ = 10°), images with local changes in illumination (pan = 60°, $tilt$ = 10°), and images with entirely different illumination (pan = 80°, $tilt$ = 10°). Each group had the same pose. We estimate the parameters of these three groups using Dynamic calibration, Two-point, TPCM, and our method, and then project the image onto the sphere using the estimated parameters. The experimental results are shown in Fig. 10. Since two images have the same poses, they should be projected in the same region. We use a color

TABLE 4 The Motion Model Errors

PTZ Camera	**Model Errors**		
Type	***Pan*(°)**	***Tilt*(°)**	***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 ± 0.162.45

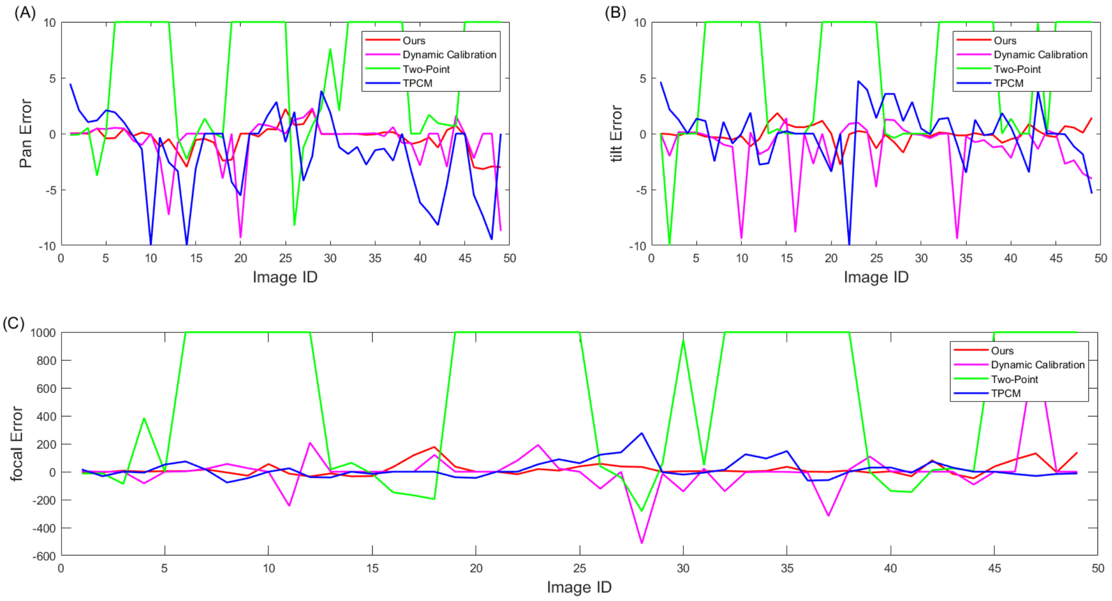


FIGURE 9 Results of several PTZ calibration methods. (A) The error in the pan angle θ. (B) The error in The Tilt angle φ. (C) The error in the focal length f.

of yellow indicating images at 6:00 p.m. and a color of purple indicating images at 2:00 p.m. The overlapping region of two images is shown in color of light red. If the estimation is correct, only the light red area will be visible. If the estimation is inaccurate, we will see yellow or purple areas. We also provide the reprojection error of matched pairs between images under every projection. When the illumination change is subtle (the first column in the Fig. 10), all four methods obtain acceptable parameter estimation results with small reprojection error. When the local illumination of the image changes (second column in the Fig. 10), the Dynamic calibration method produces a significant bias (yellow part in the third row), and the results estimated by the Two-point method and the TPCM method are both biased (purple area in the fourth row and yellow area in the fifth row). However, our method continues to produce accurate parameter estimations. When the difference in illumination is significant (third column in the Fig. 10), the Dynamic calibration method, the Two-point method, and the TPCM method all fail significantly (yellow and purple areas in the figure). Our method yet obtains acceptable results, with only a one-point increase in reprojection error compared to that of the first column.

6.4 | The Results of Visual Expansion and Real-time Calibration System

We design a visual expansion and real-time calibration system for pan-tilt-zoom cameras assisted by panoramic models. The interface of the system is shown in the figure 11 (a). The "Main View" slider can adjust the visible area of the right view to see more extensive parts of panoramic models. The "Reconnection" button can re-establish the link with the PTZ camera, and the "Zero Set" button can restore the PTZ camera to its original pose. The "Set" button can set the camera to the specified pose. The "PTZ Calibration" button can calibrate the PTZ camera using the PTZ calibration based on key-ray collection and register the PTZ images into the specified position of the panoramic model. The "Update Image" button can refresh the latest PTZ image. The buttons in the lower-left of the interface can control the PTZ camera rotation and zooming.

Because most of AVE systems doesn't provide their codes, we cannot compare performance and effect with them in our test scenes. To better comprehend the significance of our work, we still intend to conduct a comprehensive comparison of their functionalities. We compared our system with traditional PTZ camera control systems and some typical AVE systems in terms of functionality. The experimental results are shown in the table 5. From the table 5, we find that the interest in AVE technologies is concentrated on smartphone cameras, and they almost rely on inbuilt sensors. However, no sensors are mounted in PTZ cameras, and these AVE techniques cannot be applied to PTZ camera calibration. Traditional PTZ camera control focuses solely on PTZ

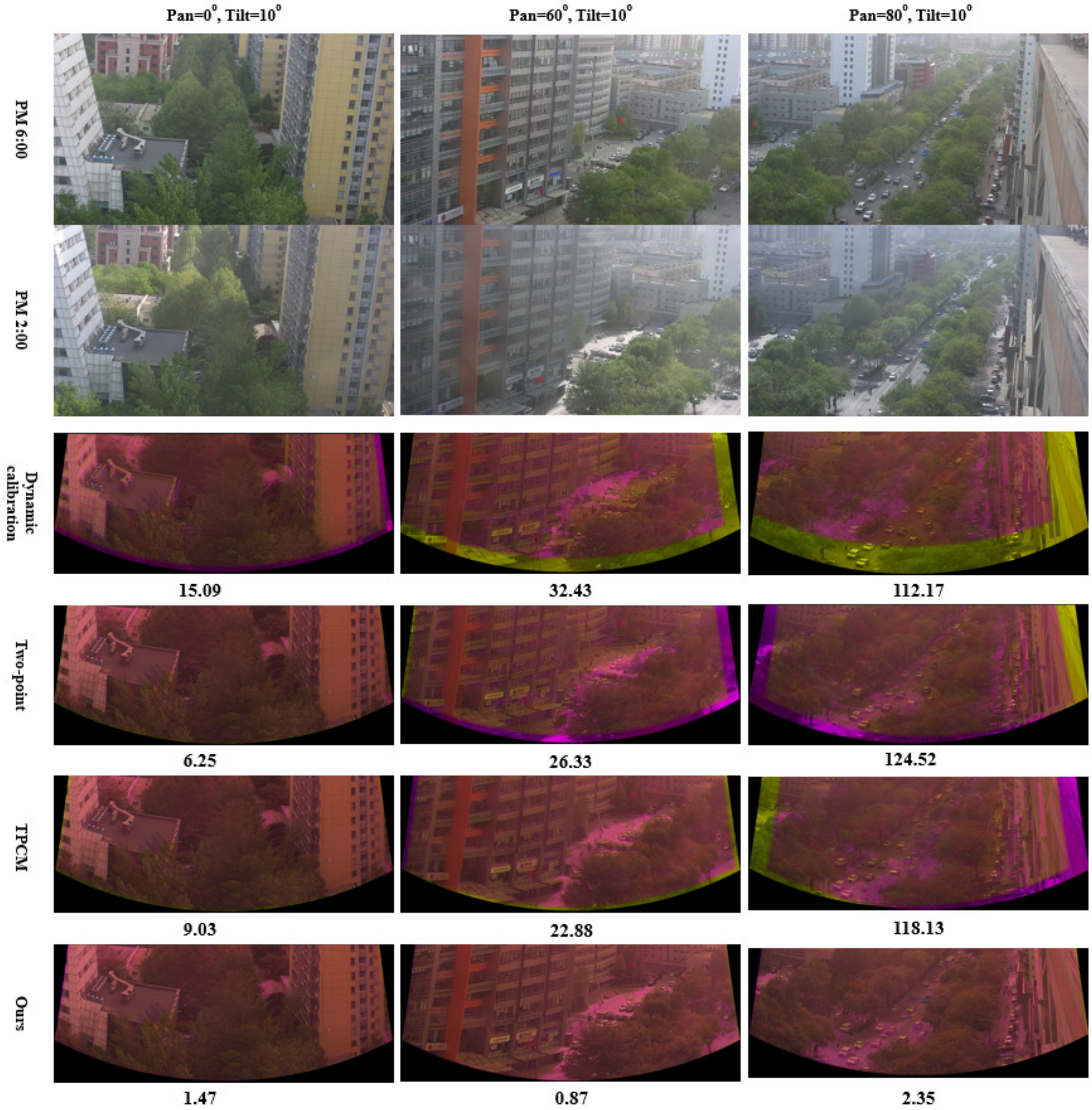
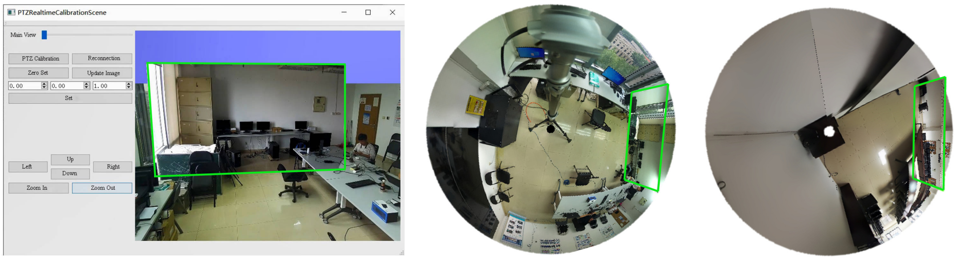


FIGURE 10 PTZ calibration results under different illumination conditions. The first column indicates that the condition with few substantial changes in illumination; The second column indicates the condition with local changes in illumination; The third column indicates that the condition with entirely different illumination. In the figure, the yellow area indicates the image at 6:00 p.m.; the purple part indicates images at 2:00 p.m.; the light red area indicates the overlap of the two images.

camera control and video transmission, disregarding the surroundings of the camera. This is a limitation of conventional PTZ camera manipulation.

We test the system in different scenarios and record the results in the figure 12. In Fig. 12, we observe that our approach still perform eminent on the less-texture region from the first row. In Outdoor, more features mean more distractors. However, our method gain a satisfactory outcome demonstrating the applicability of our method both indoors and outdoors. In the third row, we attempt to prove the robustness of our system for movable objects, such as chairs and tables, working as expected. In addition, our system can effortlessly reach the target area with the support of the panoramic model, which is difficult to achieve in other AVE systems yet.



(a) (b)

FIGURE 11 The visual expansion and real-time calibration system and panoramic models. Figure (a) is the system interface of the visual expansion and real-time calibration system. Figure (b) is the top view of Panoramic models. The green box denotes the PTZ image projection of PTZ camera.

TABLE 5 Functional comparison of different systems.

System	Environment	Camera type	Functions	Limitations
Traditional PTZ control system	Outdoor Indoor	PTZ camera	Focuses on PTZ camera control and video transmission	Accurate and rapid control of multiple cameras requires operators to memorize the environment of each camera, which is a burden.
Work of Pece et al.[38]	Indoor	Smartphone camera	Uses smartphone cameras to create a surround representation of meeting places	1. Relies on markers; 2. Only achieves static camera position.
Work of Tompkin et al.[2]	Outdoor	Smartphone camera	Shows changes of places on Google Street View and demonstrates several representations for different displays.	1. Relies on GPS location, and orientation data from multiple integrated sensors; 2. Only calibrate video frames offline.
Work of Young et al.[4]	Outdoor	Smartphone camera	Provides telepresence and remote collaboration on mobile and wearable devices	1. Relies on the device's inbuilt sensors; 2. Unaccurate camera position in the panorama
Our system	Outdoor Indoor	PTZ camera	1. Achieves accurate and rapid control of PTZ cameras via background panoramic models; 2. Realizes the alignment between PTZ images and panoramic models in real time	Not consideres the lens distortion.

7 | CONCLUSION

We propose a novel visual expansion and real-time calibration for PTZ cameras assisted by panoramic models. We first develop the two-point method into the two-ray method, which is applied to two adjacent PTZ images for parameter estimation. Second, we strengthen the feature matching and parameter estimation of the image mosaic method by exploiting the geometric property of PTZ adjacent images. The improved method achieve more excellent the parameter estimation during the spherical stitching of PTZ images. We construct the panoramic model based the improved method. Then we present a real-time PTZ camera calibration algorithm primarily composed of the PTZ motion estimation model and the camera calibration algorithm based on key-ray collection. We verify our method on both public and self-built data sets. As known from the experimental results that our method can indicate outstanding performance. We also design a visual expansion and real-time calibration system using our method, which realizes effective control of the PTZ camera through the panoramic model.

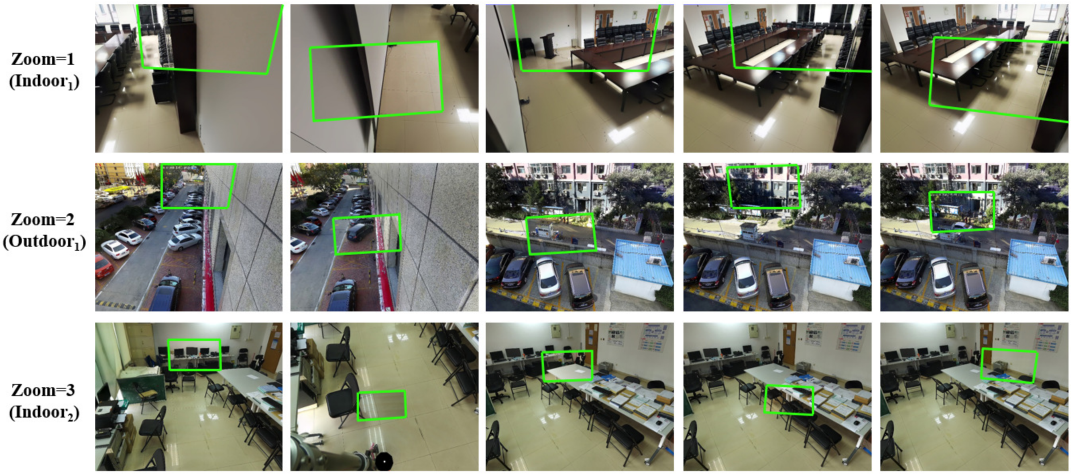


FIGURE 12 The results of our system. The green box indicates the projection of real-time PTZ images in panoramic models. Regions outside green boxes are panoramic models.

However, there are several limitations to this work. We have yet to consider the lens distortion carrying some biases during PTZ camera calibration. Also, we assume that the principal point coincides with the projection center and zooming center. However, this may not uniformly be the truth for cameras in zooming. Finally, while we presume that the optical center of the PTZ camera is fixed, it would have variations when the camera rotates in reality. Those factors require to be taken into account if we expect a more accurate result.

FUNDING INFORMATION

This work was supported in part by the Science and Technology Project of Hainan Provincial Department of Transportation under Grant HNJTT- KXC-2024-3-22-02, in part by the National Natural Science Foundation of China under Grant 62272018.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Sawhney HS, Arpa A, Kumar R, et al. Video flashlights: real time rendering of multiple videos for immersive model visualization. In: . 28. 2002:157–168.
2. Tompkin J, Pece F, Shah R, Izadi S, Kautz J, Theobalt C. Video collections in panoramic contexts. In: 2013:131–140.
3. Wang M, Shamir A, Yang GY, et al. BiggerSelfie: Selfie video expansion with hand-held camera. *IEEE Transactions on Image Processing.* 2018;27(12):5854–5865.
4. Young J, Langlotz T, Cook M, Mills S, Regenbrecht H. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics.* 2019;25(5):1908–1918.
5. Song L, Wu W, Guo J, Li X. Survey on camera calibration technique. In: . 2. IEEE. 2013:389–392.
6. Long DS. Review of Camera Calibration Algorithms. In: 2019:723–732.
7. Chen Z, Si X, Wu D, Tian F, Zheng Z, Li R. A novel camera calibration method based on known rotations and translations. *Computer Vision and Image Understanding.* 2024;243:103996.
8. Tsai RY. An efficient and accurate camera calibration technique fro 3d machine vision. In: 1986.
9. Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence.* 2000;22(11):1330–1334.
10. Zhao F, Tamaki T, Kurita T, Raytchev B, Kaneda K. Marker based simple non-overlapping camera calibration. In: IEEE. 2016:1180–1184.
11. Campbell D, Petersson L, Kneip L, Li H. Globally-optimal inlier set maximisation for camera pose and correspondence estimation. *IEEE transactions on pattern analysis and machine intelligence.* 2018;42(2):328–342.
12. Lin Y, Larsson V, Geppert M, Kukelova Z, Pollefeys M, Sattler T. Infrastructure-based multi-camera calibration using radial projections. In: Springer. 2020:327–344.
13. Ren L, Chang H, Liu C, et al. A Calibration Algorithm of 3D Point Cloud Acquisition System Based on KMPE Cost Function. *IEEE Transactions on Instrumentation and Measurement.* 2024.
14. De Ma S. A self-calibration technique for active vision systems. *IEEE Transactions on Robotics and Automation.* 1996;12(1):114–120.
15. Luong QT, Faugeras OD. Self-calibration of a moving camera from point correspondences and fundamental matrices. *International Journal of computer vision.* 1997;22(3):261–289.

16. Vasconcelos F, Barreto JP, Boyer E. Automatic camera calibration using multiple sets of pairwise correspondences. *IEEE transactions on pattern analysis and machine intelligence.* 2017;40(4):791–803.
17. Liu S, Chen J, Chang CH, Ai Y. A new accurate and fast homography computation algorithm for sports and traffic video analysis. *IEEE Transactions on Circuits and Systems for Video Technology.* 2017;28(10):2993–3006.
18. An P, Ma J, Ma T, et al. Two-point calibration method for a zoom camera with an approximate focal-invariant radial distortion model. *JOSA A.* 2021;38(4):504–514.
19. YANG H, XIAO T, WU L, DENG F. Global Image Orientation Method for PTZ Camera with Pure Rotation. *Geomatics and Information Science of Wuhan University.* 2024.
20. Wen PC, Cheng WC, Wang YS, Chu HK, Tang NC, Liao HYM. Court reconstruction for camera calibration in broadcast basketball videos. *IEEE transactions on visualization and computer graphics.* 2015;22(5):1517–1526.
21. Homayounfar N, Fidler S, Urtasun R. Sports field localization via deep structured models. In: 2017:5212–5220.
22. Sharma RA, Bhat B, Gandhi V, Jawahar C. Automated top view registration of broadcast football videos. In: IEEE. 2018:305–313.
23. Rematas K, Kemelmacher-Shlizerman I, Curless B, Seitz S. Soccer on your tabletop. In: 2018:4738–4747.
24. Citraro L, Márquez-Neila P, Savare S, et al. Real-time camera pose estimation for sports fields. *Machine Vision and Applications.* 2020;31(3):1–13.
25. Sha L, Hobbs J, Felsen P, Wei X, Lucey P, Ganguly S. End-to-end camera calibration for broadcast videos. In: 2020:13627–13636.
26. Faugeras OD, Luong QT, Maybank SJ. Camera self-calibration: Theory and experiments. In: Springer. 1992:321–334.
27. Wu Z, Radke RJ. Keeping a pan-tilt-zoom camera calibrated. *IEEE transactions on pattern analysis and machine intelligence.* 2012;35(8):1994–2007.
28. Chen J, Zhu F, Little JJ. A two-point method for PTZ camera calibration in sports. In: IEEE. 2018:287–295.
29. Lu J, Chen J, Little JJ. Pan-tilt-zoom SLAM for Sports Videos. *arXiv preprint arXiv:1907.08816.* 2019.
30. Chen J, Little JJ. Sports camera calibration via synthetic data. In: 2019:1–8.
31. Brown M, Lowe DG. Automatic panoramic image stitching using invariant features. *International journal of computer vision.* 2007;74(1):59–73.
32. Gao J, Kim SJ, Brown MS. Constructing image panoramas using dual-homography warping. In: IEEE 2011; Piscataway:49–56
33. Zaragoza J, Chin TJ, Brown MS, Suter D. As-Projective-As-Possible Image Stitching with Moving DLT. In: 2013.
34. Li J, Wang Z, Lai S, Zhai Y, Zhang M. Parallax-Tolerant Image Stitching Based on Robust Elastic Warping. *IEEE Transactions on Multimedia.* 2018:1672-1687.
35. Li J, Deng B, Tang R, Wang Z, Yan Y. Local-Adaptive Image Alignment Based on Triangular Facet Approximation. *IEEE Transactions on Image Processing.* 2019;PP(99):1-1.
36. Yong H, Huang J, Xiang W, Hua X, Zhang L. Panoramic background image generation for PTZ cameras. *IEEE Transactions on Image Processing.* 2019;28(7):3162–3176.
37. Chen SC, Lee CY, Lin CW, et al. 2D and 3D visualization with dual-resolution for surveillance. In: IEEE. 2012:23–30.
38. Pece F, Steptoe W, Wanner F, et al. Panoinserts: mobile spatial teleconferencing. In: 2013:1319–1328.
39. Yi Z, Ming M, Wei W, Zhong Z. Virtual-reality video fusion system based on video model. *Journal of System Simulation.* 2018;30(7):2550.
40. Ma J, Zhao J, Jiang J, Zhou H, Guo X. Locality preserving matching. *International Journal of Computer Vision.* 2019;127(5):512–531.
41. Lowe DG. Distinctive image features from scale-invariant keypoints. *International journal of computer vision.* 2004;60(2):91–110.
42. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM.* 1981;24(6):381–395.
43. Kang KS, Cho YW, Jin KH, Kim YB, Ryu HG. Application of one-stage instance segmentation with weather conditions in surveillance cameras at construction sites. *Automation in Construction.* 2022;133:104034.

AUTHOR BIOGRAPHY

Liangliang Cai. received the B.S. degree from the Central South University of China, Changsha, China, in 2018. He is currently pursuing the Ph.D. degree at the State Key Laboratory of Virtual Reality Technology and Systems, computer science and technology of Beihang University, Beijing, China. His current research interests include computer vision, camera localization and scene understanding.

Zhong Zhou. received the B.S. degree in material physics from Nanjing University in 1999 and the Ph.D. degree in computer science and engineering from Beihang University, Beijing, China, in 2005. He is currently a Professor and Ph.D. Adviser with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University. His main research interests include virtual reality, augmented reality, computer vision, and artificial intelligence.