

ARTICLE TYPE

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

Liangliang Cai¹ | Zhong Zhou^{1,2}¹ State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China² 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 judge their directions using contextual clues from the surrounding environment. To achieve this goal, some research projects camera videos into three-dimensional (3D) models or panoramas, and allows operators to establish spatial relationships. However, these works face several challenges in terms of real-time processing, localization accuracy, and realistic reference. To address this problem, a visual expansion and real-time calibration for PTZ cameras assisted by panoramic models is proposed. The calibration method consists of three parts: providing a real environment background by building a panoramic model, meeting the needs of real-time processing by establishing a PTZ camera motion estimation model, and achieving high-precision alignment between PTZ images and panoramic models using only two feature point pairs. Our methods were validated using both public and our Scene dataset. The experimental results indicate that our method outperforms other state-of-the-art methods in terms of real-time processing, accuracy, and robustness.

KEY WORDS

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

1 | INTRODUCTION

Pan-tilt-zoom (PTZ) cameras provide extensive surveillance capabilities through flexible control, and have been widely implemented in a variety of settings, such as schools, universities, companies, and stadiums. The broad array of emergencies in real-world scenarios necessitates rapid response from PTZ cameras to locate them accurately. To enhance the rapid response ability, numerous studies^{1,2,3,4} in the field of augmented virtual environments (AVE) have introduced background models as environmental cues, including three-dimensional (3D) models and panoramas, to assist in the operation of PTZ cameras. These approaches involve real-time image projection technologies within a virtual environment to render realistic textures onto reference models. However, the majority of studies suffer from misalignments between images and backgrounds, and lack realistic reference models, as depicted in Fig. 1.

To address these challenges, many researchers employ tilt photography models and enhance camera calibration algorithms. The camera calibration is the process of estimating the internal and external parameters of the camera, as well as determining its orientation and position relative to a reference model^{5,6,7}. Camera calibration methods are classified into pattern-based^{8,9,10}, infrastructure-based^{11,12,13}, and self-calibration methods^{14,15,16,17,18,19}. Pattern-based methods exploit the regularity and symmetry of the patterns to enable precise measurements of camera parameters. Due to the necessity of specific calibration patterns, these methods are challenging to implement in outdoor settings. Infrastructure-based approaches typically entail constructing a controlled environment with known reference points or structures for camera calibration. Nevertheless, the construction and deployment of this infrastructure often incur significant labor costs. Self-calibration methods estimate both intrinsic and extrinsic camera parameters by establishing numerous point correspondences between adjacent images. The accuracy of these methods can be influenced by factors such as illumination conditions, moving objects, and feature density.

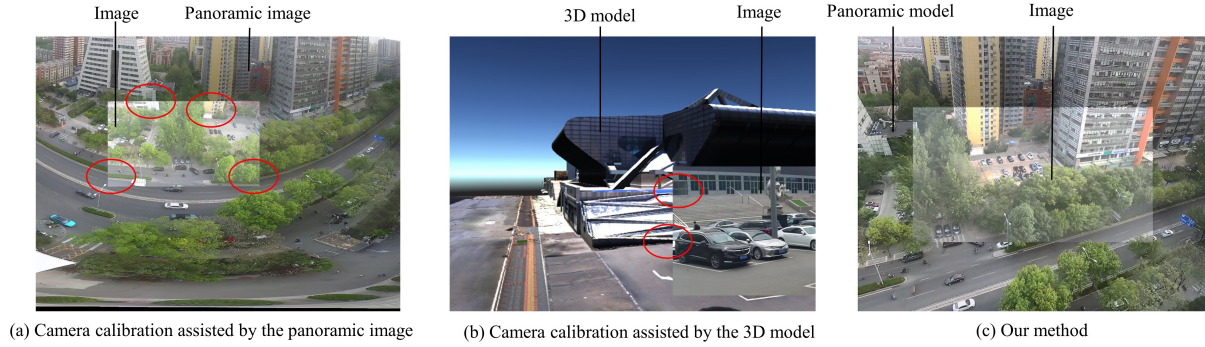


FIGURE 1 Misalignments and unrealistic reference models were observed in camera calibration processes. The red-encircled regions highlight misalignment of embedded images with the background. The panoramic images and 3D models illustrated in the figure are typical environmental models. Nonetheless, distortions in the panorama and the unnatural textures of the 3D model result in significant discrepancies between these background models and real-world scenes.

In this paper, we present a novel approach for visual expansion and real-time calibration of PTZ cameras using panoramic models. This integration harnesses the benefits of both infrastructure-based methods and self-calibration techniques. The method also takes into account the geometric properties of the PTZ camera, as illustrated in Fig. 2. We first introduce a two-ray method that requires only two feature point pairs to estimate parameters based on the geometric properties of the PTZ camera. Subsequently, we propose a novel method for constructing panoramic models based on a PTZ camera model. Initially, we collect scene images from the PTZ camera and erase the moving objects to obtain pure background images. Subsequently, we stitch these images utilizing the image stitching method, thereby generating a panorama of the current scene. To enhance the quality of image stitching, we optimize two phases using the two-ray method: feature matching and parameter estimation. Ultimately, we map the panorama onto a half-sphere model to construct the panoramic model. Also, we present a real-time PTZ camera calibration algorithm. The algorithm principally consists of a motion estimation model and a camera calibration algorithm based on key-ray collection. The motion estimation model, derived from the parameter variations during camera movement, enables rapid camera pose estimation during motion. The camera calibration algorithm based on key-ray collection performs feature extraction on the set of background image, and constructs a key-ray collection for the background model. Thereafter, the algorithm employs the two-ray method for PTZ camera calibration, yielding a series of candidate parameters for the current image, and selects the optimal camera parameters as the camera pose corresponding to the current image. The algorithm realizes high calibration accuracy for the PTZ camera by using panoramic models. The two-ray method requires only two pairs of matched points between adjacent images to estimate the pose of the PTZ camera. The reduced number of feature points utilized for parameter estimation minimizes the impact of illumination and appearance variations, which is particularly crucial in extreme weather conditions. This presents a significant advancement over previous methods.

In contrast to traditional PTZ camera calibrations and augmented virtual environment, we introduce a panoramic model, which offers a realistic environmental context with a lower labor cost. The panoramic model assists operators in determining the direction and zoom level of the camera. Additionally, our study focuses on a PTZ camera calibration algorithm that employs merely two feature pairs. Our method employs fewer feature pairs yet achieves superior calibration accuracy compared to alternative methods. Besides, the motion estimation model for the PTZ camera satisfies the requirement for real-time processing while minimizing resource usage. Our contribution can be summarized as follows:

- A two-ray method for PTZ camera calibration is developed from 2D-3D correspondences to 2D-2D correspondences. The method can enhance the pose accuracy of a PTZ camera using solely two pairs of feature points.
- A comprehensive framework of a visual expansion and real-time calibration for PTZ cameras assisted by panoramic models is designed to alleviate the burden of understanding the spatial relation and to simplify PTZ camera manipulation for operators.
- Extensive experiments demonstrate that our methods outperform other state-of-the-art methods in terms of real-time processing, accuracy, and robustness.

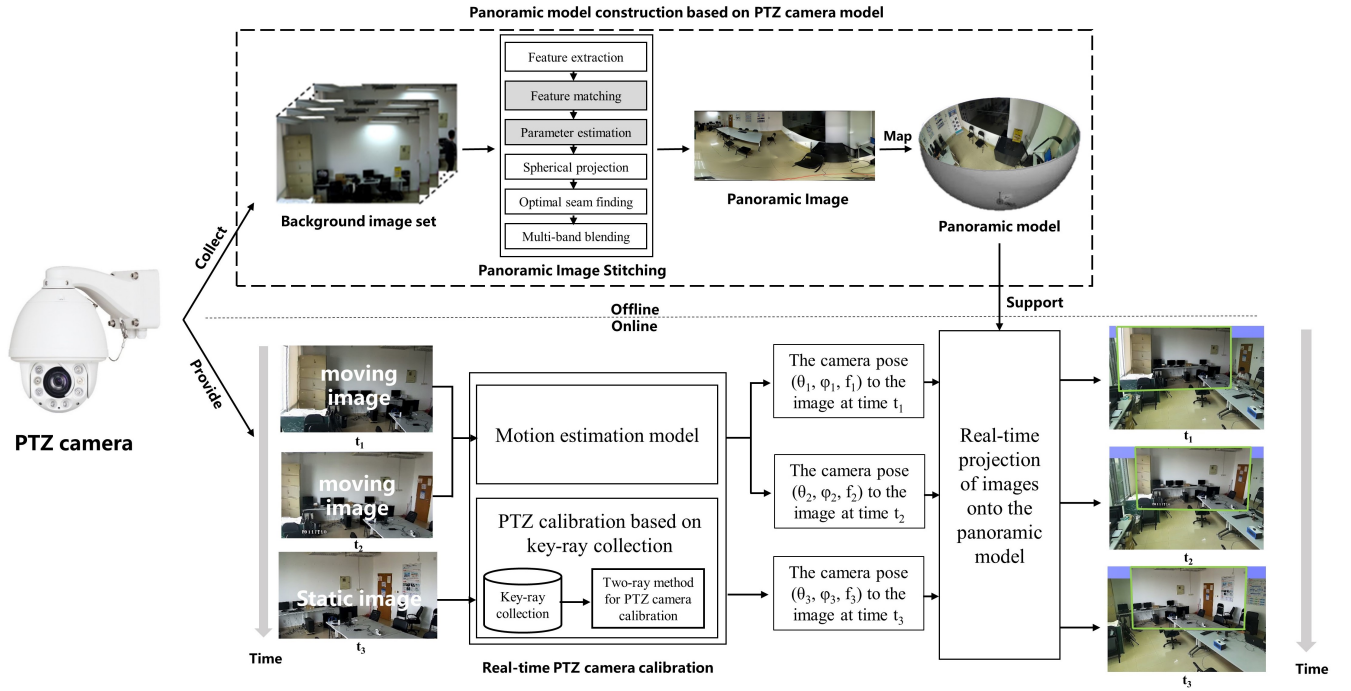


FIGURE 2 Algorithm framework of the visual expansion and real-time calibration for PTZ cameras assisted by panoramic models. PTZ cameras capture background images that are used for panorama stitching. Subsequently, a panorama model is constructed using the generated panoramic image. Finally we register PTZ camera frames onto panorama models through real-time PTZ camera calibration. Grey boxes indicate the corresponding algorithms have been improved to enhance their performance. $t_1 < t_2 < t_3$.

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 the real-time PTZ camera calibration algorithm. Section 6 shows experimental results and Section 7 concludes the paper.

2 | RELATED WORK

2.1 | Camera Calibration

The three primary categories of fixed/portable camera calibration methods are pattern-based calibration, infrastructure-based calibration, and self-calibration. Pattern-based calibration estimates camera parameters by utilizing unique calibration patterns, such as checkerboards^{8,9,10}. This method is typically employed to estimate the internal parameters of cameras. However, it relies on calibration patterns that are challenging to install in outdoor environments. Chen et al.⁷ formulated constraint equations by correlating matching points in images before and after specified camera motions. These motions can also be regarded as a specific pattern. Infrastructure-based calibration employs pointcloud or tilt photography models, establishing correlations between images and these infrastructures to estimate camera parameters. Campbell et al.¹¹ proposed a robust and globally optimal approach for inlier set maximization, which jointly estimates camera pose while accounting for the identification of cross-modality correspondences between 2D image points and 3D points. Zhu et al.²⁰ proposed an incidence field, which represented the rays between points in 3D space and pixels in the 2D imaging plane, to calibrate camera parameter with low-bias and low-variance depth estimation. Xiong et al.²¹ enhanced the calibration performance of the camera using virtual images from 3D scene models. The infrastructure-based method can accurately estimate both intrinsic and extrinsic parameters of cameras, but the construction of infrastructures requires high labor costs. In certain scenarios, such as soccer fields, basketball courts, or hockey arenas, researchers often replace simplified sports field models with point clouds or tilt photography models to reduce labor costs^{22,23,24,25,26,27}. The self-calibration method generates a large number of point correspondences between

adjacent images to estimate external camera parameters. The theory of camera self-calibration, requiring only point matches from image sequences, was first proposed by Faugeras²⁸. Ma¹⁴ performed camera calibration through a sequence of specifically designed motion generated by the active vision system. Luong and Faugeras¹⁵ determined the orientation of an uncalibrated moving camera by using point correspondences among three images and the fundamental matrices derived from these point correspondences. Vasconcelos et al.¹⁶ proposed an automatic camera calibration based on a network of calibrated cameras and only used pairwise points to estimate parameters. Liu et al.¹⁷ presented a novel algorithm for homography computation that enhances precision and reduces 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 a single image. Since these self-calibration methods only require adjacent images for the calibration process, they offer significant advantages over the other two methods. However, most of these methods require at least four feature pairs in adjacent images to calibrate camera effectively, which increases the probability that calibrated results fall under the influence of illumination, occlusion, and texture scarcity.

Unlike the calibration methods employed for fixed/handheld cameras, the calibration of PTZ cameras presents unique characteristics. Since the optical and geometric centers of a PTZ camera are presumed to be known, image localization primarily depends on pan, tilt, and zoom values. This inherent characteristic of PTZ cameras achieves a more effective calibration pipeline by simplifying the relationship between camera movements and image positions. Wu et al.²⁹ present a dynamic calibration algorithm that aligns the current image with a collection of offline-stored features. This approach facilitates the intrinsic and extrinsic parameter estimation of a PTZ camera. However, this technique is suitable only for small angle and focal length variations. Chen et al.³⁰ introduced a two-point calibration method that requires merely two point correspondences to calibrate a PTZ camera, along with a rapid random forest method for predicting pan-tilt angles without the necessity of feature matching. Furthermore, they introduced an online SLAM system based on the two-point method and PT random forest, which can track PTZ cameras in highly dynamic sporting environments³¹. Chen et al.³² employed a Siamese network to excavate compact deep features and incorporated a novel two-GAN model to identify field markings in real images. The sports venues where these technologies are applied exhibit distinct scene characteristics, including clear field boundary lines and minimal variations in lighting. Due to inherent complexity of outdoor environments, such as significant fluctuations in illumination conditions, the application of these technologies is significantly challenging.

The proposed method combines the advantages of infrastructure-based calibration, self-calibration and PTZ calibration methods, introducing an innovative self-calibration approach that employs only two pairs of image points. This method minimizes labor costs, enhances robustness to illumination variations through the simplified input of image point pairs, and enlarges the FOV of the operator.

2.2 | Image Stitching

Extensive research has been conducted on image stitching that stitch multiple images together to provide a broader visual perception of large-scale scenes. Brown and Lowe³³ first proposed a comprehensive framework for image stitching using invariant local features, such as SIFT (scale-invariant feature transform)³⁴, to extract features between images. Meanwhile, bundle adjustment, a photogrammetric technique to combine multiple images of the same scene into an accurate 3D reconstruction, has also been applied to stitch images. The essence of multi-band blending involves merging low-frequency bands across extensive spatial ranges and high-frequency bands within restricted ranges to mitigate exposure discrepancies and misalignments. Gao et al.³⁵ proposed a method for constructing a seamless panoramic image with two predominant planes: a distant background plane and a ground. Zaragoza et al.³⁶ seamlessly bridged inconsistent image regions through the application of Moving Direct Linear Transformation (Moving DLT). The method was designed to support global projection transformation while allowing for local deformation to accommodate potential violations between global and local regions. Li et al.³⁷ introduced a parallax-tolerant method for image stitching, which utilized robust elastic warping to achieve both precise alignment and efficient processing. Li et al.³⁸ developed 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.³⁹ proposed a rapid multi-band blending method to improve the efficiency in panoramic image fusion and stitching. Zhang and Xiu⁴⁰ established an attribute relationship model combining SIFT features and the human visual system (HVS), and used a dynamic process to find the optimal seamlines. Prokop and Polap⁴¹ proposed a hybridization of the classic RANSAC approach with a selected heuristic algorithm like Grey Wolf Optimizer, allowing for image adjustment. Guo et al.⁴² designed a spatial arrangement preservation-based stitching method to stitch farmland remote sensing images, and optimized the image parameters

using bundle adjustment method. Wang et al.⁴³ improved the bundle adjustment method using an exploitable block-diagonal sparsity structure, which sped up the optimization in the context of event-based cameras.

The most of above research addresses the seamline problem resulting from image misalignment, but the core reason lies in the imprecise estimation of image parameters. To address the seamline problem, enhancing the estimation accuracy of image parameters is crucial. We improved the feature matching and parameter estimation stages during panoramic model construction, thereby increasing the accuracy of image parameter estimation and significantly reducing the presence of image seamlines.

2.3 | Augmented Virtual Environment

The augmented virtual environment (AVE) technology represented a real-time image projection technology, rendering realistic-looking textures onto 3D models. Sawhney et al.¹ first introduced a video flashlight system that utilized live video textures captured by a combination of stationary and moving cameras to illuminate a static 3D graphics model. Chen et al.⁴⁴ proposed a visualization framework for surveillance systems that utilized a fixed low-resolution camera for the large-scale display area, while employing a high-resolution camera for the fovea region. Pece et al.⁴⁵ demonstrated the PanoInserts system, which integrated video content into the panorama through a combination of marker-based and image-based tracking techniques. Tompkin et al.² developed a video-collections+context interface by embedding videos into a panorama. They also created a spatio-temporal index tools for the rapid exploration of video collections in both space and time. Zhou et al.⁴⁶ proposed a novel model-guided 3D stitching system based on video models, leveraging single-image modeling technology. Young et al.⁴ introduced a system that facilitated immersive telepresence and remote collaboration on mobile and wearable devices. A live spherical panoramic representation of the real environment was constructed, allowing a remote user to independently select the viewing direction. While these methods enhance the virtual environment using images, it becomes evident that these images also gain richer context within the virtual environment. We try to enlarge spatial contexts of images by virtual environments.

In summary, the augmented virtual environment technology, through the use of 3D models, panoramic images, or video models, effectively expands the field of view of user. This technology can also integrate multi video streams into a unified space. Nevertheless, these methods still encounter challenges in precisely aligning video content with the virtual environment. Moreover, the presence of motion blur complicates the achievement of real-time calibration of videos. To address these challenges, we propose a motion estimation model capable of real-time video alignment with the panoramic model.

3 | TWO-RAY METHOD FOR PTZ CAMERA CALIBRATION

Inspired by Chen's work³⁰, we extend the two-point method from 2D-3D correspondences to 2D-2D correspondences. This modified method is called as the two-ray method. Chen's work primarily focuses on estimating camera pose through 2D-3D correspondences, which relies on known 3D points and their projections in the image. Chen et al. initially obtain the correspondence between image points and 3D points through manual annotation, and then calculate the pan angle, tilt angle, and focal length of the image using the geometric relationships between two pair of image points and 3D points, as shown in Fig. 3(a). However, Chen's work is mainly applied to stadium scenarios, where only the boundary points of the sports field need to be manually marked. In our application scenario, there are numerous objects and complex illumination conditions, very difficult to establish stable 2D and 3D correspondences by manual annotation. Therefore, directly applying Chen's method is not feasible. To overcome this limitation, we propose a new method that requires only two pairs of image points.

We first establish the transformation between the rays and the image pixels, converting the pixels in the image into rays. According to connection between the rays in the overlapping region of adjacent images, we calculate the pose of the PTZ camera.

The PTZ images conform to the geometric properties depicted in Fig. 4. The point **O** represents the optical center of the PTZ camera, and (θ, φ, f) denotes the orientation and focal length of the camera. The coordinate $\mathbf{O}_R(x_{OR}, y_{OR})$ defines the center of the PTZ image within the image coordinate system. The PTZ image is tangent to a sphere with a radius of f . For any point $\mathbf{p}(x_p, y_p)$ within the image coordinate system, the orientation of the corresponding ray \mathbf{r}_p is given by (θ_p, φ_p) . According to the main view and the top view of Fig. 4, we can obtain:

$$\theta_p = \arctan\left(\frac{x_p - x_{OR}}{f \cos(\varphi)}\right) + \theta \quad (1)$$

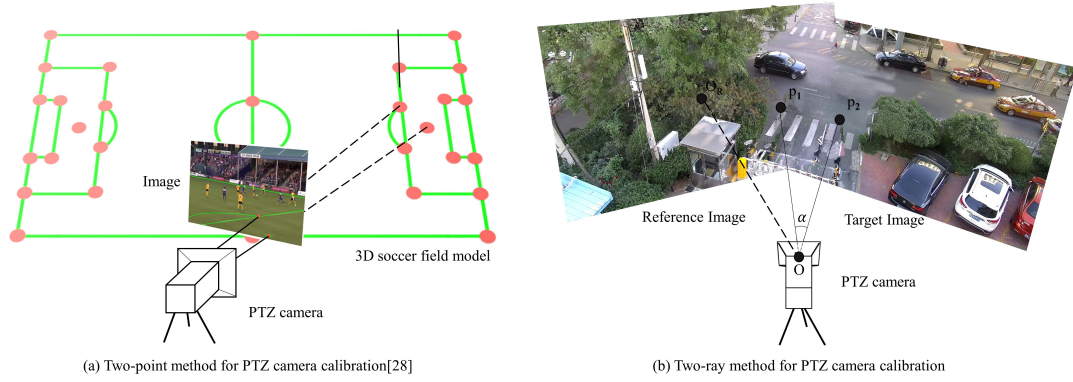


FIGURE 3 Method comparison between the two-point method and the two-ray method. Figure (a) indicates the two-point method for PTZ camera calibration from Chen's work³⁰, and figure (b) represents two-ray method proposed in this paper.

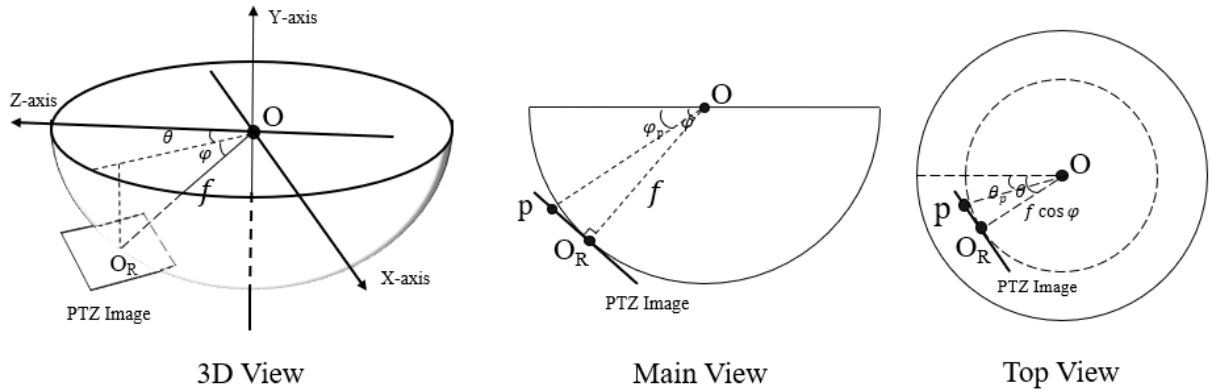


FIGURE 4 The geometric relationship between the PTZ image and the PTZ camera. The main view is obtained from the direction along the negative half axis of the X-axis. And the top view is obtained from the direction along the negative half-axis of the Z-axis.

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

We further analyze the spatial property between adjacent PTZ images, as shown in Fig. 3(b). Both the reference image and the target image are adjacent and have overlapping regions. We assume that the parameters $(\theta_r, \varphi_r, f_r)$ of the reference image are known, and the parameters $(\theta_t, \varphi_t, f_t)$ of the target image are unknown. We select two points \mathbf{p}_1 and \mathbf{p}_2 randomly within the overlapping region. This means \mathbf{p}_1 and \mathbf{p}_2 are both the reference image and the target image. The ray r_{p1} represents the ray from the PTZ camera center to the point \mathbf{p}_1 , and similarly for r_2 . The angle between ray $\mathbf{r}_{p1}(\theta_{p1}, \varphi_{p1})$ and ray $\mathbf{r}_{p2}(\theta_{p2}, \varphi_{p2})$ is denoted as α . The cosine formula for the angle α between two rays is:

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

where the K denotes the internal matrix of the PTZ camera. We calculate \mathbf{r}_{p1} and \mathbf{r}_{p2} in the reference image employing the formula 1 and 2. We assume that PTZ images are free from distortion and that their sizes are known. Consequently, the focal length f_t represents the sole unknown element in K_t of the target image. The focal length f_t can be determined by:

$$f_t^2 = \frac{2(d^2ab - c^2)}{2c - d^2(a + b) + \sqrt{(d^2(a + b) - 2c)^2 - 4(d^2ab - c^2)(d^2 - 1)}} \quad (4)$$

where

$$\begin{aligned} a &= \mathbf{p}_1^T \mathbf{p}_1 \quad b = \mathbf{p}_2^T \mathbf{p}_2 \quad c = \mathbf{p}_1^T \mathbf{p}_2 \\ d &= \cos(\theta_{p1})\cos(\varphi_{p1})\cos(\theta_{p2})\cos(\varphi_{p2}) + \sin(\theta_{p1})\cos(\varphi_{p1})\sin(\theta_{p2})\cos(\varphi_{p2}) + \sin(\varphi_{p1})\sin(\varphi_{p2}) \end{aligned}$$

the parameter f_t is caculated by Eq. 4. The remaining parameters (θ_t, φ_t) of the target image satisfies:

$$K^{-1} \mathbf{p} = \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \mathbf{R}_{\varphi_t} \mathbf{R}_{\theta_t} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_t & -\sin \varphi_t \\ 0 & \sin \varphi_t & \cos \varphi_t \end{bmatrix} \begin{bmatrix} \cos \theta_t & 0 & -\sin \theta_t \\ 0 & 1 & 0 \\ \sin \theta_t & 0 & \cos \theta_t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (5)$$

where $[XYZ]^T$ denotes the coordinates of 3D point corresponding to \mathbf{p} , and $[UV]^T$ represents the 2D coordinates of \mathbf{p} . The ray $r_p(\theta_p, \varphi_p)$ is the ray from the PTZ camera center to the point \mathbf{p} . We have:

$$\frac{X}{\sqrt{X^2 + Y^2 + Z^2}} = \sin \theta_p \cos \varphi_p \quad (6)$$

$$\frac{Y}{\sqrt{X^2 + Y^2 + Z^2}} = \sin \varphi_p \quad (7)$$

$$\frac{Z}{\sqrt{X^2 + Y^2 + Z^2}} = \cos \theta_p \cos \varphi_p \quad (8)$$

We separate $\sin \theta_t$ and $\cos \theta_t$ from the formula 5, which causes:

$$\begin{bmatrix} A_{\varphi_t} & B_{\varphi_t} \end{bmatrix} \begin{bmatrix} \cos \theta_t \\ \sin \theta_t \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (9)$$

where

$$\begin{aligned} A_{\varphi_t} &= \begin{bmatrix} -XU \cos \varphi_t - Z & X - ZU \cos \varphi_t \\ X \sin \varphi_t - XV \cos \varphi_t & Z \sin \varphi_t - ZV \cos \varphi_t \end{bmatrix} \\ B_{\varphi_t} &= \begin{bmatrix} YU \sin \varphi_t \\ Y \cos \varphi_t + YV \sin \varphi_t \end{bmatrix} \end{aligned}$$

From the formula 9, we have:

$$\cos \theta_t = \frac{-YZU + XY(\cos \varphi_t + V \sin \varphi_t)}{\det A_{\varphi_t}} \quad (10)$$

$$\sin \theta_t = \frac{XYU + YZ(\cos \varphi_t + V \sin \varphi_t)}{\det A_{\varphi_t}} \quad (11)$$

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

$$a \tan^2 \varphi_t + b \tan^2 \varphi_t + c = 0 \quad (12)$$

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 Equation 12, The parameter φ_t is calculated up to two possible solutions. One solution is discarded by restricting the valid range to $(0^\circ, 90^\circ)$. The value of θ_t can then be calculated from Equations 10 and 11. Consequently, the parameters $(\theta_t, \varphi_t, f_t)$ of the target image can be estimated using only two rays.

4 | PANORAMIC MODEL CONSTRUCTION BASED ON PTZ CAMERA MODEL

The principle of constructing a panoramic model involves stitching multiple images into a cohesive panorama. The primary processes of image stitching include feature extraction and matching, parameter estimation, spherical projection, optimal seam finding, and multi-band blending. Image stitching often results in the emergence of artifacts, commonly referred to as ghosts. Our analysis indicates that imprecise parameter estimation is a significant contributor to the occurrence of both ghosts and discontinuities. Therefore, we initially optimize the feature matching and parameter estimation processes to enhance the accuracy of parameter estimation. Subsequently, we apply the estimated parameters to stitch the images and generate the panoramic model.

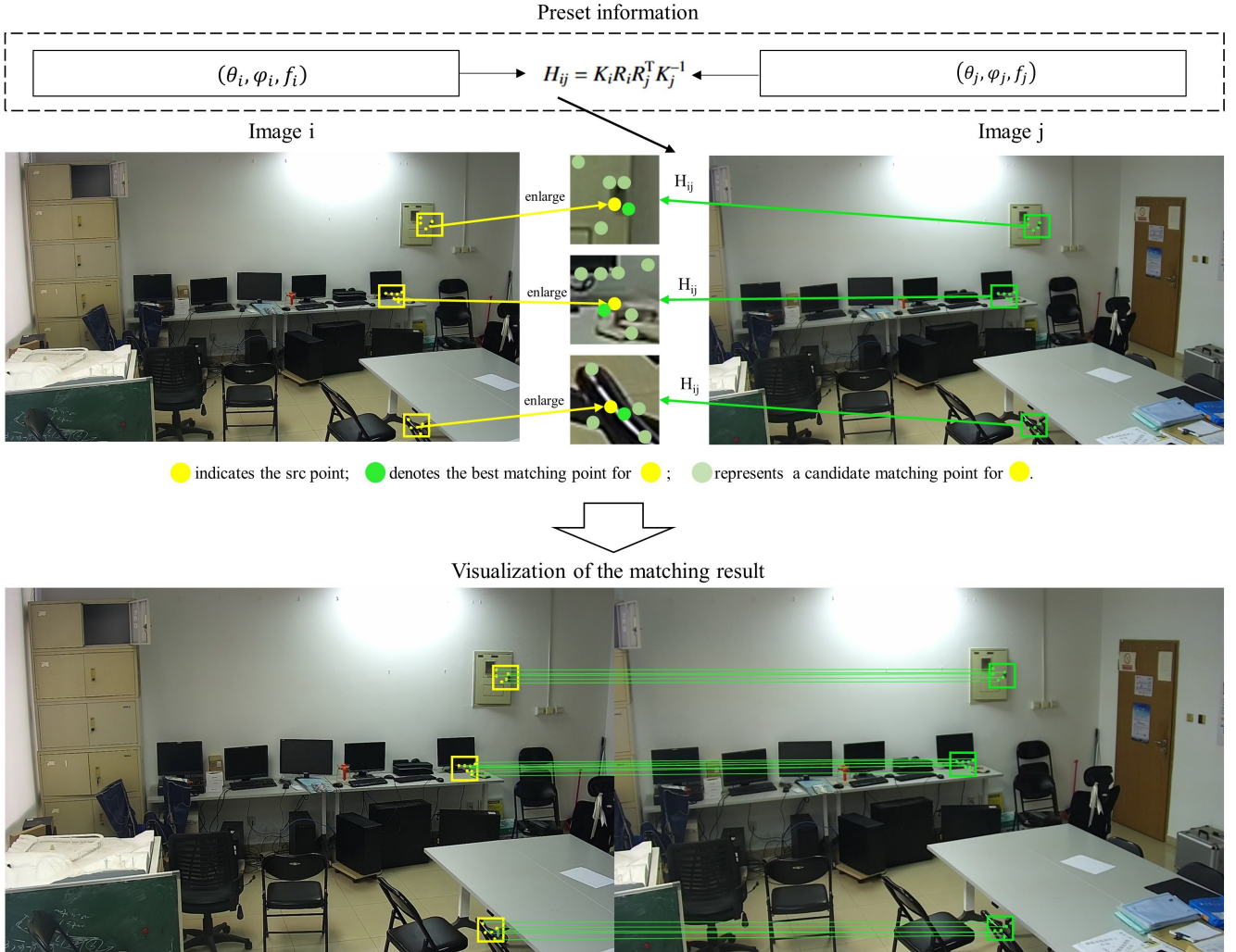


FIGURE 5 The proposed feature matching method. $(\theta_i, \varphi_i, f_i)$ and $(\theta_j, \varphi_j, f_j)$ are image parameters obtained directly from the camera. According to the projection transformation between adjacent images, we can directly calculate the homography matrix H_{ij} to help image j project into image i . We calculate the distance between the feature point and the projection point, and select the closest projection point as the matching point of the feature point.

Feature matching: Feature points are primarily determined by keypoints and their corresponding descriptors. Descriptors are typically derived from the neighboring pixels around keypoints. For instance, a descriptor of an ORB (Oriented FAST and Rotated BRIEF) feature point⁴⁷ is constructed by selecting a certain pattern of N point pairs around the keypoint, and combining the comparison results of these N point pairs to form the descriptor. In contrast, a descriptor of a SIFT feature point is calculated based on positional, scale, and orientation information. The length of descriptors is typically 128 or 256 dimensions. Traditional

feature matching approaches involve calculating the similarity between the descriptors of feature points in two images and establishing matches based on similarity scores. However, because feature descriptors only capture information from the local region around the keypoint, false matches often arise in regions with similar appearance. Some studies such as the Ma's work⁴⁸ utilize the local neighborhood structures of those potential true matches to heighten the accuracy of feature matching. They group multiple keypoints based on their spatial relationships and compute the matching loss at the group level. This approach increases the receptive field of the descriptors, but there can still exist incorrect matches using Ma's work. This is because regions with similar textures often extend beyond the enlarged receptive field of the group, leading to erroneous matches.

For PTZ (Pan-Tilt-Zoom) images, the image preset parameters, such as the pan angle θ , the tilt angle φ and the focal length f , can be obtained directly from the camera during the capture process. While these parameters may be affected by mechanical drifts, the discrepancies are typically small and do not significantly differ from the true image parameters. As a result, we can project adjacent images onto a common plane to find the optimal matching points. Since these images are captured by the same camera, their projection transformations can be considered as unique homography matrixs. The homography matrix between the image I_i and the image I_j is

$$H_{ij} = K_i R_i R_j^T K_j^{-1} \quad (13)$$

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 5. The camera preset parameters and the unique homography matrixs are both the preset information. We leverage preset information to improve the accuracy of feature matching, as illustrated in Figure 5. The proposed feature matching method does not rely on descriptors and will not cause mismatching due to apparent similarity. Using this preset information, we first compute the homography matrix \hat{H}_{ij} . Then, the feature points of image j are projected onto image i . For each feature point in image i , we define a search radius R and look for the closest projected point within this radius. The detailed computation process is:

$$(p_{jk}, p_{jm}) = \arg \min_{\hat{H}_{ij} p_{jm} \in N(p_{ik})} \|\hat{H}_{ij} p_{jm} - p_{ik}\|_2 \quad (14)$$

where p_{ik} denotes the k -th feature point in the image I_i , and p_{jm} denotes the m -th feature point in the image I_j . $N(p_{ik})$ is the set of feature points in the neighborhood of p_{ik} in the image I_i . $\hat{H}_{ij} p_{jm}$ represented the projection of p_{jm} . The radius R is set to 60 pixels. $\|\cdot\|_2$ is Euclidean distance. The feature matching method is illuminated in Algorithm 1.

Algorithm 1 Feature matching

- 1: Calculate the feature points of adjacent image i and image j , and save the feature points to F_i and F_j , respectively.
 - 2: Get image parameters $(\theta_i, \varphi_i, f_i)$ and $(\theta_j, \varphi_j, f_j)$ from the PTZ camera.
 - 3: Calculate the homography matrix H_{ij} from image j to image i according to Eq. 13.
 - 4: Convert the feature points F_j of image j into the projection point set F_{ij} through H_{ij} .
 - 5: Initialize d_{min} to 1000, and record the matching point pair set of F_i as M_i .
 - 6: **for** Select f_m from F_i **do**
 - 7: **for** Select f_n from neighborhood with a radius of R of f_m **do**
 - 8: **if** f_n in F_{ij} **then**
 - 9: Calculate the distance d_{mn} between f_m and f_n .
 - 10: **if** $d_{min} > d_{mn}$ **then**
 - 11: $d_{min} = d_{mn}$.
 - 12: Set $m = (f_m, f_n)$
 - 13: **end if**
 - 14: **end if**
 - 15: **end for**
 - 16: Set m into M_i
 - 17: **end for**
-

Parameter estimation: Currently, several methods of parameter estimation first roughly estimate parameters and then optimize these parameters using the bundle adjustment (BA) method^{33,39,43,42}. Bundle adjustment is a mathematical optimization technique used in photogrammetry to refine 3D reconstructions and camera parameters. It simultaneously minimizes the re-projection error by adjusting both the 3D coordinates of the observed points and the parameters of the cameras. It is also used to solve the camera parameters for single view image coordinate simultaneously by minimizing the sum over all key frames of the residual errors. The residual errors can be represented as:

$$\mathcal{L}(\{(\theta_i, \varphi_i, f_i)\}_{i=1}^n) = \sum_{i=1}^n \sum_{k \in M(i)} \|\mathbf{p}_{ik} - \mathbf{H}_{ij}(\theta_i, \varphi_i, f_i)\mathbf{p}_{jk}\|_2 \quad (15)$$

where n denotes the image number, $M = \{(\mathbf{p}_{ik}, \mathbf{p}_{jk})\}_{k=1}^N$ represents the matched-point pair set. Due to the uncertainty of the parameters for all images, the BA algorithm often finds it difficult to converge during optimization. The optimization result mostly causes an immense drift between the estimation and the ground truth. Moreover, the BA method ignores the geometric property among the PTZ images. To address these limitations, we propose a novel and highly accurate approach for the parameter estimation of PTZ images. A single image is designated as the reference image from the calibration set. Additionally, we incorporate the two-ray method, which we explored in Sec. 3, for parameter estimation.

We select an image from the background image set $S = \{I_1, I_2, \dots, I_n\}$ 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 from image I_i and I_j , and perform the feature matching using the formula 14. 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

$$f_j = \frac{\sum_{k_1=1}^N \sum_{k_2=k_1+1}^N f(r_{ik_1}, r_{ik_2}, p_{jk_1}, p_{jk_2})}{C_N^2} \quad (16)$$

where $f(*)$ is the focal length estimated by the formula 4, 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 10 and 12. We can estimate the θ_j and φ_j of image I_j by

$$(\theta_j, \varphi_j) = \arg \min_{(\theta_{jk}, \varphi_{jk}) \in \Phi} \sum_{m=1}^M \|H_{ij}(\theta_{jk}, \varphi_{jk}, f_j) p_{im} - p_{jm}\|_2 \quad (17)$$

We estimate the $(\theta_j, \varphi_j, f_j)$ of image I_j by the formula 16 and 17. Then we add I_j into S_c , and iterate the above steps until S_u is empty. The Parameter estimation method is presented in Algorithm 2.

Algorithm 2 Parameter estimation

- 1: Divide the background image set S into the calibrated set S_c and the uncalibrated set $S_u = S - S_c$. Put the first background image into the set S_c .
 - 2: **for** Take an image I_i from the set S_c **do**
 - 3: **if** S_u is empty set **then**
 - 4: break.
 - 5: **end if**
 - 6: Find an image I_j from the set S_u that overlaps with I_i .
 - 7: Extract feature points from I_i and I_j .
 - 8: Use the Alg. 1 to perform feature matching.
 - 9: Estimate the parameters of the image I_j using Eq. 16 and 17.
 - 10: Add image I_j to S_c
 - 11: **end for**
-

panoramic model construction: The initial orientation of the PTZ camera is $\theta = 0^\circ$ and $\varphi = 10^\circ$. 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 stitch these background images into the panorama through six stage of feature extraction, feature matching, parameter estimation, spherical projection, optimal seam finding, and multi-band blending. Our approach is used for the feature matching and the parameter estimation stage. subsequently, we project the panorama into the inner surface of the hemispherical model, which generates the panoramic model.

5 | REAL-TIME PTZ CAMERA CALIBRATION ALGORITHM

Real-time calibration requires the dynamic adjustment of PTZ camera poses at regular time intervals. Due to the image blur and illumination changes caused by camera motion, the feature-based method may fail. We determine that the camera motion is traceable and propose a novel real-time calibration algorithm that combines a motion estimation model with a ray-based method. Firstly, We collect motion trajectories for horizontal, vertical, and zoom movements. Secondly, we formulate the motion estimation model over time. The model can estimate the camera poses in real time during PTZ camera motion. Finally, we compute the precise camera pose by utilizing PTZ camera calibration based on key-ray collection when the camera is stationary. More details show as below.

5.1 | A Motion Estimation Model for PTZ Camera

Most PTZ camera models in existing research typically use the standard pinhole camera model with rotation matrices, which represents the static state of the camera. These models are unable to account for variations in PTZ camera parameters during motion. We propose a novel motion estimation model that is continuously updated as a function over time. The model can be quickly formulated through a series of straightforward initialization steps. This section provides a detailed description of the motion estimation model we proposed.

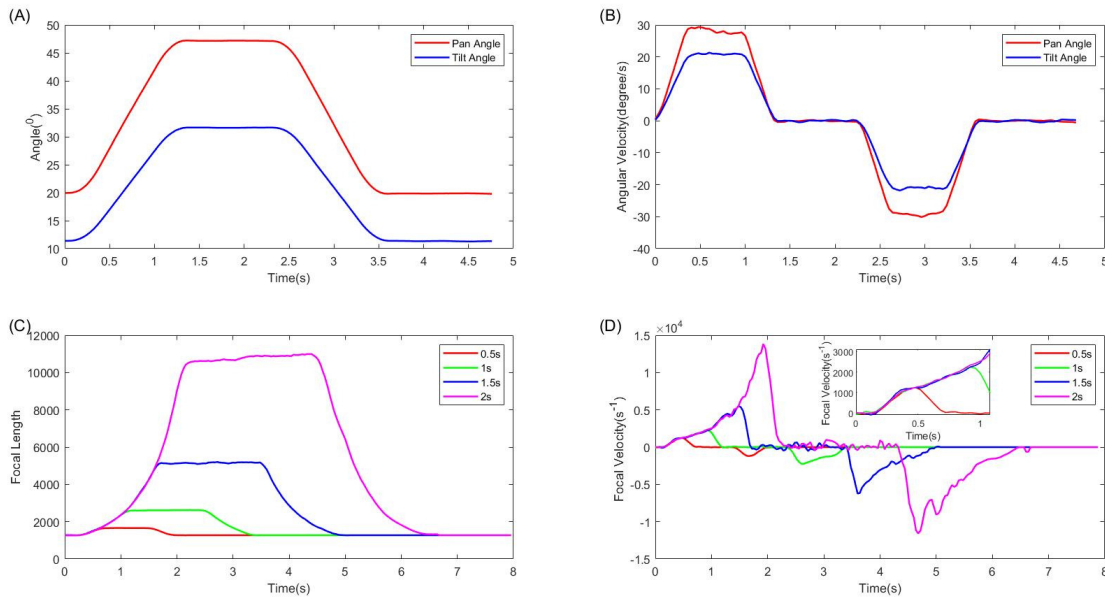


FIGURE 6 The average relationships of the pan angle θ , the tilt angle φ , and the focal length f over time. (A): pan/tilt angle as a function over time; (B): pan/tilt angle velocity as a function over time; (C) focal length as a function over time; (D) focal length velocity as a function over time.

To establish the motion estimation model, we conducted a simple experiment. The PTZ camera was controlled to pan (or tilt) for one second, then stopped and held for one second before returning to its original position. The results of the pan (or tilt) angle variation are shown in fig. 6(A)(or (B)). For the zoom function, the camera was directed to zoom for durations of 0.5, 1, and 1.5 seconds, followed by a return to the original position after a corresponding waiting period. The result of the focal length variation is portrayed in fig. 6(C). To get the accurate value of each frame, a series of images are collected before and after each motion. The parameters of these images were calibrated using the parameter estimation method described in Section 4. The average trends of the pan angle θ , the tilt angle φ , and the focal length f over time are illustrated in Fig. 6.

As evident from the graph of the first row, the rotation of the PTZ camera is through three phases of acceleration, linear motion, and deceleration. In the second row, an one-to-one correspondence is observed between focal length and time during zooming motion of the camera. Consequently, we propose a novel PTZ motion estimation model:

$$\begin{aligned} \{\theta, \varphi, f\} &= h(t \mid v_p, v_t, v_f, t_{p1}, t_{t1}) \\ &= \begin{cases} h_1(t \mid v_p, t_{p1}) \\ h_2(t \mid v_t, t_{t1}) \\ h_3(t \mid v_f) \end{cases} \end{aligned} \quad (18)$$

Where

$$h_1(t \mid v_p, t_{p1}) = \begin{cases} \frac{1}{2} v_p t^2, & (t < t_{p1}) \\ v_p t - \frac{1}{2} v_p t_{p1}, & (t_{p1} < t < t_{p2}) \\ \frac{1}{2} v_p (t + t_{p2} - t_{p1}) \\ + \frac{1}{2} v_p (1 - t)(t - t_{p1}), & (t > t_{p2}) \end{cases} \quad (19)$$

$$h_2(t \mid v_t, t_{t1}) = \begin{cases} \frac{1}{2} v_t t^2, & (t < t_{t1}) \\ v_t t - \frac{1}{2} v_t t_{t1}, & (t_{t1} < t < t_{t2}) \\ \frac{1}{2} v_t (t + t_{t2} - t_{t1}) \\ + \frac{1}{2} v_t (1 - t)(t - t_{t1}), & (t > t_{t2}) \end{cases} \quad (20)$$

$$h_3(t \mid v_f) = \frac{1}{2} v_f t^k + \sqrt{2v_f(f_1 - f_0)}t + f_1 \quad (21)$$

The parameters v_p , v_t , and v_f are constants in the motion estimation model, but they vary for each PTZ camera. t_{p1} and t_{t1} represent the acceleration time, and t_{p2} and t_{t2} denote the constant-speed time, which is determined by the user. f_0 and f_1 represent the focal lengths at $Zoom = 1$ and before zooming, respectively.

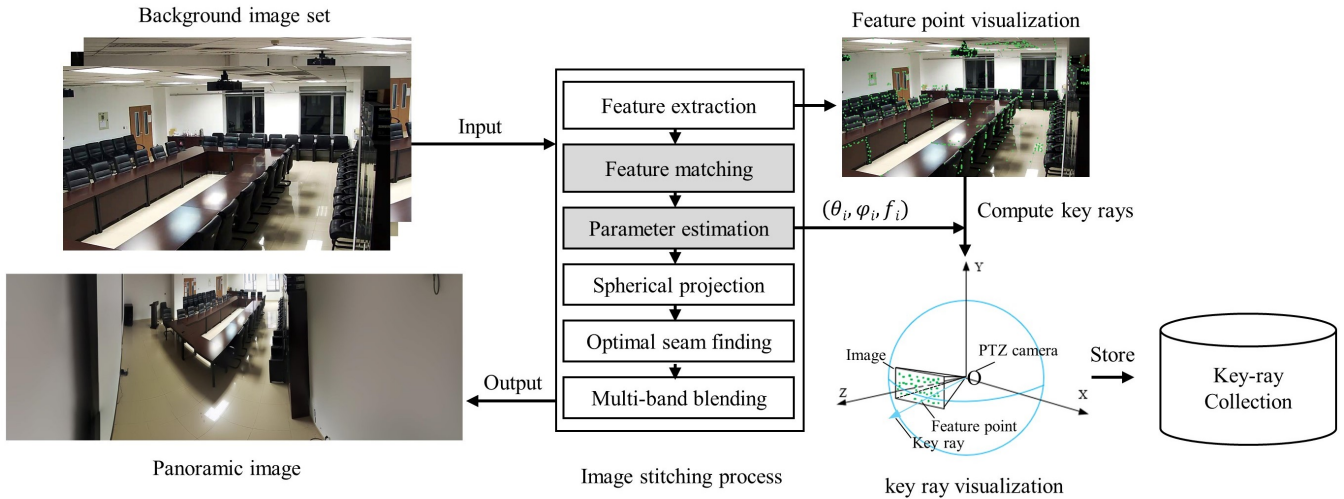


FIGURE 7 Pipeline of the key-ray collection construction. The feature points and image parameters $(\theta_i, \varphi_i, f_i)$ are derived from the image stitching process. The feature points are then converted into key rays, which are stored in the key-ray collection.

5.2 | PTZ Calibration Based on Key-ray Collection

The motion estimation model can predict the motion of the PTZ camera in real time. However, it is unrealistic to expect the results of the model in high precision. We propose a PTZ calibration method to ensure accurate camera calibration. Our approach involves constructing a key-ray collection of the current scene and calibrating online images using the two-ray method and the RANSAC strategy⁴⁹.

The process for constructing the key-ray collection is illustrated in Fig. 7. During the construction of the panoramic image, we extract the SIFT feature sequence and obtain the estimated parameters $(\theta_i, \varphi_i, f_i)$ for each image. The extracted SIFT features are converted into key rays using Formulas 1 and 2. Key rays are rays emitted from the PTZ camera toward the feature points. The feature points and their corresponding key rays are combined as pairs. Feature pairs from different images that highly overlap in both descriptors are merged, ensuring that each feature pair appears only once in the collection. All feature pairs are stored in a collection, forming the key-ray collection of the scene.

A current image is captured from the PTZ camera, and the parameters $(\hat{\theta}, \hat{\varphi})$ of the current image are estimated using the motion estimation model. Key rays around the pose $(\hat{\theta}, \hat{\varphi})$ are extracted from the key-ray collection to form a sub-collection, denoted as M'_r . The set of SIFT features, denoted as M' , is extracted from the current image. Each feature in M' is then matched to the corresponding features in M'_r , resulting in a set of putative matches, denoted as $M_r = (p_k, r_k)_{k=1}^N$, where p_k belongs to M' and r_k belongs to M'_r . The feature matches are computed using Brute-force matching between SIFT descriptors.

The parameters (θ, φ, f) of the current image can be estimated using the idea of RANSAC⁴⁹. Firstly, two feature pairs are selected from M_r , and the two-ray method is applied to calculate the candidate parameters $(\hat{\theta}, \hat{\varphi}, \hat{f})$, which serves as a set of candidate parameters of (θ, φ, f) . The inlier set P of the candidate parameters is

$$P_{inlier} = \{p_k | (\hat{\theta} + \arctan \frac{x_k - u}{\hat{f} \cos \hat{\theta}} - \theta_k)^2 + (\hat{\varphi} + \arctan \frac{y_k - v}{\hat{f}} - \varphi_k)^2 \leq \varepsilon\}_{k=1}^N \quad (22)$$

where $(x_k, y_k)^T$ represents the coordinates of p_k , $(\theta_k, \varphi_k)^T$ represents the coordinates of r_k , and (u, v) denotes the image center of the current image. The accuracy of the candidate parameters is directly proportional to the number of inliers it corresponds to. The above steps are repeated N_s times ($N_s = 1000$), and the candidate parameters with the largest number of inliers are selected as the best estimation for the image parameters. We present PTZ calibration based on key-ray collection in Algorithm 3.

Algorithm 3 PTZ calibration based on key-ray collection

- 1: Collect current frame I from PTZ camera
 - 2: Extract feature points M' from the current frame I
 - 3: Match the feature points between the M' and the key-ray collection M'_r using Brute-force matching, and obtain a matching pair set M_r .
 - 4: Set the maximum number of internal points $N_{max} = 0$.
 - 5: **for** i in $1:1000$ **do**
 - 6: Select two feature pairs from the M_r .
 - 7: Estimate parameter $(\hat{\theta}, \hat{\varphi}, \hat{f})$ using the two-ray method.
 - 8: Calculate the number N_{inlier} of interior points using Eq. 22.
 - 9: **if** $N_{inlier} > N_{max}$ **then**
 - 10: Set $N_{max} = N_{inlier}$, $P_{best} = (\hat{\theta}, \hat{\varphi}, \hat{f})$
 - 11: **end if**
 - 12: **end for**
 - 13: The best parameter estimation of the current frame are P_{best} .
-

6 | EXPERIMENTS

Experiments were conducted to evaluate the proposed model and algorithm using both a public dataset and our dataset. Our dataset was called as Scene dataset. In scene dataset, all cameras used were Hikvision PTZ cameras. on a computer equipped with an Intel i7-9750H CPU, an NVIDIA GeForce GTX 2070M graphics card, and a 16 GB memory.

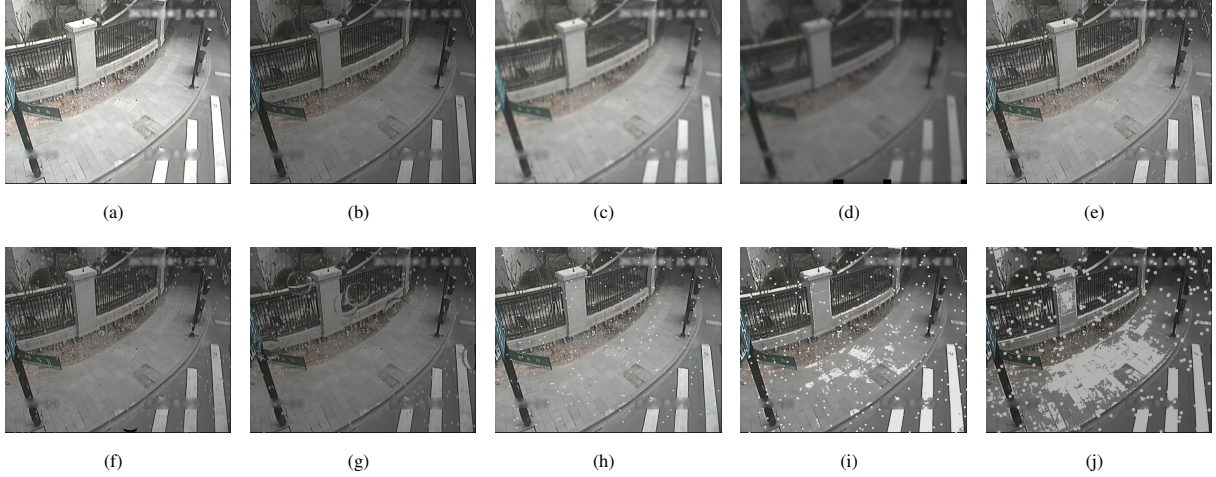


FIGURE 8 Visualization of weather augmentation: (a) brightness; (b) darkness; (c) mist; (d) fog; (e) light rain; (f) moderate rain; (g) heavy rain; (h) light snow; (i) moderate snow; (j) heavy snow.

dataset: There are very few PTZ camera image datasets publicly available. Yong et al.³⁹ 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 constructed Scene dataset including four real scenes for the experiments.

PTZ image dataset: This dataset contains four image groups from various natural traffic scenes, such as intersections and overpasses. The image groups are labeled as 803, 878, 8425, and 8505, respectively. Each image group consists of 52 images. Each group contains 52 images. For detailed information about the dataset, please refer to the original source³⁹.

Data augmentation with weather conditions: We develop a weather augmentation algorithm inspired by the work of Kang et al.⁵⁰. To simulate extreme weather conditions, the RGB (red, green, blue) image is converted to the HLS (hue, saturation, lightness) color space, and the lightness value is adjusted to reflect different weather conditions, as illustrated in Fig. 8.

Brightness and Darkness. By adjusting the lightness values of the images, we can simulate various lighting conditions, such as sunny days and early evenings.

Rainy. We simulate raindrops using randomly generated gray lines (with RGB values of 160, 166, 166), where the size and density of the lines correspond to the intensity and quantity of rainfall. The fisheye effect is applied to distort local regions of the image, simulating the appearance of raindrops on the camera lens. Three types of rainfall are simulated: light rain, moderate rain, and heavy rain.

Snowy. Snowflakes are represented by randomly generated white dots, with the size and density of the dots corresponding to the size and quantity of the snowflakes. Snow on the ground is simulated by extracting pixels with a specific lightness value from the image. Three levels of snowfall are simulated: light snow, moderate snow, and heavy snow.

Foggy. Fog can alter the brightness and sharpness of an image, resulting in a blurred appearance. Fog is simulated as mist and haze using Gaussian blur and adjustments for lightness, which enhance the realism of the images.

Scene Dataset: The scene dataset includes four scenes, as shown in Fig. 9. The first scene "Office" is an $8m \times 6m \times 3m$ laboratory equipped with a Hikvision DS-2DC4223IW-D PTZ camera and furnished with chairs and tables. The second scene "Meeting room" is a $12m \times 8m \times 3m$ meeting room equipped with a Hikvision DS-2DE7172-A PTZ camera and containing conference tables and chairs with minimal features. The "Community" and "Industrial Zone" scenes are $50m \times 50m \times 50m$ outdoor environments, each containing pedestrians and vehicles, and equipped with Hikvision DS-2DC4223IW-D PTZ cameras.




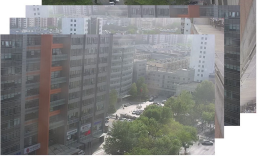
Scene Name	Image dataset	PTZ Camera Type	Dataset Size	Scene Characteristics
Office		Hikvision DS-2DC4223IW-D	36	An office includes tables, chairs, large screens, computers, notice boards, cables.
Meeting room		Hikvision DS-2DE7172-A	36	A meeting room includes conference tables, chairs, cabinets, a podium and a projector.
Community		Hikvision DS-2DC5220IW-A	36	A community scene includes vehicles, pedestrians, trees, and buildings. All images are collected from a PTZ camera at a lower height.
Industrial Zone		Hikvision DS-2DC5220IW-A	36	An industrial zone includes office buildings, shopping malls, banks, structures, trees, vehicles, and pedestrians. All images are collected from a PTZ camera at a higher height.

FIGURE 9 Description of our Scene dataset. Our dataset contains four groups of images, which are obtained from four scenes, and each group consists of 36 views.

Baselines: In the remainder of this section, we compare our method with existing state-of-the-art approaches for image stitching and camera calibration tasks. For the image stitching task, we use the parameter estimation method from the work of Wang⁴³, known as the BAY method, as a benchmark. This method has also been applied in many studies^{39,42}. This method optimizes only the three parameters: pan angle, tilt angle, and focal length. The BAY method aims to minimize the residual error. This approach may produce abnormal results. Therefore, we modify the BAY method so that the first input image does not participate in the optimization process. We refer to the modified BAY method as the BAYs method. The core idea of the BAYs method is to optimize the other images using the first image as a reference. We use the BAYs method as the other benchmark. For the calibration task, we compare our approach with several PTZ camera calibration methods that directly compute the camera parameters using feature point pairs: dynamic calibration²⁹, the two-point method³⁰, and the TPCM method¹⁸, the LAC method⁵¹.

Metrics: In the absence of ground truth, we report parameter errors as the Euclidean distance between the results of 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. Therefore, large offsets indicate that the method produces incorrect parameter estimations. In addition, we use the results of parameter estimation to construct a homography matrix and compute the reprojection error for each feature pair between adjacent images. The score of the reprojection error reflects the precision of parameter estimation.

6.1 | Experiment Results

Accuracy analysis: We first validate our parameters estimation algorithm on the PTZ image dataset. The experimental results are presented in Table 1. The first input image is taken as the reference image. Assuming the (θ, φ) parameters are known, we estimate only the focal length of the reference image using the bundle adjustment method. The estimated focal length is used as the focal length of the reference image. In each scene, we then select three consecutive images with different tilt angles and perform parameter estimation using our method, the BAY method, and the BAYs method. Table 1 shows that, although the

TABLE 1 Experiment comparison of different parameter estimation methods on the PTZ image dataset. **BOLD/BLUE** indicates the best/second-best performance.

Scene label	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	BAY ⁴³	267.83	267.78	267.83	3.53	2.58
				BAYs	12.24	12.25	12.26	9.63	6.10
				Ours	8.72	8.76	8.75	7.82	6.06
	27	28	29	BAY ⁴³	293.72	293.70	293.73	7.28	6.76
				BAYs	13.01	13.01	13.05	10.64	7.66
				Ours	10.41	10.43	10.46	10.43	5.52
	41	42	43	BAY ⁴³	1572.43	1572.41	1572.44	29.17	17.13
				BAYs	0.00	1.60	2.70	35.74	14.98
				Ours	7.56	7.87	7.87	9.45	7.01
878	1	2	3	BAY ⁴³	293.15	293.09	293.13	4.74	2.07
				BAYs	1.46	1.60	1.67	6.14	5.03
				Ours	4.65	4.71	4.72	5.65	4.85
	27	28	29	BAY ⁴³	221.88	221.86	221.87	9.85	4.54
				BAYs	0.76	1.29	1.56	17.79	7.00
				Ours	3.59	3.94	4.08	6.65	4.54
	41	42	43	BAY ⁴³	881.46	881.58	881.77	5.87	11.46
				BAYs	37.58	37.58	37.58	21.86	8.12
				Ours	8.92	9.31	9.46	6.59	6.93
8425	1	2	3	BAY ⁴³	242.38	242.34	242.40	1.27	1.45
				BAYs	9.47	9.47	9.48	6.29	4.58
				Ours	8.17	8.19	8.18	4.54	4.12
	27	28	29	BAY ⁴³	385.76	385.73	385.76	13.70	2.77
				BAYs	1.52	1.70	1.67	10.25	8.42
				Ours	1.42	2.04	1.81	10.19	5.80
	41	42	43	BAY ⁴³	163.46	163.53	163.69	3.14	9.28
				BAYs	22.95	23.01	22.97	4.59	8.57
				Ours	9.61	9.62	9.71	16.37	22.32
8505	1	2	3	BAY ⁴³	230.45	230.41	230.46	2.79	2.68
				BAYs	1.54	1.55	1.57	4.85	5.16
				Ours	5.79	5.81	5.80	4.72	4.65
	27	28	29	BAY ⁴³	357.70	357.66	357.69	8.23	5.98
				BAYs	1.12	1.30	1.56	11.12	6.17
				Ours	4.16	4.21	4.24	5.74	4.71
	41	42	43	BAY ⁴³	1559.39	1559.34	1559.35	16.50	11.65
				BAYs	33.92	33.94	33.93	25.66	11.48
				Ours	8.91	9.21	9.22	10.82	8.63
All scenes	\	\	\	BAY ⁴³	539.13	539.11	539.17	8.83	6.57
				BAYs	11.29	11.52	11.66	13.71	7.77
				Ours	6.82	7.00	7.02	8.24	6.26

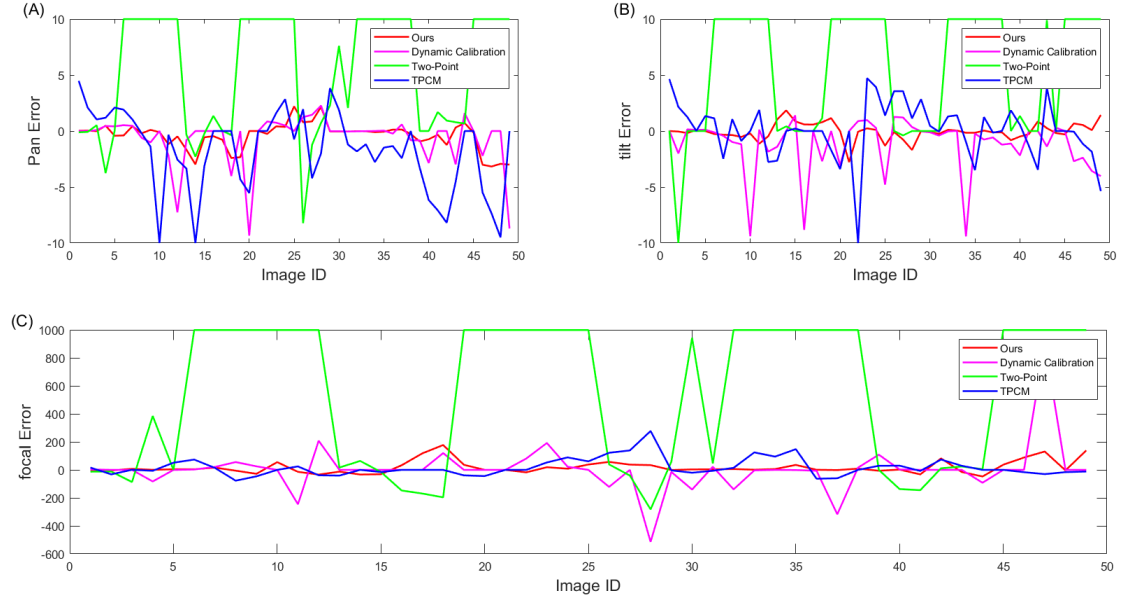
TABLE 2 Reprojection errors of different parameter estimation methods in Scene dataset. **BOLD** indicates the best performance.

Image number		Method	Scene			
Img 1	Img 2		Office	Meeting room	Community	Industrial Zone
1	2	Ours	4.29	6.25	4.90	5.19
		BAY ⁴³	6.19	8.33	6.78	8.84
		BAYs	10.17	8.20	12.10	10.24
1	5	Ours	2.72	2.28	2.91	3.80
		BAY ⁴³	4.31	4.06	4.36	4.64
		BAYs	9.14	7.96	8.40	7.33

BAY method achieves the smallest reprojection errors in many conditions, its offset is significantly larger than that of the other two methods. One reason for this phenomenon is the lack of a reference image during the optimization stage, while another is the focus on minimizing the reprojection error. The BAYs method overcomes this limitation, but its accuracy is inferior to that of our method. Our method demonstrates excellent performance in both parameter estimation offset and reprojection error, indicating that it outperforms the other two methods in parameter estimation. Additionally, we computed the average error

TABLE 3 Experiment results of several PTZ calibration methods on four scenes. **BOLD** indicates the best performance. '-' indicates estimation failure.

Scene	Parameter Error	Method					
		Dynamic calibration ²⁹	Two-point method ³⁰	TPCM ¹⁸	LAC ⁵¹	Our motion estimation model	Our PTZ calibration method
Office	Pan angle	3.45	10.39	6.54	-	0.35	0.12
	Tilt angle	5.61	12.56	1.23	-	0.14	0.09
	Focal length	198.73	876.55	56.23	318.45	167.33	23.65
	FPS	0.3	0.6	1.5	1.9	60	2.3
Meeting room	Pan angle	2.67	8.56	4.33	-	0.47	0.23
	Tilt angle	3.21	10.68	1.43	-	0.33	0.18
	Focal length	166.78	1009.56	86.52	456.12	67.45	33.56
	FPS	0.4	0.5	1.3	2.1	60	2.4
Community	Pan angle	3.56	7.63	5.63	-	0.35	0.28
	Tilt angle	2.35	12.79	3.82	-	0.22	0.16
	Focal length	265.89	1256.33	197.68	556.33	162.455	56.45
	FPS	0.3	0.4	1.4	1.3	60	1.7
Industrial zone	Pan angle	6.18	6.23	6.77	-	0.27	0.13
	Tilt angle	4.58	8.33	4.12	-	0.18	0.11
	Focal length	312.77	1123.05	332.54	443.23	183.53	48.12
	FPS	0.2	0.4	1.2	1.1	60	1.3

**FIGURE 10** Parameter error visualization of several PTZ calibration methods on the office scene. (A) The error in the pan angle θ . (B) The error in The Tilt angle φ . (C) The error in the focal length f .

across all scenarios. Across all scenarios, our method consistently achieves the smallest reprojection error and parameter offset, demonstrating its superiority.

We also evaluate the performance of our parameter estimation methods using Scene dataset. We select the first and second images, as well as the first and fifth images, from each scene for parameter estimation. The first and second images have a horizontal overlapping region. The first and fifth images have a vertical overlapping region. These images are typical representatives for image stitching tasks. In this experiment, we focus solely on testing the reprojection error. The experimental results are shown in Table 2. The experimental results show that our method significantly outperforms the other two methods. Compared to the PTZ image dataset, our scenes includes more texture-less regions and more pronounced illumination variations. These regions are where traditional feature matching methods often fail, leading to poorer performance of the BAY and BAYs methods.

We evaluated the Dynamic Calibration, Two-Point Method, TPCM, LAC, our motion estimation model, and our PTZ calibration method across four scenes: Office, Meeting room, Community, and Industrial zone. The camera was controlled to move horizontally and vertically for one second in each scene, repeated 20 times. We manually aligned the camera video frames with background images to obtain the parameters of each image as ground truth. The experimental results are presented in Table 3. The experimental results indicate that our PTZ calibration method achieved the best performance. As the scene transitions from indoor to outdoor, the parameter errors of the other methods increase significantly, while our PTZ calibration method remains stable. This stability can be attributed to the use of fewer feature points to estimate the pose, which reduces the influence of environmental factors, such as illumination. The LAC method solves the focal length of the PTZ camera through the homography of multiple images, but cannot estimate the orientation of the camera. Since the LAC method relies on the offset between the camera center and the rotation center, the focal length calculated by the LAC method has a large error when the offset is too small. Given these limitations, we will exclude the LAC method from further discussion.

We collected 50 images at the "Office" scene, capturing different pan, tilt, and zoom values, and visualize the parameter estimation errors using these images. Since these three methods rely on accurate feature correspondence, we first eliminate obviously mismatched pairs using RANSAC. Figure 10 illustrates the biases between these methods and the ground truth. The reported bias for each parameter is computed by $\|param_{est} - param_{ground-truth}\|_2$. From the results, we analyze that the two-point method exhibits a large bias because it relies on a random forest that depends on trained images. The more images the random forest is trained on, the better the predicted results become. Since we only used the background dataset to train the random forest, the predicted results exhibit significant errors. The TPCM method amplifies errors through multiple additions and multiplications, leading to significant fluctuations in the results.

Robustness analysis: We further assess the robustness of our approach under extreme weather conditions. We selected two images, numbered 18 and 19, from different weather conditions and scenes, and applied our method, the BAY method, and the BAYs method to estimate their parameters. The reprojection error between the images was then calculated. The experimental results are presented in Table 4.

TABLE 4 Reprojection errors of parameter estimation in extreme weather. **BOLD** indicates the best performance, **RED** indicates the enormous error, and '-' indicates estimation failure.

Scene	Method	Original tion	Bright ness	Dark ness	Light Rain	Mod. Rain	Heavy Rain	Light Snow	Mod. Snow	Heavy Snow	Mist	Fog
803	BAY ⁴³	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
	Ours	7.29	8.25	10.90	7.19	24.46	5.63	8.01	6.52	12.17	9.52	5.53
878	BAY ⁴³	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
	Ours	14.72	8.28	14.91	13.80	15.95	15.76	13.50	21.64	16.99	19.39	31.80
8425	BAY ⁴³	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
	Ours	4.73	4.74	5.05	4.68	6.35	5.75	4.98	3.40	12.31	4.73	4.25
8505	BAY ⁴³	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
	Ours	6.08	6.43	5.97	5.99	6.14	3.75	6.58	12.09	17.44	6.34	21.50

As shown in Table 4, all three methods demonstrate outstanding performance under conditions of brightness, darkness, light rain, and light snow. This indicates that the three algorithms are capable of mitigating interference caused by minor weather changes. However, as weather severity increases, such as in moderate rain and mist, the BAY method and the BAYs method begin to produce incorrect estimates in the 878 scene. Under conditions of moderate snow, heavy rain, heavy snow, and heavy fog, the BAY and BAYs methods nearly fail, either producing completely incorrect estimates (e.g., the 8505 scene in moderate snow and the 878 scene in fog), or failing to estimate the parameters altogether (e.g., the 878 scene in heavy rain and the 8425 scene in heavy snow). In contrast, our method maintains strong robustness under these weather conditions. Even when the other methods fail, the estimation results of our method show only a few pixel offsets compared to the normal case (e.g., the 878 scene in heavy rain and the 8425 scene in heavy snow).

Real-time performance analysis: We evaluated the real-time performance analysis of different methods, as shown in Tab. 3. The frame rate of PTZ camera was set to 60 frames per second. As shown in the table, our motion estimation model achieves a

frame rate of 60 frames per second, with predicted errors close to the ground truth. In contrast, other methods can only achieve frame rates of 1 to 2 frames per second, failing to meet the real-time requirements. To balance the real-time requirements with the effectiveness of the method, we combine the motion estimation model with the PTZ calibration method. When the model is in motion, the pose is estimated using the motion model. When the model is stationary, the PTZ calibration method is invoked to calibrate the camera.

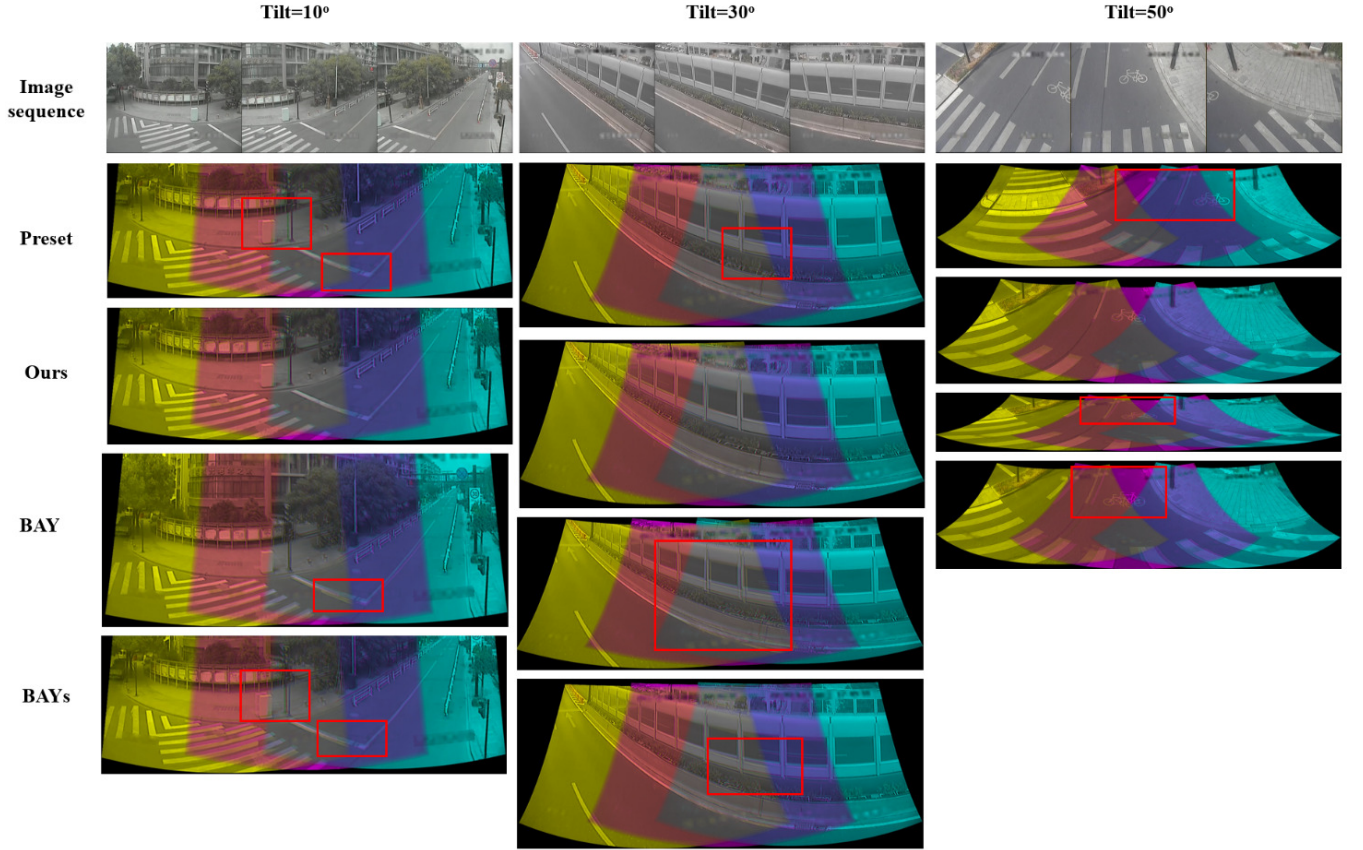


FIGURE 11 Result visualization of parameter estimations in PTZ image dataset. The image sequence in the first column consists of images (1, 2, 3) from the 803 scene, the second column contains images (27, 28, 29) from the 8505 scene, and the third column includes images (40, 41, 42) from the 878 scene. Yellow represents the first image, green represents the third image, and other colors denote the overlapping areas between adjacent images. Red boxes highlight significant misalignments.

6.2 | Experiment Visualization

Visualization of accuracy experiments: We project several of the parameter estimation results from Table 1 onto a spherical surface, as shown in Fig. 11. Compared to the other two methods, our approach effectively eliminates artifacts caused by inaccurate preset parameters (Red boxes in Fig. 11). Additionally, the result of the BAY method in the third column differs from the preset value in terms of image size. This indicates that the image size of the result from the BAY method changed due to the lack of a reference in the optimization stage. These findings are consistent with the results presented in Table 1.

We visual the calibration results from our PTZ camera calibration algorithm on four scenes, shown as Fig. 12. In Fig. 12, we observe that our approach still accurately projected the image onto the panoramic model with the less-texture regions from the second row. In Outdoor scenes, more features mean more interference. However, our method gains satisfactory calibration results demonstrating the good application of our method in various scenes. In addition, our method can effortlessly reach the target area with the support of the panoramic model, which is a challenging task for other AVE method yet.

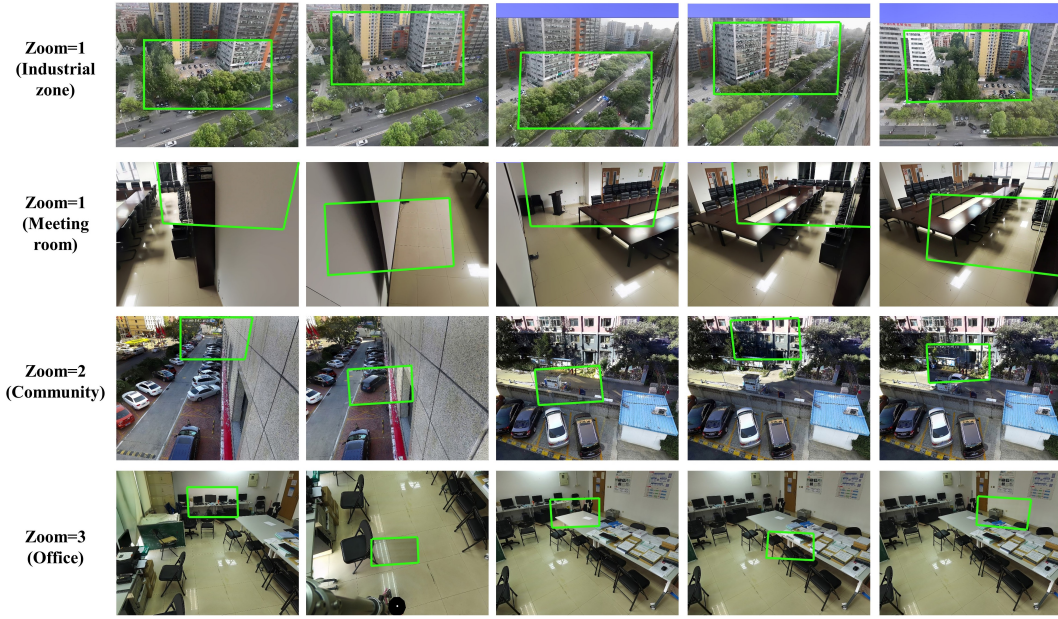


FIGURE 12 The visualization results of our PTZ camera calibration algorithm. The green box indicates the projection of PTZ images on panoramic models. Regions outside green boxes are panoramic models.

Visualization of robustness experiments: We also collected images from the outdoor scene "Industrial Zone" at 2:00 p.m. and 6:00 p.m. to evaluate the robustness of illumination on camera calibration. We selected three groups of images: (1) images with minimal changes in illumination (pan = 0° , tilt = 10°), (2) images with local changes in illumination (pan = 60° , tilt = 10°), and (3) images with significant variations in illumination (pan = 80° , tilt = 10°). All images within each group shared the same camera pose. We estimated the parameters for each of these groups using Dynamic Calibration, the Two-Point method, TPCM, and our method, followed by projecting the images onto a spherical surface using the estimated parameters. The experimental results are presented in Fig. 13. Since the two images share the same pose, they should be projected onto the same region. Yellow is used to indicate images captured at 6:00 p.m., and purple represents images taken at 2:00 p.m. The overlapping region between the two images is highlighted in light red. If the estimation is correct, only the light red area should be visible. Inaccurate estimations will result in visible yellow or purple areas. The reprojection errors of matched pairs were also provided between images for each projection. When the illumination change is subtle (shown in the first column of Fig. 13), all four methods produce acceptable parameter estimates with minimal reprojection errors. When local illumination changes (shown in the second column of Fig. 13), the Dynamic Calibration method produces significant bias (yellow areas in the third row), while both the Two-Point method and the TPCM method show biased results (purple area in the fourth row and yellow area in the fifth row). In contrast, our method consistently produces accurate parameter estimation results. When the illumination difference is significant (shown in the third column of Fig. 13), the Dynamic Calibration method, the Two-Point method, and the TPCM method all fail (yellow and purple areas in the figure). However, our method still produces acceptable results, with only a minor increase in reprojection error compared to the first column.

7 | CONCLUSION

We propose a novel visual expansion and real-time calibration for PTZ cameras assisted by panoramic models. Firstly, we extend the two-point method to the two-ray method, which is then applied to parameter estimation. Secondly, we improve the feature matching and parameter estimation in the image stitching method by leveraging the geometric properties of adjacent PTZ images. The improved method achieves superior parameter estimation during the spherical stitching of PTZ images. We construct the panoramic model based on the improved method. Besides, we present a real-time PTZ camera calibration algorithm, which primarily consists of a PTZ motion estimation model and a camera calibration algorithm based on key-ray collection. We

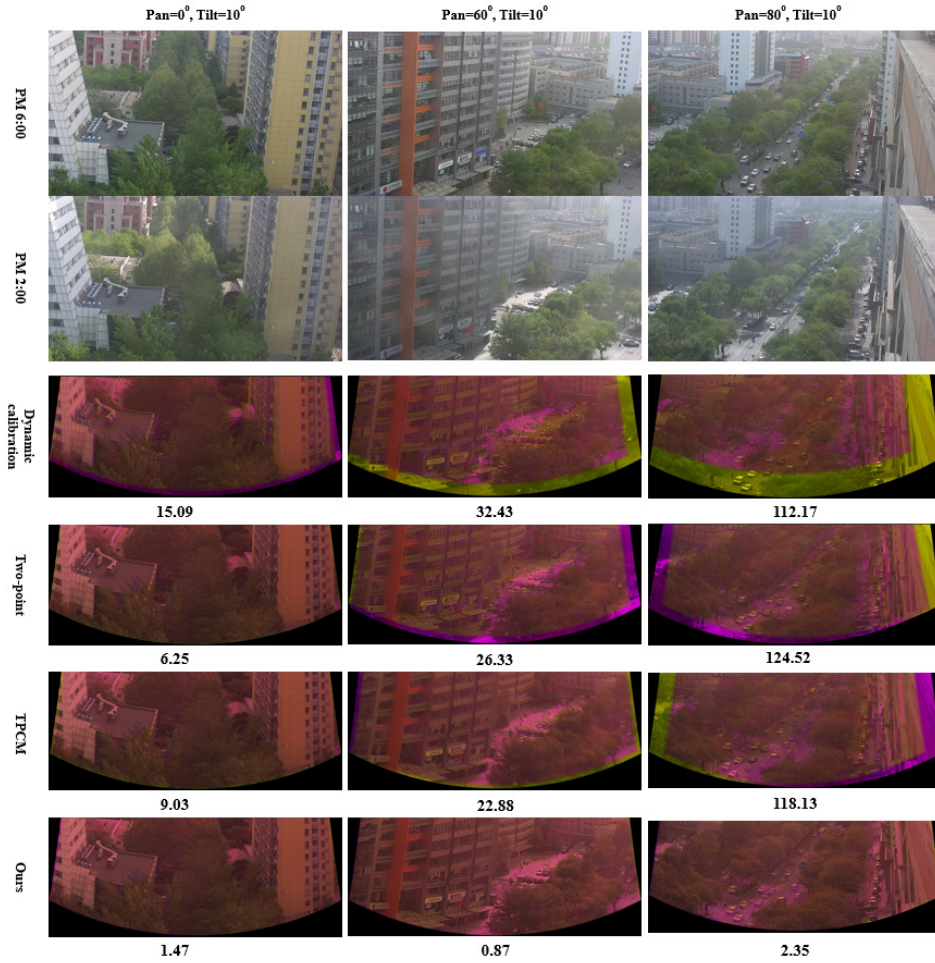


FIGURE 13 PTZ calibration results under different illumination conditions. The first column indicates that the condition with few substantial changes in illumination; The second column indicates the condition with local changes in illumination; The third column indicates that the condition with entirely 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.

validate our method using both public and our datasets. Experimental results demonstrate that our method exhibits outstanding performance.

However, there are several limitations in this work. We have not yet considered the lens distortion, which may introduce biases during PTZ camera calibration. Additionally, we assume that the principal point coincides with both the projection center and the zoom center. However, this assumption may not hold uniform for zooming cameras. Finally, while we presume that the optical center of the PTZ camera is fixed, it would have variations when the camera rotates.

AUTHOR CONTRIBUTIONS

Liangliang Cai: conceptualization (lead), methodology, investigation, software, validation, formal analysis, investigation, original draft (lead), writing – review and editing (lead), visualization. **Zhong Zhou:** conceptualization (supporting), resources, data curation, writing – original draft (supporting), writing – review and editing (supporting), supervision, funding acquisition.

FUNDING INFORMATION

This work was supported in part by the Science and Technology Project of Hainan Provincial Department of Transportation under Grant HNJT- 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: ACM International Conference Proceeding Serie. 2002:157–168.
2. Tompkin J, Pece F, Shah R, Izadi S, Kautz J, Theobalt C. Video collections in panoramic contexts. In: Proceedings of the 26th annual ACM symposium on User interface software and technology. 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: 5th International Conference on Intelligent Human-Machine Systems and Cybernetics. 2013:389–392.
6. Long DS. Review of Camera Calibration Algorithms. In: Advances in Computer Communication and Computational Sciences. 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: IEEE Conference on Computer Vision and Pattern Recognition. 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 International Conference on Image Processing (ICIP). 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: European Conference on Computer Vision. 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. *Journal Of The Optical Society Of America A-optics Image Science And Vision*. 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. Zhu S, Kumar A, Hu M, Liu X. Tame a Wild Camera: In-the-Wild Monocular Camera Calibration. In: Advances in Neural Information Processing Systems (NeurIPS). 2023:45137–45149.
21. Xiong Y, Wang J, Zhou Z. VirtualLoc: Large-scale visual localization using virtual images. *ACM Transactions on Multimedia Computing, Communications and Applications*. 2023;20(3):1–19.
22. 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.
23. Homayounfar N, Fidler S, Urtasun R. Sports field localization via deep structured models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:5212–5220.
24. Sharma RA, Bhat B, Gandhi V, Jawahar C. Automated top view registration of broadcast football videos. In: IEEE Winter Conference on Applications of Computer Vision (WACV). 2018:305–313.
25. Rematas K, Kemelmacher-Shlizerman I, Curless B, Seitz S. Soccer on your tabletop. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:4738–4747.
26. 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.
27. Sha L, Hobbs J, Felsen P, Wei X, Lucey P, Ganguly S. End-to-end camera calibration for broadcast videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:13627–13636.
28. Faugeras OD, Luong QT, Maybank SJ. Camera self-calibration: Theory and experiments. In: European conference on computer vision. 1992:321–334.
29. Wu Z, Radke RJ. Keeping a pan-tilt-zoom camera calibrated. *IEEE transactions on pattern analysis and machine intelligence*. 2012;35(8):1994–2007.
30. Chen J, Zhu F, Little JJ. A two-point method for PTZ camera calibration in sports. In: IEEE Winter Conference on Applications of Computer Vision (WACV). 2018:287–295.
31. Lu J, Chen J, Little JJ. Pan-tilt-zoom SLAM for Sports Videos. *arXiv preprint arXiv:1907.08816*. 2019.
32. Chen J, Little JJ. Sports camera calibration via synthetic data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019:1–8.
33. Brown M, Lowe DG. Automatic panoramic image stitching using invariant features. *International journal of computer vision*. 2007;74(1):59–73.
34. Lowe DG. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*. 2004;60(2):91–110.

35. Gao J, Kim SJ, Brown MS. Constructing image panoramas using dual-homography warping. In: IEEE Conference on Computer Vision and Pattern Recognition 2011:49–56
36. Zaragoza J, Chin TJ, Brown MS, Suter D. As-Projective-As-Possible Image Stitching with Moving DLT. In: IEEE Conference on Computer Vision and Pattern Recognition 2013.
37. 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.
38. 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.
39. 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.
40. Zhang J, Xiu Y. Image stitching based on human visual system and SIFT algorithm. *The Visual Computer*. 2024;40(1):427–439.
41. Prokop K, Polap D. Heuristic-based image stitching algorithm with automation of parameters for smart solutions. *Expert Systems with Applications*. 2024;241:122792.
42. Guo S, Gallego G. Event-Based Mosaicing Bundle Adjustment. In: European Conference on Computer Vision. 2025:479–496.
43. Wang J, Du P, Yang S, Zhang Z, Ning J. A Spatial Arrangement Preservation Based Stitching Method via Geographic Coordinates of UAV For Farmland Remote Sensing Image. *IEEE Transactions on Geoscience and Remote Sensing*. 2024.
44. Chen SC, Lee CY, Lin CW, et al. 2D and 3D visualization with dual-resolution for surveillance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012:23–30.
45. Pece F, Steptoe W, Wanner F, et al. Panoinsets: mobile spatial teleconferencing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2013:1319–1328.
46. 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.
47. Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. In: IEEE International conference on computer vision. 2011:2564–2571.
48. Ma J, Zhao J, Jiang J, Zhou H, Guo X. Locality preserving matching. *International Journal of Computer Vision*. 2019;127(5):512–531.
49. 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.
50. 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.
51. Liu Y, Zhang H. Linear Auto-calibration of Pan-Tilt-Zoom Cameras With Rotation Center Offset. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). 2023:11461–11467.

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.