Facilitating PTZ Camera Auto-Calibration to Be Noise Resilient With Two Images

RUI WANG [ID]1, RAN HUANG [ID]1, AND JIAORU YANG2

[1]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China
[2]China Aerospace Science and Technology Corporation, Beijing 100048, China

Corresponding author: Rui Wang (wangr@buaa.edu.cn)

ABSTRACT This paper presents a simple but effective technique to calibrate a PTZ (Pan/Tilt/Zoom) camera by using only two images for five intrinsic parameters: focal length, aspect ratio, the principle point coordinates and the distortion coefficient. In our approach, the SCC-SURF (Shape-Color Combined SURF) descriptor is first employed to obtain robust point correspondences in a pair of color images taken before and after the camera undergoing an arbitrary pan-tilt rotation respectively. Based on the radial lens distortion division model, the point correspondences between these two images are applied to calculate the homography and the distortion coefficient simultaneously. The estimated homography is proved more precise with our novel framework CWRLD (Covariance Weighted Ransac under Lens Distortion), which employs a covariance matrix in the presence of feature location noise. Finally, the remaining four intrinsic parameters are solved using directly decomposing estimated homography with a series of Givens rotations. Both synthetic and real data are provided to verify that our proposed technique is precise, convenient, and applicable for online calibration without regard for a specific imaged environment.

INDEX TERMS PTZ camera calibration, SCC-SURF, lens distortion, CWRLD, Givens rotation.

I. INTRODUCTION

Owing to their pan, tilt and zooming abilities, PTZ(Pan/Tilt/Zoom) cameras, which can observe a larger field of view and act as high-resolution sensors, have a wide range of applications such as object tracking [1], intelligent teaching systems [2],as well as in early detection of building destruction [3]. However, the dynamic changes in internal and external parameters have made it difficult to precisely determine spatial point coordinates associated with the world coordinate system, and traditional off-line calibration (3D/2D/1D object-based) methods [5-8] are no longer suitable for this situation. Therefore, it is necessary to find a fully automatic online calibrating method which needs no calibration targets or specific structures in a scene for the PTZ camera as an omnidirectional sensor.

In many PTZ camera applications, self-calibration, first introduced by Faugeras *et al.* [9] in the early 1990s, is employed as one of the most popular methods. Compared with other means, rotation-based self-calibration approaches have remained as a major concern for a PTZ camera, which

can pan/tilt around its own optical center. For example, Hartley [10] used rotational motion to obtain overlapping images, and computed homography matrix by obtaining matching points in three images. He then utilized factorization to calculate the parameters of the calibration matrix. Moreover, Frahm and Koch [11] applied video sequences and accurate camera's rotation angles to estimate homography, and then obtained camera parameters. However, high-precision rotational data is difficult to achieve in the practical application of PTZ camera. Unfortunately, both Frahm and Hartley's methods are susceptible to noise input. Sinha and Marc [12] and Seo and Hong [13] addressed the defect by using Gaussian noise as input. Sinha calculated the intrinsic parameters accurately by capturing an extended panorama. Seo performed calibration by analyzing inter-image homographies. An improved camera motion model was then proposed by Davis and Chen [14], who modeled the pan and tilt as rotations around arbitrary axes in space. The approach, however, is still flawed, for it depends on a well-calibrated tracking system. Agapito *et al.* [15] and Lourdes *et al.* [16] extended Hartley's method [10] by introducing an auto-calibration approach for rotating and zooming cameras, which makes no assumptions about

scene content. Hence it is different from dynamic calibration methods [4,17-19] which are based on vanishing point extraction and limited to environments featuring reliable straight lines or special geometry patterns that can be precisely extracted.

Recently, Junejo and Foroosh [20], [21] proposed a state of the art auto-calibration method for PTZ cameras. Like Agapito's work, this approach is based on direct matrix decomposition of the infinite homography and no assumptions about scene content are needed for the camera calibration. All these algorithms are perfect theoretically. However, they fail to yield good calibration in practice when they are applied in low-resolution cameras and poor image quality, because the present detectors and the location noise of feature points cannot guarantee an accurate homography estimation between two views. In addition, the above studies fail to consider the radical distortion(RD) of lens [22]. The truth is that although lenses are now usually manufactured by precise rotational equipment to ensure radial symmetry, it is impossible to avoid optical component eccentricity and tilt in regard to one another in the camera assembling process [23], where the radial symmetry is broken and thus nonlinear radius will map from the object to the image. Therefore, the lens distortion model should be included in the calibration process to improve the computed camera model. Wu and Radke [24] thereupon extended the single-parameter division model to estimate lens distortion, which is first proposed by Fitzgibbon [25], and to simultaneously compute the homography and distortion coefficient. In Ziyan Wu's work, the lens distortion coefficient as well as the intrinsic parameters (i.e., focal length, aspect ratio, and principal point coordinates) are proceeded in several steps on ten images. Another camera self-calibration and automatic radial Distortion correction method is proposed by Zheng Tang *et al.* [26], but this approach need walking humans as calibration targets and depends on reliable human body segmentation to robust track object.

In our work, for the first time, a complete solution for on-line PTZ camera self-calibration which can calibrate five internal parameters of camera including the distortion coefficient with only two images is proposed and realized. To be specific, Givens rotation matrix to decompose the homography between the two images into a projective equivalent pair of upper-triangular matrices is constructed. To get high accuracy estimation of homography and lens distortion distortion coefficient, a novel framework called as CWRLD(Covariance Weighted Ransac under Lens Distortion), which adopted covariance matrix to define feature localization error in the presence of feature uncertainty, is created. Meanwhile, we proposed a new descriptor—SCC-SURF(Shape-Color Combined SURF) for more robust and fast point matching. All of these creative works guarantee our proposed scheme is precise, convenient, and applicable for online determining the internal parameters of PTZ camera, which is free to rotate and zoom.

This paper is structured as follows: Section II describes camera projection concepts and notations. Section III presents the proposed PTZ camera calibration method in detail. Experimental results of both synthetic and real data are presented in Section IV, and concluding remarks are given in Section V.

II. BACKGROUND AND NOTATIONS

Perspective cameras are typically in a pinhole camera model using projective mapping from point $M = [X \ Y \ Z \ 1]^T$ in 3D space to a point $m = [x \ y \ 1]^T$ on the 2D projective plane (image plane) by $m = PM$. The mapping matrix P can be represented as 3×4 matrix, decomposed as $P = K[R \mid t] = K[r_1 \ r_2 \ r_3 \ t]$ [10], thus the Euclidean transformation between camera and world coordinate system can be represented by the rotation $R = [r_1 \ r_2 \ r_3]$ and translation t. As an upper triangular matrix, K includes several intrinsic camera parameters in the form:

$$K = \begin{bmatrix} \alpha f & \tau & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad K^{-1} = \begin{bmatrix} \dfrac{1}{\alpha f} & -\dfrac{\tau}{\alpha f^2} & \dfrac{-\tau v_0 + f u_0}{\alpha f^2} \\ 0 & \dfrac{1}{f} & -\dfrac{v_0}{f} \\ 0 & 0 & 1 \end{bmatrix}$$

(1)

where f is focal length and α is aspect ratio. (u_0, v_0) represent the principal point and τ is a skew parameter as an angle function between sensor array horizontal and vertical axes (generally, $\tau = 0$ is assumed). The camera is usually positioned in the center of the absolute coordinate frame, which makes $t = 0$ and $m = KRM$.

Camera calibration aims to determine the calibration matrix K. For a PTZ camera, it is assumed that K will not change when the focal length f remains invariant, as shown in Fig. 1. Instead of being worked out directly [27], K is acquired by first solving the symmetric matrix presented as $\omega = K^{-T}K^{-1}$ (i.e. IAC: Imaged Absolute Conic) and then decomposing ω using the Cholesky Decomposition [28].

III. CALIBRATING A PTZ CAMERA WITH LENS DISTORTION FROM TWO IMAGES

Accurate camera parameter calibration relies on precise homography matrix estimation. In the last few years, Zhu *et al.* [29] proposed a new method for estimating homography, which is achieved by means of an order-preserving constraint and a similarity measurement of the quadrilateral formed by four pairs of matching points obtained by SURF (Speeded Up Robust Features) [30] algorithm. Although the calculation speed is fast, the correct matching rate of the proposed method is lower than that of the RANSAC method when the proportion of interior points is higher than 45%, and the effect of radial distortion is not considered.

For our automatic calibrating PTZ camera, homography estimation with an image pair [31] is a crucial step, which mainly includes four steps shown as Fig.2: keypoint
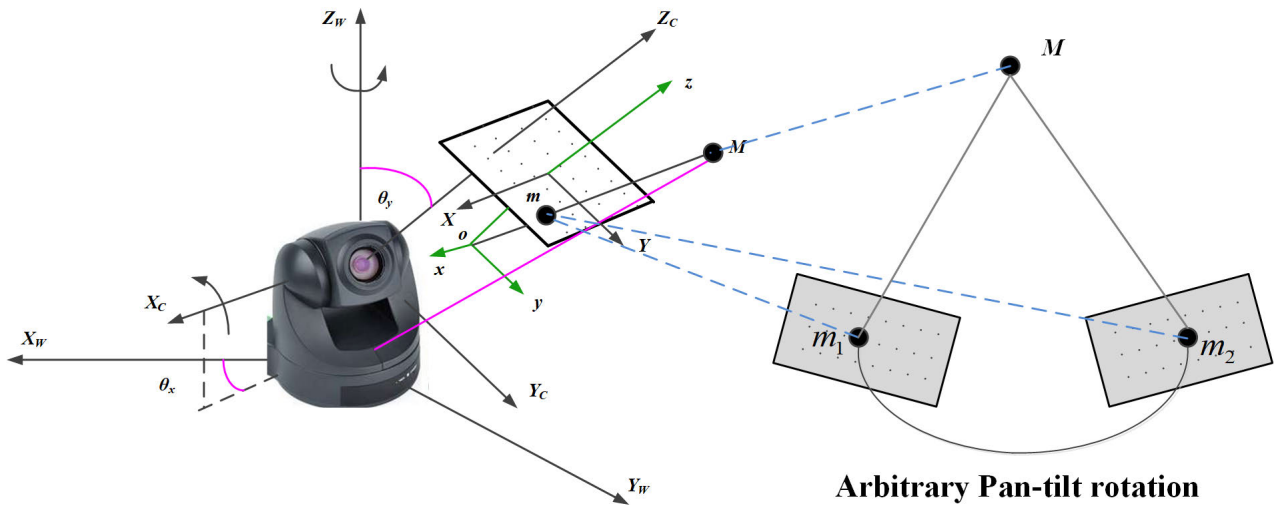
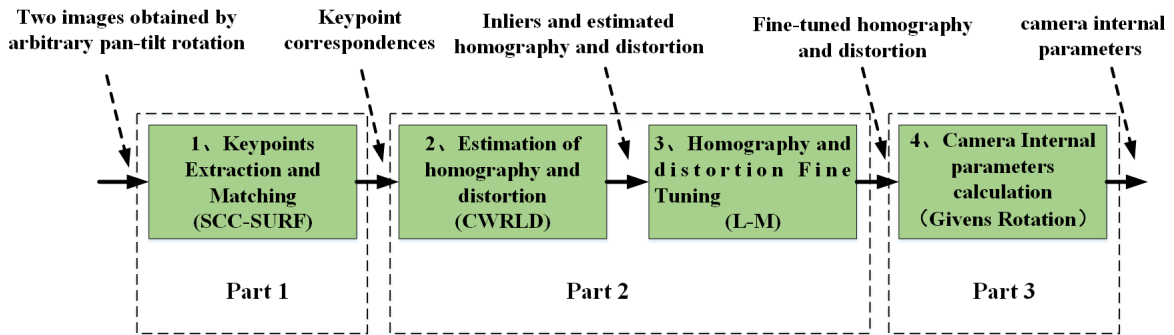**FIGURE 1.** Auto-calibration model of PTZ camera.

FIGURE 2. The overall procedure of the proposed PTZ camera calibration.

extraction in two images, estimation of homography and distortion, fine-tune the homography and distortion, and finally calculating the Givens rotation, as well as camera internal parameters.

Generally homography estimation can be decomposed into two stages: feature point matching and then homography computation by minimizing the cost function. With the remarkable development of key-point features, the SURF algorithm has recently gained popularity for its quick feature point matching. However, this algorithm is meant for gray images under distortion less perspective projections and lacks global information feature points. Moreover, these detected feature points are uncertain due to the influence of light conditions, noise, feature extraction algorithms, and other factors, which directly affects feature location stability and produces location error. To get around this, a common trait of these uncertainty estimation approaches is put forward here to represent feature location noise by using a 2D covariance matrix, resulting in an anisotropic covariance shape other than isotropic one. Thus, a novel framework is proposed in this work to accurately estimate homography by subtly combining

the SCC-SURF (Shape-Color Combined SURF) descriptor and CWRLD (Covariance Weighted Ransac under Lens Distortion) to substantially improve detection repeatability and matching effectiveness under radial distortion.

A. KEYPOINTS EXTRACTION AND MATCHING BY SCC-SURF

In the first stage of calibration, it is crucial to obtain exact matching points between the two images. Though SURF has been guaranteed to perform better than the other existing local invariant feature descriptors, and the operation is efficient, it does not have sufficient distinctiveness or robustness in the case of many locally similar areas appear and color transformations and may cause improper image matching. Inspired by SCARF[32] algorithm, we have chosen to focus primarily on the SURF descriptor extension SCC-SURF to obtain robust point correspondences, which is quick in calculation, as well as integrating the color invariant space and shape context with SURF.

The SCC-SURF algorithm is given in three stages: 1) Feature detection and localization. By using SURF method,

the determinant value of Hesseian matrix is used to detect feature points and the integral graph is used to accelerate the operation; 2) Orientation assignment; 3) Key point descriptor computation for every interest point detected combined with shape-color feature. Thus, a three-component vector is built consisting of a SURF descriptor representing local properties and a global context vector to disambiguate locally similar features:

$$SCC - SURF \sim \begin{bmatrix} \beta_1 S \\ \beta_2 G \\ (1 - \beta_1 - \beta_2) \, C \end{bmatrix} \quad (2)$$

where S is 64-dimension SURF descriptor, G and C are 32-dimension global and color descriptor respectively, here β_1 and β_2 are relative weighting factors satisfying $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_1 + \beta_2 < 1$. In fact, $\beta_1 = \beta_2 = 1/3$ is usually reasonable to handle the matching problem in the INRIA database, in which there is no uniform particularity. Hence, we used a value of $\beta_1 = \beta_2 = 1/3$ in this paper.

B. ESTIMATION OF HOMOGRAPHY AND DISTORTION BASED ON CWRLD

With the detected points given a robust matching point sequence by exploited SCC-SURF in the first stage, an appropriate method should be designed to compute homography. Hence, the novel framework CWRLD, which is based on the derived covariance estimate in the presence of feature uncertainty under the lens distortion, is defined in the following sections.

1) COVARIANCE MATRIX ESTIMATION IN THE PRESENCE OF FEATURE LOCALIZATION UNCERTAINTY

Feature points always have a level of uncertainty, whether they are extracted by an image operator or by hand. Let $(\bar{x}, \bar{y})$ be the true position of a feature point (x, y). If the errors $\Delta x = x - \bar{x}$ and $\Delta y = y - \bar{y}$ are regarded as random variables, their covariance matrix is written as:

$$\begin{pmatrix} E[\Delta x^2] & E[\Delta x \Delta y] \\ E[\Delta x \Delta y] & E[\Delta y^2] \end{pmatrix} \quad (3)$$

where E[ ] denotes expectation.

Usually, feature localization noise is considered simply as isotropically and identicallly distributed when feature detectors and descriptors are studied in terms of stability and repeatability. Thus, the covariance matrix is usually expressed as $\Gamma^0 = \Sigma^0 \sigma^0$, which is in accordance with the concentric circles model shown in Fig.3(a). Here σ^0, as a constant, stands for the noise level, and Σ^0 describes the normalized covariance matrix in the following form.

$$\Gamma^0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (4)$$

However, this assumption does not take into account that the noise distribution of variable feature points is different. Here, our work aims to parameterize the localization error and obtain a novel quantitative stability measure model for the
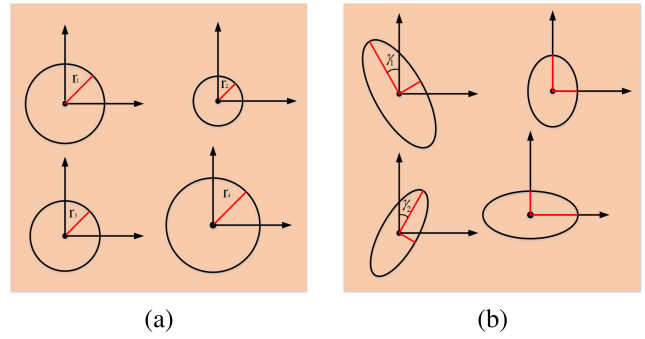

FIGURE 3. Classification of feature location noise. (a) is isotropic identical distributed noise, (b) is anisotropic non identical distributed noise.

localization uncertainty of each region, as identified by scale-invariant feature detectors, including SCC-SURF. To the best of our knowledge, covariance shape will be, in general, anisotropic, due to the focus on interest regions. Thus, the feature point localization noise in the presence of the 2D covariance matrix is no longer consistent with the concentric circles model, resulting in the elliptical shapes shown in Fig.3(b). A general representation of anisotropic and non-identical localization error is the covariance matrix[33], which is formulated as:

$$\begin{cases} \Gamma = \kappa \cdot R_r \begin{bmatrix} \beta & 0 \\ 0 & 1 - \beta \end{bmatrix} \cdot R_r^T \\ R_r = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix} \end{cases} \quad (5)$$

Here, $\kappa \in [0, +\infty)$ is the scale, $\beta \in (1/2, 1)$ is the eccentricity and $\gamma \in [0, \pi)$ represents the rotation angle of matrix Γ.

In literature [34,35], Bernhard Zeisl introduced a general framework via anisotropic covariance matrix to determine feature point uncertainty and applied it to the well-known SIFT(Scale Invariant Feature Transform)[36] and SURF algorithms. We try to extend this idea to our proposed SCC-SURF algorithm to estimate the feature localization uncertainty in this paper. Thus, feature location uncertainty represented in the form of covariance matrix Γ can then be decomposed into the following form because of its symmetry:

$$\Gamma = V \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} V^T \quad (6)$$

Here, λ_1, λ_2 are the eigenvalues of matrix Γ and V is an invertible matrix of eigenmatrices. The matrix represents an ellipse, where λ_1 and λ_2 are the squared length of long and short axes, and the direction of axes aligned with eigenvectors in V respectively.

2) LENS DISTORTION MODEL

The proposed SCC-SURF algorithm, as well as other scale-invariant feature detection approaches, is designed for perspective images and cannot deal with camera lens distortion. At the image level, radial distortion(RD) causes non-uniform pixel position displacement along radial directions and

toward the center, which leads to the compression of image structures that affect the scale-invariant detection in multiple ways, resulting in false key-points based on spurious image artifacts (e.g., straight lines that become curves). Moreover, since RD also changes image gradients, the SURF description varies in terms of the position where the feature is projected. This has a pernicious effect on matching results [37]. Accordingly, RD has a great impact on camera self-calibration accuracy. Fortunately, the one-parameter division model proposed by Fitzgibbon [25] concerns the above-mentioned RD problem, which is able to handle even pronounced RD at a much lower order. By using this model here, it follows that:

$$\tilde{m}_u = \frac{\tilde{m}_d}{1 + \eta \parallel \tilde{m}_d \parallel^2} \qquad (7)$$

where m_u is undistorted image coordinate, m_d is correspondence distortion image coordinate, $\tilde{m}_d = m_d - O$ and $\tilde{m}_u = m_u - O$. Here, the O is homogeneous coordinate of the center-of-distortion (COD). We can safely assume the COD occurs at the principal point or center of the image. In this paper, the center of the image is regarded as the COD. In addition, η is the distortion coefficient.

3) CWRLD FRAMEWORK AND INTRINSIC PARAMETERS CALIBRATION

The aforementioned linear estimating division model of lens distortion, introduced by Fitzgibbon[25] in fundamental matrix estimation, was extended by Wu and Radke [24], who simultaneously computes the homography and distortion coefficient of a PTZ camera with an arbitrary pure panning. Inspired by this idea, we developed a new framework called as CWRLD, which is a learning technique in a voting way and estimates H as well as η by random sampling the input dataset of all corresponding feature points detected by SCC-SURF.

More specially, the formula expressing of CWRLD is detailed below:

Supposing (m_u, m'_u) is a feature correspondence detected by SCC-SURF between two ideal, undistorted images taken by the PTZ camera undergoing an arbitrary pan-tilt rotation at the same zoom scale. Since the camera center is supposed to remain constant, the corresponding feature points are related by the 3×3 infinite homography matrix H, $m'_u \sim H m_u$. This can also be given by:

$$m_d' \times H m_u = 0 \qquad (8)$$

When lens distortion effects are considered, we combine (8) with the one-parameter division model in (7) to obtain the following:

$$(m'_d + \eta z_d') \times H(m_d + \eta z_d) = 0$$
$$m_d' \times H m_d + \eta(z_d' \times H m_d + m_d' \times H z_d) + \eta^2(z_d' \times H z_d) = 0 \qquad (9)$$

where $z_d = [0\ 0\ \parallel m_d \parallel^2]^T$, $z'_d = [0\ 0\ \parallel m'_d \parallel^2]^T$. Expanding everything out on (9), The following equality constraints on

homography H and distortion coefficient η can be expressed as follows:

$$(D_1 + \eta D_2 + \eta^2 D_3)h = 0 \qquad (10)$$

where h is a column vector collecting the nine elements of H. D_1, D_2, D_3 are 2×9 matrixes related to the distorted image coordinate m_d and m'_d, which is stated in Appendix A.

After obtaining the overdetermined equations as shown in formula (10), the normalized covariance weighted residuals, which considers the anisotropic and non-identity distributed feature localization error derived before, can be written as:

$$r_i = \frac{2(\lambda_1(v_1^T(m'_{ui} - \hat{m}_{ui}))^2 + \lambda_2(v_2^T(m'_{ui} - \hat{m}_{ui}))^2)}{\lambda_1 + \lambda_2} \qquad (11)$$

where m_{ui} and m'_{ui} are the ith pair of point correspondences between two ideal undistorted images, and $i = 1,2,...,n$. $\hat{m}_{ui}$ is the re-projection coordinate of m_{ui} in terms of candidate homography H. In addition, λ_1, λ_2 are eigenvalues of the covariance matrix Γ defined in (6), while $V = [v_1, v_2]$ are corresponding eigenvectors of λ_1 and λ_2.

Combining (10) and (11), we iteratively repeat the CWRLD until the final consensus set with the largest number of inliers is obtained and the estimates for H and η are calculated.

In order to further improve the estimation accuracy, we optimize the estimates for H and η with LM (Levenberg-Marquardt), thus the final accurate values for H and η, could then be obtained. Consistent with the residual error model expressed in (11), the least square cost function of optimization method is redefined in the covariance weighted form:

$$E = argmin \sum_{i=1}^{n_{max}} \left[\frac{(v_1^T(m'_{ui} - \hat{m}_{ui}))^2}{\lambda_1} + \frac{(v_2^T(m'_{ui} - \hat{m}_{ui}))^2}{\lambda_2} \right] \qquad (12)$$

where n_{max} is the number of the largest number of inliers. Up to this point, we accomplished the homography and distortion coefficient computation task of Part 2 of in Fig.2. By enumerating all key steps in Fig.4, we can systematically illustrate the whole CWRLD framework for estimation of H and η.

C. CAMERA INTERNAL PARAMETERS CALCULATION USING GIVENS ROTATION

The aforementioned high precision homography matrix obtained by CWRLD framework, can then be decomposed to get the camera internal parameters. Motivated by practical considerations with respect to PTZ cameras, we have focused principally on another solution to recover the unknown parameters ($\tau = 0$ is assumed) in K presented by Junejo and Foroosh [20], [21], who defined a sequence of Givens rotations which can selectively nullify homography entries and restored it to an upper-triangular matrix. This principle is briefly described as follows:

Let R_{12} denote the relative rotation (pan and tilt) from orientation 1 to orientation 2 of the PTZ camera with the
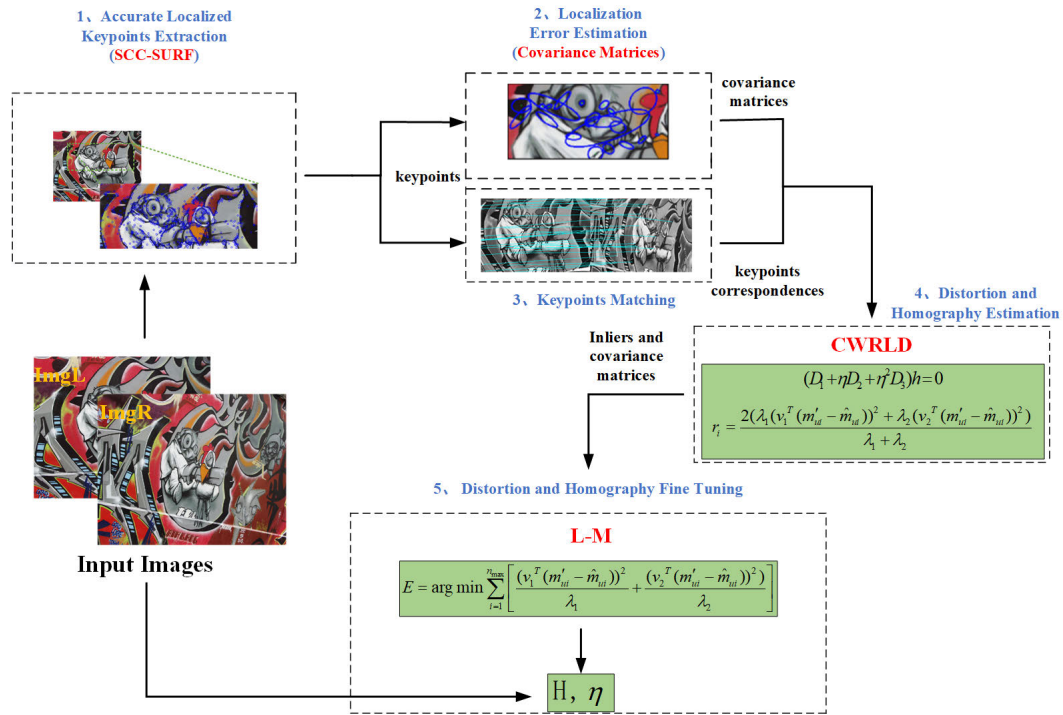
FIGURE 4. Workflow diagram of CWRLD.

constant f. Without requiring the special structure in the scene, two images taken by the PTZ camera with arbitrary pan-tilt rotation are related by infinite homography, which can be expressed in the following form:

$$H_{12} \sim KR_{12}K^{-1} \tag{13}$$

This equation can be rewritten as:

$$K^{-1}H_{12} \sim R_{12}K^{-1} \tag{14}$$

Obviously, the relative rotation R_{12} should be eliminated from (14) to determine K. Address this, Our solution uses three derived Givens rotations Q_3, Q_2 and Q_1, thus the left side of (13) could be decomposed to an upper-triangular matrix, the right side of (13) is restored as merely a camera intrinsic matrix, so:

$$Q_3Q_2Q_1K^{-1}H_{12} \sim K^{-1} \tag{15}$$

Details of (15) can be found in Appendix B. At the same time, the infinite homography H can be decomposed into projective equivalent pairs of upper-triangular matrices which offer four constraints directly from the two images with an arbitrary pan-tilt rotation as follows:

$$Q_3Q_2Q_1K^{-1}H_{12} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ 0 & \rho_{22} & \rho_{23} \\ 0 & 0 & \rho_{33} \end{bmatrix} \tag{16}$$

In this paper, $\tau = 0$ is assumed and the principal point is supposed to remain fixed between different views. From

(13-16), we can figure out the unknown parameters in K^{-1} with the four independent equality constraints to:

$$\begin{cases} \rho_{13} + u_0\rho_{11} = 0 \\ \rho_{23} + v_0\rho_{22} = 0 \\ \rho_{22} - \alpha\rho_{11} = 0 \\ \rho_{12} = 0 \end{cases} \tag{17}$$

Although these equations in (17) are non-linear, all lead to four order polynomials. The solution causing positive numbers is regarded as the correct one, by yielding the four unknown parameters (α, f, u_0, v_0). Combined with the radial distortion coefficient η obtained in section III.B, we have completed the calibration of five internal parameters (α, f, u_0, v_0, η) of PTZ camera.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present an extensive set of experimental results in two parts, i.e. the proposed homography estimation and the whole PTZ camera calibration. Specifically, either of them are evaluated by both computer simulated data and real data, respectively.

A. HOMOGRAPH ESTIMATION EXPERIMENT RESULTS

In this experiment, we study the effects of the interior point ratio as well as noise level on the homography estimation and compare our homography estimation method derived from CWRLD framework with other related methods. Considering that the radical distortion is not involved in estimating homography in relevant studies, we mainly evaluate the performance

on both computer synthetic data and real data supposing that the lens distortion is zero here.

1) SYNTHETIC EXPERIMENTS

First of all, the proposed CWRLD framework was applied to simulation data to determine its performance in the presence of various noise magnitudes and different inlier ratio levers. Firstly, we randomly generated 1000 true key-point correspondences using a known homography matrix H_{true}. Secondly, we selected noise level σ in [0.1, 1] with an interval of 0.1 and inliers ratios in [0.3, 0.9] with an interval of 0.1 to generate noisy keypoint correspondences contaminated with outliers. Then with the assumption that the lens distortion is zero (distortion coefficient $\eta = 0$), the homograph H can be calculated in the CWRLD framework described in section III.B.. In the methods of estimating homography, Random Sample Consensus (RANSAC) [40] have been proposed to select inliers and estimate transformation. Numerous RANSAC-derived methods such as M-SAC (M-estimator SAC) [38] and MLESAC (Maximum Likeli-hood SAC) [38] have been proposed to improve the per-formance of RANSAC by modifying its cost function. Nowadays, a novel feature-based homography estimation method named CW-MLESAC[39] is proposed, which is weighted by the covariance matrix of localization error on the basis of MLESAC. For fairness in comparative experiments, the above iteration algorithms, RANSAC, MSAC, MLESAC, CW MLESAC, and our CWRLD, cease to run after reaching the same iteration count on the same data set. And the RMSE (Root Mean Squares of Error) and FN (Frobenius Norm) are used for homography accuracy evaluation in this paper. Taking (m_{1i}, m_{2i}) as a feature correspondence between two images, the RMSE is computed by the discrepancy between the re-projection image point $\hat{m}_{2i}$ of m_{1i} and the ground true point m_{2i}, as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \| m_{2i} - \hat{m}_{2i} \|^2 / n} \qquad (18)$$

Here, n is the number of feature points.The FN is defined as follows:

$$FN = \| H_{true} - H_{est} \|_F \qquad (19)$$

Here, the $\| \cdot \|_F$ represents the Frobenius Norm of the matrix.

With the average performance over 1000 independent experiments for each inlier ratio and noise level. To describe the overall performance of the proposed method, the statis-tical results of RMSE and FN under different noise levels are listed in Table 1. Meanwhile, the statistical results under different inlier ratios are listed in Table 2. From the results, it can be seen that the proposed method achieves the best performance under different noise levels and inlier ratios.

In addition, the RMSE and FN for different algorithms under different inlier ratio and noise level are displayed in Fig.5 and Fig.6, where the result of the proposed CWRLD

TABLE 1. RMSE and FN vary with different noise level when inlier ratios is 0.3.

Noise Level	RANSAC RMSE	RANSAC FN	MSAC RMSE	MSAC FN	MLESAC RMSE	MLESAC FN	CWMLESAC RMSE	CWMLESAC FN	CWRLD RMSE	CWRLD FN
0.1	0.200	0.233	0.200	0.236	0.199	0.231	0.140	0.205	**0.093**	**0.137**
0.2	0.317	0.346	0.301	0.322	0.288	0.324	0.215	0.298	**0.145**	**0.194**
0.3	0.354	0.404	0.354	0.412	0.343	0.383	0.246	0.337	**0.202**	**0.236**
0.4	0.435	0.628	0.465	0.711	0.397	0.485	0.303	0.455	**0.265**	**0.307**
0.5	0.526	0.658	0.523	0.672	0.450	0.483	0.307	0.431	**0.284**	**0.365**
0.6	0.599	0.739	0.641	0.864	0.498	0.564	0.356	0.465	**0.327**	**0.397**
0.7	0.734	0.881	0.668	0.898	0.536	0.612	0.379	0.574	**0.358**	**0.432**
0.8	0.775	1.157	0.784	1.169	0.633	0.739	0.397	0.568	**0.392**	**0.451**
0.9	0.790	1.085	0.799	1.204	0.604	0.743	0.514	0.647	**0.416**	**0.517**
1.0	0.917	1.118	0.852	1.078	0.755	0.824	0.581	0.637	**0.472**	**0.568**

TABLE 2. RMSE and FN varie with different inlier ratios when noise level is 0.1.

Inlier Ratio	RANSAC RMSE	RANSAC FN	MSAC RMSE	MSAC FN	MLESAC RMSE	MLESAC FN	CWMLESAC RMSE	CWMLESAC FN	CWRLD RMSE	CWRLD FN
0.3	0.924	1.212	0.800	1.354	0.644	0.839	0.457	0.619	**0.402**	**0.583**
0.4	0.582	0.805	0.581	0.824	0.465	0.569	0.364	0.517	**0.315**	**0.502**
0.5	0.539	0.792	0.556	0.760	0.453	0.504	0.334	0.430	**0.298**	**0.407**
0.6	0.510	0.660	0.521	0.692	0.447	0.480	0.299	0.429	**0.271**	**0.395**
0.7	0.497	0.576	0.513	0.625	0.442	0.487	0.352	0.439	**0.302**	**0.398**
0.8	0.496	0.567	0.499	0.570	0.437	0.470	0.318	0.402	**0.291**	**0.375**
0.9	0.485	0.613	0.504	0.644	0.438	0.491	0.314	0.440	**0.284**	**0.368**
1.0	0.487	0.575	0.497	0.582	0.438	0.473	0.311	0.418	**0.271**	**0.357**

framework is drawn in red. We can easily and clearly find that our method performance is better and more robust in terms of mapping between image points and the accuracy of the estimation results of the homography than the other methods. Therefore, it can be indicated from the experiment that our CWRLD framework is capable of remaining reason-able performance with the presence of outliers and noise in homograph estimation.

2) REAL EXPERIMENTS

In order to support the qualitative and quantitative analysis of our improving method with real data, the sequence of images in the Institut National de Recherche en Infomatique et Automatique (INRIA) datasets (http://www.robots.ox.ac.uk/~vgg/research/affine/) are employed to test our method. And the datasets contain eight groups of datasets with the ground truth of homograph between the images. In this paper, the four groups of datasets including *bark*,*graffiti*, *wall* and *boat* are selected to experiment (see Fig.7).

Firstly, image features and correspondences for each group of dataset are obtained by using the SCC-SURF algorithm. Then the homography H is estimated by the proposed CWRLD framework in section III.B. Here, assuming that $\eta = 0$ is to make our algorithm can comparing with other algorithms.

With the average performance over 1000 independent trials for each image pair, Table 2 displays the RMSE of differ-ent algorithms for different image pairs. From the result, we can figure out that the proposed CWRLD coupling with the SCC-SURF point extracting method performs better than the other three algorithms. Compared with RANSAC, as well as two existing algorithms, e.g. MSAC and MLESAC, which assume the feature location noise to be isotropic
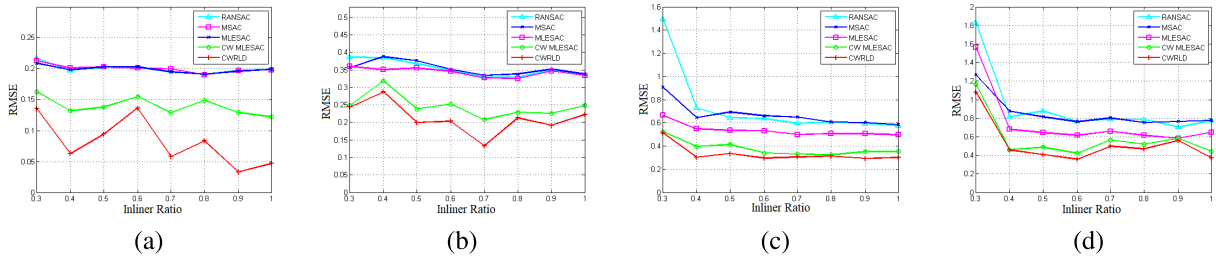
FIGURE 5. RMSE for different algorithms with inliers ratio in different noise lever. (a) - (d) correspond to noise levels of 0.1, 0.3, 0.7 and 1, respectively.
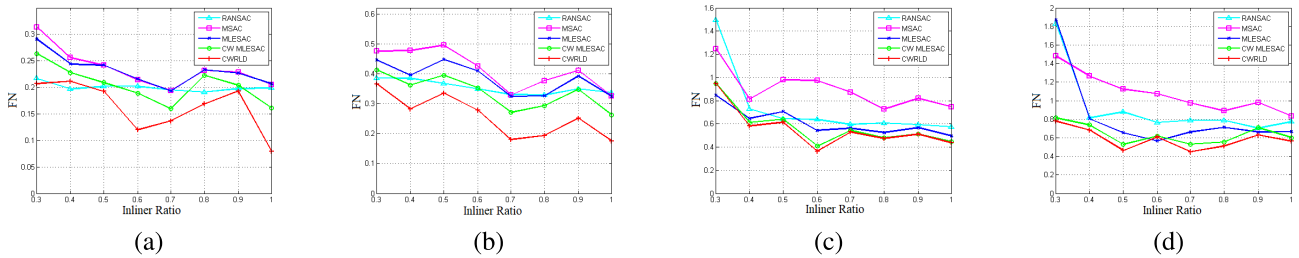

FIGURE 6. FN for different algorithms with inliers ratio in different noise lever. (a) - (d) correspond to noise levels of 0.1, 0.3, 0.7 and 1, respectively.
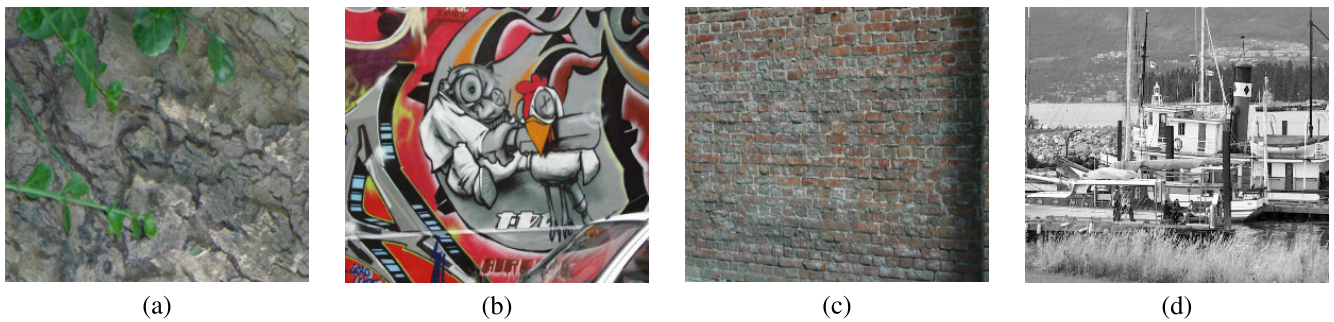

FIGURE 7. Sample images of the Institut National de RechercheenInfomatique et Automatique (INRIA) datasets used for experiments. (a) Bark; (b) graffiti; (c) wall; (d) boat.

and identical distributed, our CWRLD framework derives a new covariance estimate, which parameterizes the localization error closer to actual one as anisotropic non identical distributed, resulting in the accuracy of homography estimation improved. In addition, despite the wall dataset has plenty of similar local features which are tend to induce incorrect matches, its empirical data shown in Table 3(wall) indicate that the SCC-SURF outperforms SURF. It occurs possibly because the SCC-SURF technique which integrates the global scope and color descriptor with SURF is able to eliminate multiple regions with locally similar appearance.

In the case of distortion, not only the above advantages are maintained in our CWRLD method, but also the distortion coefficient formulated by one-parameter division model is obtained in the homography estimation. Therefore, we can easily deduce that our CWRLD method is capable of coming out on top in the presence of distortions.

TABLE 3. Average RMSE of different image pairs for different algorithms.

Image pairs		RANSAC	MSAC	MLESAC	CWMLESAC	CWRLD
Bark	1→2	1.616	1.445	1.427	1.377	**1.361**
	1→3	1.968	1.899	1.869	1.608	**1.589**
	1→4	1.210	1.165	1.149	1.077	**1.074**
Graf	1→2	1.237	0.826	0.766	0.753	**0.735**
	1→3	2.543	2.601	2.580	2.136	**2.164**
	1→4	1.966	**1.358**	1.546	1.377	1.375
Boat	1→2	1.010	0.509	0.429	0.403	**0.398**
	1→3	0.844	0.445	0.445	0.455	**0.444**
	1→4	0.915	0.722	0.703	0.694	**0.684**
Wall		SURF / SCC-SURF	SURF / SCC-SURF	SURF / SCC-SURF	SURF / SCC-SURF	SURF / SCC-SURF
	1→2	1.679 / 1.647	1.379 / 1.367	1.388 / 1.367	1.362 / 1.353	**1.324** / **1.315**
	1→3	1.268 / 1.225	0.810 / 0.753	0.789 / 0.759	0.641 / 0.596	**0.635** / **0.577**
	1→4	2.146 / 2.127	**1.908** / **1.885**	1.935 / 1.912	2.058 / 2.048	2.019 / 2.016

B. PTZ CAMERA CALIBRATION EXPERIMENT RESULTS

1) SIMULATED EXPERIMENTS

In this section, the detailed experimentation is performed to discuss the influence of noise on camera parameters
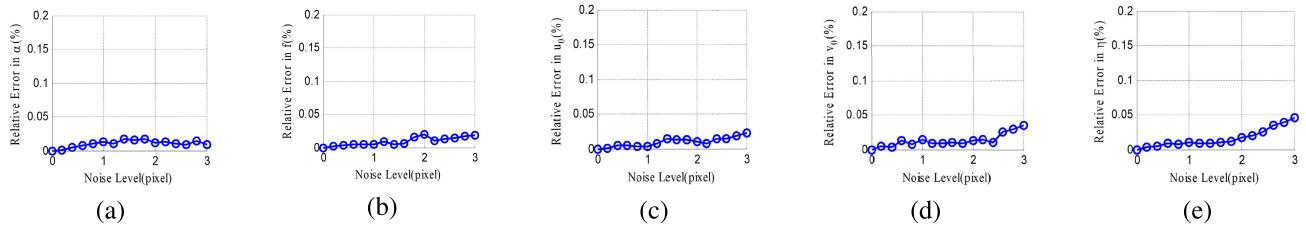
FIGURE 8. Effects of pixel coordinates noise on intrinsic parameters using the proposed method. (a) Relative error in α; (b) relative error in f; (c) relative error in u_0; (d) relative error in v_0; (e) relative error in η.
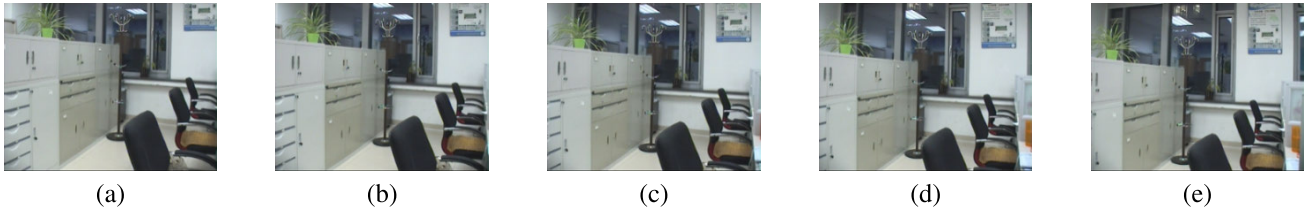

FIGURE 9. Real image sequences.

calibration. The parameters of camera are simulated as follows: the aspect ratio $\alpha = 1$, skew τ is assumed to 0, and the principal point $(u_0, v_0) = (512, 384)$. In addition, a single-parameter radial distortion division model is simulated with the coefficient $\eta = -0.4$ pixel^{-2} and the distortion center is the same as the centre of image for simplification. In this experiment, 1000 points within a unit cube are first produced and we project them into each of the image planes to generate image point correspondences, while adding the Gaussian noise with varying scale parameter σ ranging from 0.1 to 1 pixel with interval of 0.1.

For each noise level, 1000 independent trials were performed, and the results shown are the average. Then the estimated camera intrinsic parameters and lens distortion parameter $(\alpha, f, u_0, v_0, \eta)$ are compared with the ground truth and the relative errors of calibrated camera intrinsic parameters are displayed in Fig.8. The results show that our proposed CWRLD framework combined with Givens rotation decomposition for camera calibration is capable of remaining reasonable performance with the presence of noise. In the case of maximum noise, the relative error of α and focal length f are less than or equal to 0.02% and the errors in u_0, v_0 are less than 0.04%, respectively. These errors are slightly better than what reported in [21], that when the PTZ camera undergoes pan-tilt motion and the lens distortion is ignored, the error for the three estimated parameters (u_0, v_0, f) are less than 0.04%, 0.05%, 0.04%, respectively. What's more, the relative error of η illustrated in Fig. 8(e) is also less than 0.05%. Obviously, such result benefiting from the improvements called as CWRLD shows competitive performance to state-of-art techniques contributed by Wu and Radke [24], who employed the one-parameter division model in the same way and their estimated distortion coefficient error in simulated data is less 5% in the worst case. By the way, what is most striking is that we are using only two images
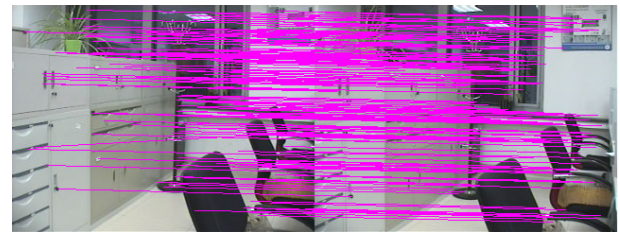

FIGURE 10. The result of SCC-SURF feature point match.

in the whole camera calibration, compared to ten in their case.

2) REAL EXPERIMENTS

We also conduct the real images on a SONY D70P PTZ camera with a resolution of 320×240 to test the proposed calibration method. Eleven images (the first five images are illustrated in Fig. 9) are captured while the pan-tilt rotation angles between the successive frames are the same. And the unknown focal length is kept constant for the camera. In this experiment, we first use the SCC-SURF algorithm to get the image features and correspondences for each pair of images. Fig.10 illustrates the matching result of the first two images taken by the PTZ camera with one-shot arbitrary Pan-Tilt rotation. Then the homography H and lens distortion coefficient η are solved by the proposed CWRLD framework. Finally, we decompose the homography using the product of three Givens rotation $Q_3 Q_2 Q_1$ to obtain the other four camera intrinsic parameters (α, f, u_0, v_0). The same procedure is repeated to all the combinations of ten pairs of images from the entire of eleven images to estimate the intrinsic parameters and lens distortion parameter $(\alpha, f, u_0, v_0, \eta)$. And, a comparison with Zhang's method [8] is carried out.

Similar to the evaluation approach of [8], i.e., the uncertainty of the estimated camera parameters over eleven images

TABLE 4. Estimated results of intrinsic parameters.

Image pairs	α	$f/pixel$	$u_0/pixel$	$v_0/pixel$	$\eta/pixel^{-2}$
1-2	0.8793	389.34	165.38	106.43	-0.374
2-3	1.0896	393.54	146.25	103.56	-0.402
3-4	1.1543	375.66	158.78	106.75	-0.386
4-5	0.9872	370.29	143.66	91.14	-0.392
5-6	0.9535	400.34	186.32	136.53	-0.414
6-7	1.1032	408.26	180.35	129.45	-0.426
7-8	1.0285	371.26	160.27	98.32	-0.389
8-9	1.1153	402.89	139.34	115.53	-0.418
9-10	0.9657	362.13	173.46	150.21	-0.359
10-11	1.1658	390.53	145.29	90.75	-0.421
Mean	1.0442	386.42	159.91	112.86	-0.3981
M.dev	0.0826	13.6150	14.0850	12.1950	0.0190
Zhang	0.9909	380.41	149.14	93.65	—

FIGURE 11. The errors of focal length between our method and Zhang's in different Pan-Tilt rotation angle.

(a)   (b)

FIGURE 12. Image pairs and corresponding matching results under different focal lengths. The focal length of (a) is 3.2967mm, The focal length of (b) is 14.7631mm.

FIGURE 13. Absolute error of focal length calibration results.

is used for the evaluation standard. And the uncertainty can be characterized by their median deviation (M.Dev). All the estimated camera parameters displayed in Table 4 clarify that the camera intrinsic parameters for different pair of images are very consistent, which can be demonstrated from M.Dev by low variance (less than 15 pixels, and the physical size of 1 pixel is 0.01mm). Moreover, from the results we observe that when using our method, the aspect ratio α is estimated as 1.0442 and the focal length is estimated as 386.42, which are respectively approaching 0.9909 and 380.41 obtained by Zhang's method. Besides, the error of principal point estimated by two methods is within 10 pixels. Especially, as one of the biggest concerns in computer vision, the focal length is only about 6 pixels difference between the Mean of our method and Zhang's method. Note that, for the reason that Zhang's method adopts the Polynomial Model to describe lens distortion and our method is based on the one-parameter Division Model, the obtained distortion coefficients are not comparable. In particular, the first order radial distortion coefficient obtained by Zhang's method is $\eta_1 = -0.2730$ pixel^{-2}, while the second order is $\eta_2 = 0.3626$ pixel^{-4}. Nevertheless, from the M.Dev by low variance (0.0190) it can be seen that the distortion coefficients obtained from different pairs of images by our method are small fluctuation and have great repeatability.

Another experiment was designed to further quantify the performance and feasibility of our proposed strategy. Keeping the same focal length, the PTZ camera is directed to several different positions in a kind of pan-tilt rotation mode in the pan range $[-20, 20]$ degree and tilt range $[-18, 18]$ degree. Corresponding to the rotation range, the control time range is set from -200ms to $+200$ms. Then we conducted the whole proposed calibration procedure with each different pair of images obtained by this test. The absolute error of estimated focal length between our method and Zhang's is depicted in Fig.11. We observe that the absolute errors range from 2 to 22 pixels, of which 80% concentrate in 3 to 14 pixels.

To verify the broad applicability of the proposed CWRLD framework, we chose another PTZ camera(SONY EVI-D90P) for the experiment under 19 different focal lengths (corresponding to group serial number 1-19). The
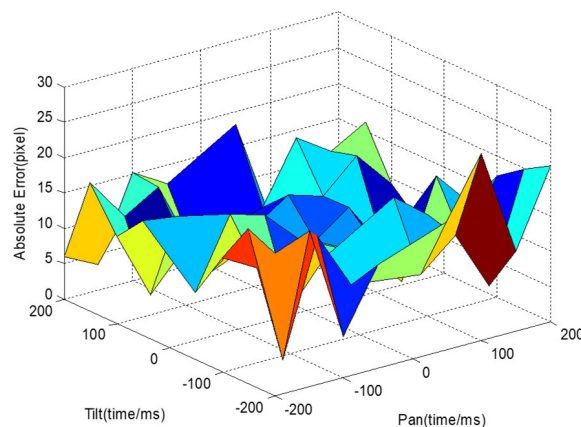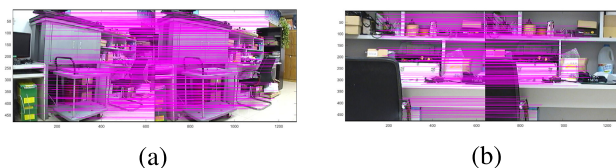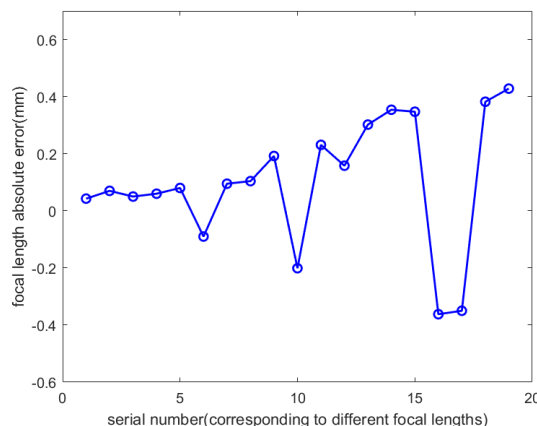
focal length value increases along with the serial number). In the experiment, 190 images (10 in each group, 19 groups in total) were taken, Adjacent images of each group were captured by PTZ camera panning 10°and tilting 2°in a rotational motion. Fig.12 shows the point matching results of adjacent images before and after a rotational motion under different focal lengths.

Furthermore, we still consider Zhang's method[8] as ground truth to evaluate the performance of the proposed CWRLD framework. Fig.13 shows the absolute error of focal length calibration results under different focal length or serial numbers. Combined with Table 4 and Fig.11, it can be
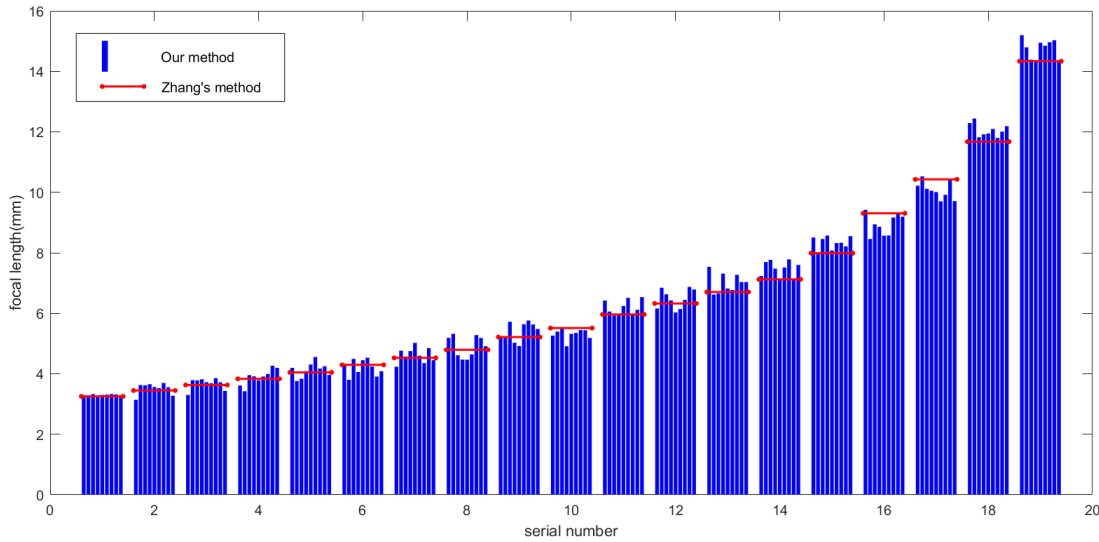
FIGURE 14. The focal length calibration results distribution histogram obtained by our method.
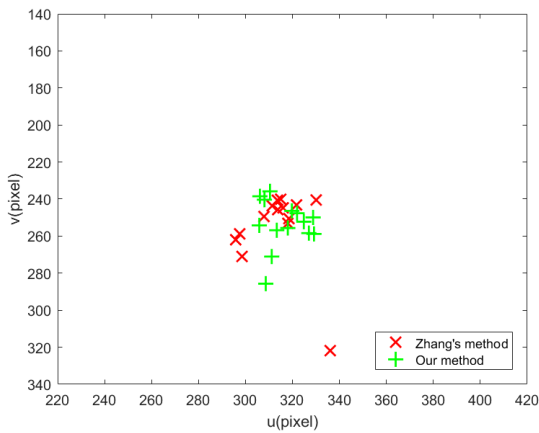

FIGURE 15. Principal point coordinates calibration results distribution.

found that the absolute error of the focal length calibration results of both PTZ cameras does not exceed 0.22 mm at the initial focal length. In general, the absolute error increases as the focal length increases and the maximum absolute error is 0.428mm in the experiment. The distribution histogram of the focal length calibration results are illustrated in Fig.14. We know from Fig.14 that the error limit of focal length calibration results in our method and Zhang's method moves from 0.065mm to 0.958mm and from 0.066mm to 0.894mm, respectively. Hence we can conclude that the focal length calibration results using our method are comparable with the results of Zhang's method which is often recognized as a standard means in most references.

Figure 15 shows the distribution diagram of the principal points. The horizontal and vertical axes correspond to the u-axis and the v-axis of the image coordinate system, respectively. For the convenience of observation, the two

axes respectively take a 100 pixels value before and after the center. As can be seen from the graph, our method is stable when the focal length is changed.

What is perhaps most remarkable here, the low variances of the camera intrinsic parameters for different pair of images can testify that our method has great robustness. Therefore, the PTZ camera could be conveniently calibrated using only two images in the practical application.

V. CONCLUSION

A novel framework requiring only two images has been explored to solve the PTZ camera full automatic calibration problem. Once the PTZ camera is adjusted to an uncalibrated fixed focal length, it is only required to undergo an arbitrary pan-tilt rotation to generate two images, which are then used to perform the eigendecomposition of infinite homography with a series of Givens rotations and yield four constraints about the intrinsic parameters. The camera is thus considered calibrated when focal length, principal point offset, and lens distortion parameters are known. In order to make the proposed methodology possess not only simple operating, but also the high noise immunity and accuracy in practical use, we elegantly combined the estimation of lens distortion coefficients and homography in a novel framework named CWRLD, which adopts covariance matrix to define feature localization error in the presence of feature uncertainty. In addition, a more distinctive SCC-SURF descriptor is employed to obtain the robust point correspondences between two views. Both simulations and real experiments prove that the proposed method is reliable and effective. The newly developed technique is more practical and flexible than previous methods, with the capacity for application in online PTZ camera calibration when measurement is needed at any time.

It's worth noting that, a static background is crucial to our calibration method because scene features in an image pair used for calibration should remain unchanged. We plan to incorporate the proposed model and algorithms into real applications in the future to improve both 2D and 3D identification and localization performance.

APPENDIX A
DETAILS OF (10)

Let $m_d = (x, y)$, $m'_d = (x', y')$, and $r = \| m_d \|$, $r' = \| m'_d \|$, D_1, D_2, D_3 can be represented as the following form:

$$D_1 = \begin{bmatrix} 0 & 0 & 0 & -x' & -y' & -1 & yx' & yy' & y \\ x' & y' & 1 & 0 & 0 & 0 & -xx' & -xy' & -x \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 0 & 0 & 0 & -rx' & -ry' & -r-r' & 0 & 0 & yr' \\ rx' & ry' & r+r' & 0 & 0 & 0 & 0 & 0 & -xr' \end{bmatrix}$$

$$D_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -rr' & 0 & 0 & 0 \\ 0 & 0 & rr' & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

APPENDIX B
DETAILS OF (15)

Let $K^{-1} = [k_1 \ k_2 \ k_3]^T$, where k_i^T (i = 1,2,3) are rows of K^{-1}. $H_{12} = [h_1 \ h_2 \ h_3]$, and h_i (i = 1,2,3) are infinite homography columns. The Givens rotation Q_3, Q_2 and Q_1 can be defined by:

$$Q_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta_1 & sin\theta_1 \\ 0 & -sin\theta_1 & cos\theta_1 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} cos\theta_2 & sin\theta_2 & 0 \\ -sin\theta_2 & cos\theta_2 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$Q_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta_3 & sin\theta_3 \\ 0 & -sin\theta_3 & cos\theta_3 \end{bmatrix}$$

where

$$cot\theta_1 = \frac{k_2^T h_1}{k_3^T h_1}, \quad cot\theta_2 = \frac{k_1^T h_1}{(k_2^T h_1 \ h_1^T k_2 + k_3^T h_1 \ h_1^T k_3)^{\frac{1}{2}}}$$

$$cot\theta_3 = \frac{k_3^T h_2 \ sin\theta_1 \ cos\theta_2 + k_2^T h_2 \ cos\theta_1 \ cos\theta_2 - k_1^T h_2 \ sin\theta_2}{k_3^T h_2 \ cos\theta_1 - k_2^T h_2 \ sin\theta_1}$$

According to the theorem of Givens rotation, the rotation Q_3, Q_2 and Q_1 can nullify the elements of homography and decomposed it to an upper-triangular matrix.

ACKNOWLEDGMENT

REFERENCES

[1] Y. Cai and G. Medioni, "Persistent people tracking and face capture using a PTZ camera," *Mach. Vis. Appl.*, vol. 27, no. 1, pp. 397–413, Apr. 2016.

[2] R. Wang, H. Dong, H. X. Tony, and L. Mei, "Robust tracking via monocular active vision for an intelligent teaching system," *Vis. Comput.*, vol. 32, no. 11, pp. 1379–1395, Nov. 2016.

[3] Y. Jeong, D. Park, and K. H. Park, "PTZ camera-based displacement sensor system with perspective distortion correction unit for early detection of building destruction," *sensors*, vol. 17, no. 3, p. 430, 2017.

[4] H. Song, C. Li, F. Wu, Z. Dai, W. Wang, Z. Zhang, and Y. Fang, "3D vehicle model-based PTZ camera auto-calibration for smart global village," *Sustain. Cities Soc.*, vol. 46, Apr. 2019, Art. no. 101401.

[5] F. C. Wu, Z. Y. Hu, and H. J. Zhu, "Camera calibration with moving one-dimensional objects," *Pattern Recognit.*, vol. 38, no. 5, pp. 755–766, 2016.

[6] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. 3, no. 4, pp. 323–344, Aug. 1987.

[7] G. Xu, A. Q. Zheng, X. T. Li, and J. Su, "3D reconstruction of laser projective point with projection invariant generated from five points on 2D target," *Sci. Rep.*, vol. 7, no. 1, Oct. 2016, Art. no. 7049.

[8] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[9] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank, "Camera self-calibration: Theory and experiments," in *Proc. Eur. Conf. Comput. Vis.*, in Lecture Notes in Computer Sciences, vol. 588, 1992, pp. 321–334.

[10] R. I. Hartley, "Self-calibration from multiple views with a rotating camera," in *Proc. ECCV*, Stockholm, Sweden, 1994, pp. 471–479.

[11] J. M. Frahm and R. Koch, "Camera calibration with known rotation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Nice, France, Oct. 2003, pp. 1418–1425.

[12] S. N. Sinha and M. Pollefeys, "Pan–tilt–zoom camera calibration and high-resolution mosaic generation," *Comput. Vis. Image Understand.*, vol. 103, no. 3, pp. 170–183, 2006.

[13] Y. Seo and K. S. Hong, "About the self-calibration of a rotating and zooming camera: Theory and practice," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Kerkyra, Greece, Sep. 1999, pp. 183–189.

[14] J. Davis and X. Chen, "Calibrating pan-tilt cameras in wide-area surveillance networks," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Los Alamitos, CA, USA, 2003, pp. 144–149.

[15] L. de Agapito, R. I. Hartley, and E. Hayman, "Linear self-calibration of a rotating and zooming camera," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* Oxford, U.K.: Oxford Univ., 1999, pp. 15–21.

[16] L. Agapito, E. Hayman, and I. Reid, "Self-calibration of rotating and zooming cameras," *Int. J. Comput. Vis.*, vol. 45, no. 2, pp. 107–127, Nov. 2001.

[17] T. N. Schoepflin and D. J. Dailey, "Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 2, pp. 90–98, Jun. 2003.

[18] K.-T. Song and J.-C. Tai, "Dynamic calibration of pan–tilt–zoom cameras for traffic monitoring," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 5, pp. 1091–1103, Oct. 2006.

[19] R. Wang, X. Li, and G.-J. Zhang, "A linear algorithm for determining intrinsic parameters of zoomed monocular camera in the vision based landing of an UAV," *Acta Aeronautica Astronautica Sinica*, vol. 27, no. 4, pp. 676–681, 2006.

[20] I. N. Junejo and H. Foroosh, "Practical PTZ camera calibration using Givens rotations," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 1936–1939.

[21] I. N. Junejo and H. and Foroosh, "Optimizing PTZ camera calibration from two images," *Mach. Vis. Appl.*, vol. 23, no. 2, pp. 375–389, 2012.

[22] Z. Chun-Ming and X. Yong-Chun, "A method of analyzing image distortion based on conjugate nodal plane," in *Proc. 31st Chin. Control Conf.*, Hefei, China, Jul. 2012, pp. 3690–3695.

[23] J. Wang, F. Shi, J. Zhang, and Y. Liu, "A new calibration model of camera lens distortion," *Pattern Recognit.*, vol. 41, no. 2, pp. 607–615, Feb. 2008.

[24] Z. Wu and R. J. Radke, "Keeping a pan–tilt–zoom camera calibrated," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1994–2007, Aug. 2013.

[25] A. W. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Kauai, HI, USA, Dec. 2001, pp. 125–132.

[26] Z. Tang, Y.-S. Lin, K.-H. Lee, J.-N. Hwang, and J.-H. Chuang, "ESTHER: Joint camera self-calibration and automatic radial distortion correction from tracking of walking humans," *IEEE Access*, vol. 7, no. 1, pp. 10754–10766, 2019.

[27] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2003, pp. 634–645.

[28] W. H. Press, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 1992, pp. 926–929.

[29] H. Zhu, X. Wen, F. Zhang, X. Wang, and G. Wang, "Homography estimation based on order-preserving constraint and similarity measurement," *IEEE Access*, vol. 6, pp. 28680–28690, 2018.

[30] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[31] J. Yang and R. Wang, "High-accuracy homography estimation for robust pan–tilt–zoom camera calibration," in *Proc. 3rd Int. Conf. Comput. Intell. Commun. Technol.*, Ghaziabad, India, Feb. 2017, pp. 1–5.

[32] R. Wang, Z. Zhu, and L. Zhang, "Improving scale invariant feature transform-based descriptors with shape–color alliance robust feature," *J. Electron. Imag.*, vol. 24, no. 3, May 2015, Art. no. 033002.

[33] M. J. Brooks, W. Chojnacki, D. Gawley, and A. van den Hengel, "What value covariance information in estimating vision parameters?" in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, Vancouver, BC, Canada, Jul. 2017, pp. 302–308.

[34] R. M. Steele and C. Jaynes, "Feature uncertainty arising from covariant image noise," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, Jun. 2005, pp. 1063–1070.

[35] Z. Bernhard, F. G. Pierre, S. Florian, G. S. Eckehard, and N. Nassir, "Estimation of location uncertainty for scale invariant features points," in *Proc. BMVC*, London, U.K., 2009, pp. 1–12.

[36] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[37] R. O. Castle, D. J. Gawley, G. Klein, and D. W. Murray, "Towards simultaneous recognition, localization and mapping for hand-held and wearable cameras," in *Proc. IEEE Int. Conf. Robot. Automat. Robot. Automat.*, Roma, Italy, Apr. 2007, pp. 4102–4107.

[38] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understand.*, vol. 78, no. 1, pp. 138–156, 2000.

[39] C. Zhao and H. Zhao, "Accurate and robust feature-based homography estimation using HALF-SIFT and feature localization error weighting," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 288–299, Oct. 2016.

[40] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

RUI WANG received the B.S. and M.S. degrees in optical instruments from the Department of Precision Instruments and Mechanology, Tsinghua University, China, in 1988 and 1990, respectively, and the on-the-job Ph.D. degree in machine vision of precision instrument and mechanics from the School of Instrumentation and Opto-Electronics Engineering, Beihang University, in 2006, where she is currently an Associate Professor. She has authored more than 70 articles and two books. Her current research interests include computer vision, signal processing, pattern recognition and analysis, and machine learning.

RAN HUANG received the B.E. degree in measurement and control technology and instruments from the China University of Geosciences, Beijing, China, in 2017. He is currently pursuing the M.E. degree in measurement technology and instruments with the Laboratory of Precision Opto-Mechatronics Technology, Beihang University, Beijing. His current research interests include camera calibration and pose measurement.

JIAORU YANG received the B.E. degree from Xidian University, in 2014, and the M.E. degree from the Beihang University of Measurement Technology and Instruments, in 2017. She is currently with the Beijing Aerospace Automatic Control Institute, China Aerospace Science and Technology Corporation. Her research interests include image processing and camera calibration.

● ● ●