Calibrating Pan-Tilt Cameras with Telephoto Lenses

Conference Paper · November 2007

DOI: 10.1007/978-3-540-76386-4_11 · Source: DBLP

3 authors, including:

Jizhou Gao
University of Kentucky
13 PUBLICATIONS **435** CITATIONS

Calibrating Pan-Tilt Cameras with Telephoto Lenses

Xinyu Huang, Jizhou Gao, and Ruigang Yang

Graphics and Vision Technology Lab (GRAVITY)
Center for Visualization and Virtual Environments
University of Kentucky, USA
{xhuan4,jgao5,ryang}@cs.uky.edu
http://www.vis.uky.edu/~gravity

Abstract. Pan-tilt cameras are widely used in surveillance networks. These cameras are often equipped with telephoto lenses to capture objects at a distance. Such a camera makes full-metric calibration more difficult since the projection with a telephoto lens is close to orthographic. This paper discusses the problems caused by pan-tilt cameras with long focal length and presents a method to improve the calibration accuracy. Experiments show that our method reduces the re-projection errors by an order of magnitude compared to popular homography-based approaches.

1 Introduction

A surveillance system usually consists of several inexpensive wide fields of view (WFOV) fixed cameras and pan-tilt-zoom (PTZ) cameras. The WFOV cameras are often used to provide an overall view of the scene while a few zoom cameras are controlled by pan-tilt unit (PTU) to capture close-up views of the subject of interest. The control of PTZ camera is typically done manually using a joystick. However, in order to automate this process, calibration of the entire camera network is necessary.

One of our driving applications is to capture and identify subjects using biometric features such as iris and face over a long range. A high-resolution camera with a narrow field of view (NFOV) and a telephoto lens is used to capture the rich details of biometric patterns. For example, a typical iris image should have 100 to 140 pixels in iris radius to obtain a good iris recognition performance [1]. That means, in order to capture the iris image over three meters using a video camera (640×480), we have to use a 450mm lens assuming sensor size is 4.8×3.6 mm. If we want to capture both eyes (e.g., the entire face) at once, then the face image resolution could be as high as 5413×4060 pixels–well beyond the typical resolution of a video camera. In order to provide adequate coverage over a practical working volume, PTZ cameras have to be used. The simplest way to localize the region of interest (ROI) is to pan and tilt the PZT camera iteratively until the region is approximately in the center of field of view [2]. This is time-consuming and only suitable for still objects. However, if the PTZ cameras are fully calibrated, including the axes of rotation, the ROI can be localized rapidly with a single pan and tilt operation. In this paper we discuss the degeneracy caused by cameras with telephoto lenses and develop a method to calibrate such a system with significantly improved accuracy.

The remaining of this paper is organized as the following. We first briefly overview the related work in section 2. In section 3, we describe our system and a calibration

method for long focal length cameras. Section 4 contains experimental results. Finally, a summary is given in section 5. We also present in the appendix a simple method to calculate the pan and tilt angle when the camera coordinate is not aligned with the pan-tilt coordinate.

2 Related Work

It is generally considered that camera calibration reached its maturity in the late 90's. A lot of works have been done in this area. In the photogrammetry community, a calibration object with known and accurate geometry is required. With markers of known 3D positions, camera calibration can be done efficiently and accurately (e.g., [3], [4]). In computer vision, a planar pattern such as a checkerboard pattern is often used to avoid the requirement of a 3D calibration object with a good precision (e.g., [5], [6]). These methods estimate intrinsic and extrinsic parameters including radial distortions from homographies between the planar pattern at different positions and the image plane. Self-calibration estimates fixed or varying intrinsic parameters without the knowledge of special calibration objects and with unknown camera motions (e.g., [7], [8]). Furthermore, self-calibration can compute a metric reconstruction from an image sequence. Besides the projective camera model, the affine camera model in which camera center lies on the plane at infinity is proposed in ([9], [10]). Quan presents a self-calibration method for an affine camera in [11]. However, the affine camera model should not be used when many feature points at different depths [12].

For the calibrations of PTZ cameras, Hartley proposed a self-calibration method for stationary cameras with purely rotations in [13]. Agapito extended this method in [14] to deal with varying intrinsic parameters of a camera. Sinha and Pollefeys proposed a method for calibrating pan-tilt-zoom cameras in outdoor environments in [15]. Their method determines intrinsic parameters over the full range of zoom settings. These methods above approximate PTZ cameras as rotating cameras without translations since the translations are very small compared to the distance of scene points. Furthermore, These methods are based on computing absolute conic from a set of inter-image homographies. In [16], Wang and Kang present an error analysis of intrinsic parameters caused by translation. They suggest self-calibrate using distant scenes, larger rotation angles, and more different homographies in order to reduce effects from camera translation. The work most similar to ours is proposed in [17, 18]. In their papers, Davis proposed a general pan-tilt model in which the pan and tilt axes are arbitrary axes in 3D space. They used a bright LED to create a virtual 3D calibration object and a Kalman filter tracking system to solve the synchronization between different cameras. However, they did not discuss the calibration problems caused by telephoto lens. Furthermore, their method cannot be easily applied to digital still cameras with which it will be tedious to capture hundreds or even thousands of frames.

3 Method

In this section, we first describe the purpose of our system briefly. Then, we discuss the calibration for long focal length camera in details.

3.1 System description

The goal of our system is to capture face or iris images over a long range with a resolution high enough for biometric recognitions. As shown in Fig. 1, a prototype of our system consists of two stereo cameras and a NFOV high resolution (6M pixels) still camera. The typical focal length for the pan-tilt camera is 300mm, while previous papers dealing with pan-tilt cameras have reported the use of lenses between 1mm to 70mm. When a person walks into the working area of the stereo cameras, facial features are detected in each frame and their 3D positions can be easily determined by triangulation. The pan-tilt camera is steered so that the face is in the center of the observed image. A high-resolution image with enough biometric details then can be captured. Since the field of view of pan-tilt camera is only about 8.2 degrees, the ROI (e.g., the eye or entire face) is likely to be out of the field of view if the calibration is not accurate enough.

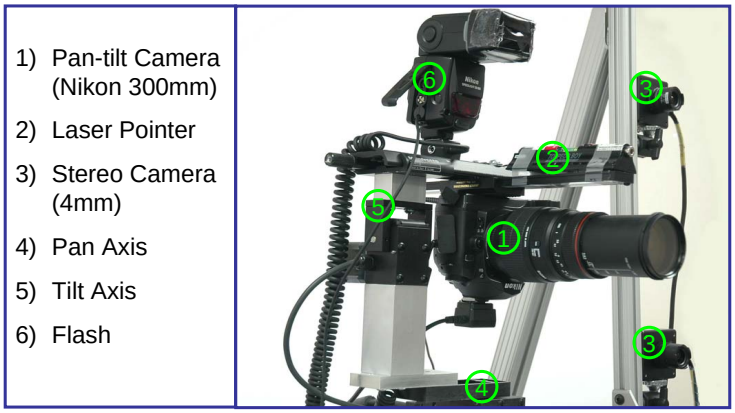

Fig. 1. System setup with two WFOV cameras and a pan-tilt camera.

3.2 Calibration

In [5], given one homography $\mathrm{H} = [\mathrm{h}_1, \mathrm{h}_2, \mathrm{h}_3]$ between a planar pattern at one position and the image plane, two constraints on the absolute conic ω can be formulated as in Eq.(1).

$$\begin{aligned} \mathrm{h}_1^T \omega \mathrm{h}_2 &= 0 \\ \mathrm{h}_1^T \omega \mathrm{h}_1 &= \mathrm{h}_2^T \omega \mathrm{h}_2 \end{aligned} \tag{1}$$

By imaging the planar pattern n times at different orientations, a linear system $\mathrm{A}c = 0$ is formed, where A is a $2n \times 6$ matrix from the observed homographies and c represents ω as a 6×1 vector. Once c is solved, intrinsic matrix K can be solved by Cholesky factorization since $\omega = (KK^T)^{-1}$.

Equivalently, one could rotate the camera instead of moving a planar pattern. This is the key idea in self-calibration of pan-tilt cameras ([15], [13]). First, inter-image homographies are computed robustly. Second, the absolute conic ω is estimated by a

linear system $\omega = (H^i)^{-T}\omega(H^i)^{-1}$, where H^i is the homography between each view i and a reference view. Then, Cholesky decomposition of ω is applied to compute the intrinsic matrix K. Furthermore, a Maximum Likelihood Estimation (MLE) refinement could be applied using the above close-form solution as the initial guesses. However, the difference is small between the close-form solution and that from MLE refinement [12].

As mentioned in [5], the second homography will not provide any new constraints if it is parallel to the first homography. In order to avoid this degeneracy and generate a over-determined system, the planar pattern has to be imaged many times with different orientations. This is also true for the self-calibration of rotating cameras. If conditions are near singular, the matrix A formed from the observed homographies will be ill-conditioned, making the solution inaccurate.

Generally, the degeneracy is easy to avoid when the focal length is short. For example, we only need to change the orientation for each position of planar pattern. However, this is not true for long focal length cameras. When the focal length increases and the filed-of-view decreases, the camera's projection becomes less projective and more orthographic. The observed homographies contain a lot of depth ambiguities that make the matrix A ill-conditioned and solution is very sensitive to a small perturbation. If the projection is purely orthographic, then observed homographies can not provide any depth information no matter where we put the planar pattern or how we rotate the camera. In summary, traditional calibration methods based on observed homographies are in theory not accurate for long focal length camera. We will also demonstrate this point with real data in the experiments section.

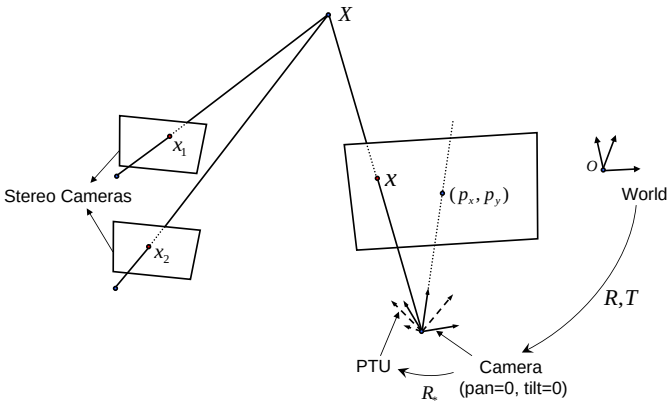

Fig. 2. Pan-tilt camera model.

The best way to calibrate a long focal length camera is to create 2D-3D correspondences directly. One could use a 3D calibration object but this approach is not only costly, but also un-practical given the large working volume we would like to cover. In our system 3D feature points are triangulated by stereo cameras, therefore it will not induce any ambiguities caused by the methods based on observed homographies. With a set of known 2D and 3D features, we can estimate intrinsic parameters and the relative transformation between the camera and the pan-tilt unit. The pan-tilt model is shown in

Fig. 2 and is written as

$$x = KR_*^{-1}R_{tilt}R_{pan}R_*[R|t]X \tag{2}$$

where K is intrinsic parameters, R and t are extrinsic parameters of pan-tilt camera at reference view that is $pan = 0$ and $tilt = 0$ in our setting. X and x are 3D and 2D feature points. R_{pan} and R_{tilt} are rotation matrices around pan and tilt axes. R_* is the rotation matrix between coordinates of the camera and the pan-tilt unit. We did not consider the pan and tilt axes as two arbitrary axes in 3D space as in [17] since translation between the two coordinates are very small (usually only a few millimeters in our setting) and a full-scale simulation shows that adding the translational offset yield little accuracy improvement. Based on the pan-tilt model in Eq.(2), we could estimate the complete set of parameters using MLE to minimize the re-projected geometric distances. This is given by the following functional:

$$\text{argmin}_{R_*,K,R,t} \sum_{i=1}^{n}\sum_{j=1}^{m} \|x_{ij} - \hat{x}_{ij}(K, R_*, R_{pan}, R_{tilt}, R, t, X_{ij})\|^2 \tag{3}$$

The method of acquiring of calibration data in [17] is not applicable in our system because that our pan-tilt camera is not a video camera that could capture a video sequence of LED points. Typically a commodity video camera does not support both long focal length and high-resolution image. Here we propose another practical method to acquire calibration data from a still camera. We attach a laser pointer close enough to the pan-tilt camera as shown in Fig.1. The laser's reflection on scene surfaces generates a 3D point that can be easily tracked. The laser pointer rotates with the pan-tilt camera simultaneously so that its laser dot can be observed by the pan-tilt camera at most of pan and tilt settings. In our set-up, we mount the laser pointer on the tilt axis. A white board is placed at several positions between the near plane and the far plane within the working area of two wide-view fixed cameras. For each pan and tilt step, three images are captured by the pan-tilt camera and two fixed cameras respectively. A 3D point is created by triangulation from the two fixed cameras. The white board does not need to be very large since we can always move it around during the data acquisition process so that 3D points cover the entire working area. The calibration method is summarized as Algorithm 1.

In order to compare our method with methods based on observed homographies, we formulate a calibration framework similar to the methods in [15] and [12]. The algorithm is summarized in Algorithm 2.

An alternative of step 4 in Algorithm 2 is to build a linear system $\omega = (H^i)^{-T}\omega(H^i)^{-1}$ and solve ω. Intrinsic matrix K is solved by Cholesky decomposition $\omega = KK^T$. However, this closed-form solution often fails since infinite homography is hard to estimate with narrow fields of view.

After calibration step, R_*, intrinsic and extrinsic parameters of three cameras are known. Hence, We can solve the pan and tilt angles easily (see Appendix A for the details) for almost arbitrary 3D points triangulated by stereo cameras.

Algorithm 1 our calibration method for a pan-tilt camera with a long focal length.

Input: observed laser point images by three cameras.
Output: intrinsic matrix K extrinsic parameters R, t, and rotation matrix R_* between coordinates of camera and PTU.

1. Calibrate the stereo cameras and reference view of the pan-tilt camera using [19].
2. Rectify stereo images such that epipolar lines are parallel with the y-axis (optional).
3. Capture laser points on a 3D plane for three cameras at each pan and tilt setting in the working area.
4. Based on blob detection and epipolar constraint, find two laser points in the stereo cameras. Generate 3D points by triangulation of two laser points.
5. Plane fitting for each plane position using RANSAC.
6. Remove outliers of 3D points based on the fitted 3D plane.
7. Estimate R_*, K, R, t by minimizing Eq.(3).

Algorithm 2 calibration method for a pan-tilt camera with a long focal length based on homographies.

Input: images captured at each pan and tilt setting.
Output: intrinsic matrix K and rotation matrix R_* between coordinates of camera and PTU.

1. Detect features based on Scale-invariant feature transform (SIFT) in [20] and find correspondences between neighboring images.
2. Robust homography estimation using RANSAC.
3. Compute homography between each image and reference view ($pan = 0, tilt = 0$).
4. Estimate K using Calibration Toolbox [19].
5. Estimate R_* and refine intrinsic matrix K by minimizing

$$\text{argmin}_{R_*,K} \sum_{i=1}^{n} \sum_{j=1}^{m} \|x_j^i - K R_*^{-1} R_{tilt} R_{pan} R_* K^{-1} x_{ref}^i\|^2 \tag{4}$$

where x_j^i and x_{ref}^i are i_{th} feature point at j_{th} and reference view respectively.

4 Experiments

Here we present experimental results from two fixed cameras (Dragonfly2 DR2-HICOL with resolution 1024×768) and a pan-tilt still camera (Nikon D70 with resolution 3008×2000). First, we compare the calibration accuracy with short and long focal length lenses using traditional homograph-based method. Then, we demonstrate that our calibration method significantly improves accuracy for telephoto lenses. In order to validate the calibration accuracy, we generate about 500 3D testing points that are randomly distributed cover the whole working area following step 2 to step 4 in Algorithm 1, i.e., tracking and triangulating the laser dot. The testing points are different from the points used for calibration.

First we present the calibration results of the still camera with a short (18mm) and a long focal length (300mm) lenses. For simplicity, we assume the coordinate systems of pan-tilt camera and pan-tilt unit are aligned perfectly. This means R_* is an identity matrix. We use the Calibration Toolbox [19] to do the calibration for the reference view

of pan-tilt camera and stereo cameras. In order to reduce the ambiguities caused by the long focal length, we capture over 40 checkerboard patterns at different orientations for pan-tilt camera. Table 1 shows results of the intrinsic matrix K and RMS of calibration data. The uncertainties of the focal length with the 300mm lens is about 10 times larger than that with a 18mm lens although the RMS of calibration data for both cases are similar.

focal length	α	β	μ_0	ν_0	RMS (in pixels)
300mm	40869.2 ± 1750.2	41081.7 ± 1735.1	$1503.5 \pm *$	$999.5 \pm *$	3.85
18mm	2331.2 ± 9.1	2339.7 ± 9.1	1550.8 ± 12.1	997.9 ± 14.4	2.11

Table 1. The comparison between short and long focal length cameras. α and β are focal length. μ_0 and ν_0 are principal point. The uncertainties of principal point for 300mm camera cannot be estimated by Calibration Toolbox [19].

Fig.3 shows distributions of re-projection errors for the 500 testing points with 18mm and 300mm cameras. From this figure, we find that calibration is quite accurate for short focal length camera even that we assume R_* is an identity matrix. Given the high resolution image, the relative errors from 18mm and 300mm cameras are about 1.3% and 30% respectively. This is computed as the ratio of the mean pixel error and the image width. Furthermore, many of the test points are out of field of view of the 300mm camera.

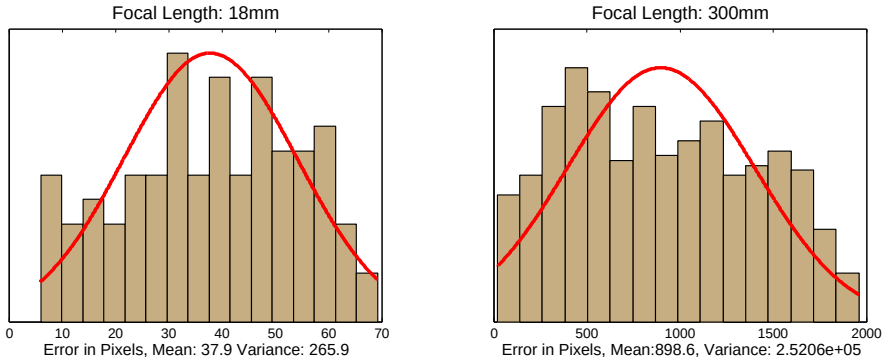

Fig. 3. Distributions of re-projection error (in pixels) based on 500 testing data for 18mm and 300mm pan-tilt cameras.

We then recalibrate the 300mm case with methods outlined in Algorithm 1 and 2, both of which include the estimation of R_*. About 600 3D points are sampled for calibration over the working area in Algorithm 1. We pre-calibrate the reference view for the pan-tilt camera as the initial guess. After calibration, we validate the accuracy with 500 3D points. Fig. 4 shows the distributions of re-projection errors from the two different methods. Our method is about 25 times better than the homography-based one. The relative errors from Algorithm 1 and 2 are about 1.2% and 29% respectively. It should be noted that R_* can not be estimated accurately from observed homographies. Hence, the percentage error from Algorithm 2 remains very large. In fact, the improvement

over assuming an identity R_* is little. Table 2 shows the results for intrinsic matrix K, θ_x, and θ_y after MLE refinement. Here we decompose R_* into two rotation matrices. One is the rotation around x axis for θ_x degree, and the other is the rotation around y axis for θ_y degree.

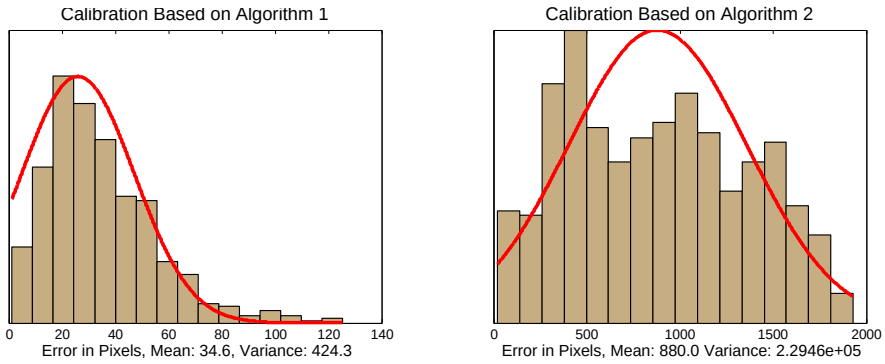

Fig. 4. Distributions of re-projection error based on 500 testing data (in pixels) for Algorithm 1 and 2.

Algorithm	α	β	μ_0	ν_0	θ_x	θ_y
1	40320.9	39507.7	1506.6	997.7	-0.14	-1.61
2	39883.3	40374.6	1567.5	1271.5	1.99	1.41

Table 2. The comparison between Algorithm 1 and 2. α and β are focal length. μ_0 and ν_0 are principal point. θ_x and θ_y are rotation angles between pan-tilt camera and pan-tilt unit.

5 Conclusion

This paper shows that calibration methods based on observed homographies are not suitable for cameras with telephoto (long-focal-length) lenses. This is caused by the ambiguities induced by the near-orthographic projection. We develop a method to calibrate a pan-tilt camera with long focal length in a surveillance network. In stead of using a large precisely-manufactured calibration object, our key idea is to use fixed stereo cameras to create a large collection of 3D calibration points. Using these 3D points allows full metric calibration over a large area. Experimental results show that the re-projection relative error is reduced from 30% to 1.2% with our method. In future work, we plan to extend our calibration method to auto-zoom cameras and build a complete surveillance system that can adjust zoom settings automatically by estimating the object's size.

References

1. Daugman, J.: How Iris Recognition Works. In: ICIP. (2002)

2. Guo, G., Jones, M., Beardsley, P.: A System for Automatic Iris Capturing. In: MERL TR2005-044. (2005)
3. Tsai, R.Y.: A Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology Using Off-The-Shelf TV Cameras and Lenses. IEEE Journal of Robotics and Automation **4**(3) (1987) 323–344
4. O.Faugeras: Three-Dimensional Computer Vision: a Geometric Viewpoint. MIT Press (1993)
5. Zhang, Z.: A Flexible New Technique for Camera Calibration. PAMI **22** (2000) 1330–1334
6. Heikkila, J., Silven, O.: A Four-Step Camera Calibration Procedure with Implicit Image Correction. In: proceedings of CVPR. (1997) 1106–1112
7. Pollefeys, M., Koch, R., Gool, L.V.: Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. In: proceedings of ICCV. (1997) 90–95
8. Pollefeys, M.: Self-calibration and metric 3D reconstruction from uncalibrated image sequences. PhD thesis, K.U.Leuven (1999)
9. Mundy, J., Zisserman, A.: Geometric Invariance in Computer Vision. MIT Press (1992)
10. Aloimonos, J.Y.: Perspective Approximations. In: Image and Vision Computing. Volume 8. (August 1990) 177–192
11. Quan, L.: Self-Calibration of an Affine Camera from Multiple Views. International Journal of Computer Vision **19**(1) (1996) 93–105
12. Hartley, R.I., Zisserman, A.: Multiple View Geometry. Cambridge University Press (2000)
13. Hartley, R.I.: "self-calibration of stationary cameras". International Journal of Computer Vision **1**(22) (1997) 5–23
14. de Agapito, L., Hayman, E., Reid, I.: Self-Calibration of a Rotating Camera with Varying Intrinsic Parameters. In: BMVC. (1998)
15. N.Sinha, S., Pollefeys, M.: Towards Calibrating a Pan-Tilt-Zoom Camera Network. In: OMNIVIS, the fifth Workshop on Omnidirectional Vision (in conjunction with ECCV 2004). (2004)
16. Wang, L., Kang, S.B.: Error Analysis of Pure Rotation-Based Self-Calibration. PAMI **2**(26) (2004) 275–280
17. Davis, J., Chen, X.: Calibrating pan-tilt cameras in wide-area surveillance networks. In: proceedings of ICCV. Volume 1. (2003) 144–150
18. Chen, X., Davis, J.: Wide Area Camera Calibration Using Virtual Calibration Objects. In: proceedings of CVPR. (2000)
19. Bouguet, J.Y.: Camera Calibration Toolbox for Matlab. `http://www.vision.caltech.edu/bouguetj/calib_doc/`
20. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **20** (2003) 91–110

Appendix A. Solving Pan and Tilt Angles

Here we discuss how to solve the pan and tilt angles so that the projection of an arbitrary point X in the 3D space is in the center of the image plane. We assume there is a rotation between the pan-tilt coordinate system and the camera's. Because of the dependency of the pan-tilt unit, that is, the tilt axis depends on the pan axis, the solution is not as simple as it appears. In order to address this problem, we back project the image center to a line $\tilde{L}_2$. The center of projection and point X form another line $\tilde{L}_1$. After the calibration steps described in Section 3, $\tilde{L}_1$ and $\tilde{L}_2$ are transformed into L_1 and L_2 in the coordinate system of the pan-tilt unit. Hence, the problem is simplified to panning around y-axis and tilting around x-axis to make L_1 and L_2 coincident or as close as possible to each other. If L_1 and L_2 are represented by the Plüker matrices, one

method to compute the transformation of an arbitrary 3D line to another line by performing only rotations around x and y axes could be a minimization of the following functional,

$$\text{argmin}_{R_x,R_y,\lambda} \|\mathbf{L_1} - \lambda \mathbf{L_2}\|^2 \tag{5}$$

where λ is a scalar, $\mathbf{L_2}$ is a 6×1 vector of the Plüker coordinates of L_2, and $\mathbf{L_1}$ is the 6×1 Plüker coordinates of the multiplication of $(R_y R_x) L_1 (R_y R_x)^T$, where R_x and R_y are rotation matrices around the x and y axes.

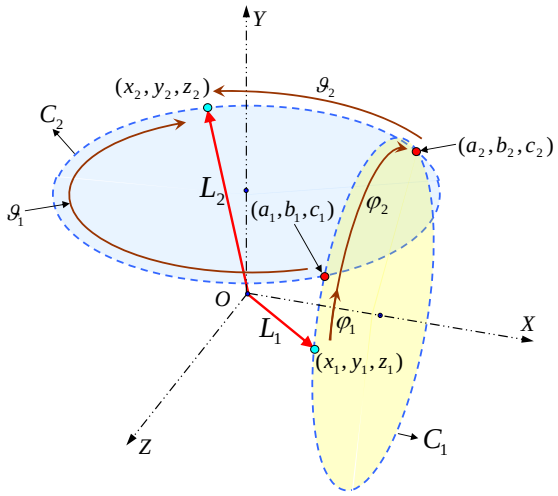

Fig. 5. Solve pan and tilt angles from L_1 to L_2.

However, the problem can be further simplified because L_1 and L_2 are intersected in the origin of the pan-tilt unit in our model. As shown in Fig. 5, we want to pan and tilt line L_1 to coincide with another line L_2. Assuming both of the two lines have unit lengths, the tilt angles are first computed by Eq. (6).

$$\begin{aligned} \varphi_1 &= \arctan\left(\frac{y_1}{z_1}\right) - \arctan\left(\frac{y_2}{r}\right) \\ \varphi_2 &= \arctan\left(\frac{y_1}{z_1}\right) - \arctan\left(\frac{y_2}{-r}\right) \\ r &= \sqrt{y_1^2 + z_1^2 - y_2^2} \end{aligned} \tag{6}$$

If $(y_1^2 + z_1^2 - y_2^2)$ is less than 0, two conics C_1 and C_2 are not intersected that means no exact solution exists. However, it almost never happens in practice since the rotation between the pan-tilt unit and the camera is small. After tilting, (x_1, y_1, z_1) is rotated to (a_1, b_1, c_1) or (a_2, b_2, c_2). Then the pan angles are computed by Eq. (7).

$$\begin{aligned} \vartheta_1 &= \arctan\left(\frac{z_2}{x_2}\right) - \arctan\left(\frac{c_1}{a_1}\right) \\ \vartheta_2 &= \arctan\left(\frac{z_2}{x_2}\right) - \arctan\left(\frac{c_2}{a_2}\right) \end{aligned} \tag{7}$$

Hence, two solutions, (φ_1, ϑ_1) and (φ_2, ϑ_2), are obtained. We choose the minimum rotation angles as the final solution.