A SPHERICAL RECTIFICATION FOR DUAL-PTZ-CAMERA SYSTEM

Dingrui Wan, Jie Zhou, Senior Member, IEEE,

Tsinghua University
Department of Automation
Beijing 100084, China

David Zhang, Senior Member, IEEE

Hong Kong Polytechnic University
Department of Computing
Kowloon, Hong Kong

ABSTRACT

In this paper we proposed a novel stereo rectification method for dual-PTZ-camera system, which is called spherical rectification. This method can be divided into two parts, offline pre-settings and online rectification. The offline pre-settings include camera calibration and building the spherical coordinate system. The online rectification only requires the pan, tilt and zoom values and does not need any image information such as feature points. So, compared with traditional rectification approaches, our method is more convenient and time-saving.

Index Terms— Stereo vision

1. INTRODUCTION

Stereo vision plays an important role in extracting 3D scene structure, in which the simplest case, dual-stereo vision, has been deeply studied for decades. The traditional standard stereo vision assumes that the stereo system satisfy nonverged geometry [1], i.e, the epipolar lines are parallel to each other. However, this assumption does not hold for many stereo systems. Stereo rectification is a method to make arbitrary stereo image pairs (i.e., with verged geometry) to become nonverged geometry [1]. Most algorithms of stereo rectification are based on epipolar constraint to map the epipolar lines to image scan lines and ensure the same scan lines in two images correspond to a specific epipolar line pair. This step is not necessary in stereo vision, but it is very useful in that it makes the search for corresponding image features to be confined to one dimension, and, hence, make the problem simplified [2].

Many stereo rectification approaches have been proposed in the past years [3]. Most of them are homography based (also called planar rectification) [4, 5]. Simplicity is a typical merit for this kind of approaches, while one of the shortcomings is that it does not preserve distances along epipolar lines. [6, 7] use more general warping functions to solve this problem. [6] proposed a cylinder rectification approach instead of the planar one, and [7] proposed a polar rectification method, which only requires the fundamental matrix. These approaches could preserve distances along epipolar lines but always be computationally expensive and do not preserve lines.

As PTZ cameras have been widely used in visual surveillance system for its flexibility of perspective and scale changing, dual-PTZ-camera system might be utilized instead of traditional dual-camera system. The dual-PTZ-camera system could obtain both global and local image features, and if stereo vision could be applied in dual-PTZ-camera system, the application scope will become much broader. But as far as we know, few articles considered stereo vision for dual-PTZ-camera system. We believe that a proper rectification approach might help solving this problem. In dual-PTZ-camera system, traditional rectification approaches as listed above could not be directly used, because first it's difficult to guarantee the precision under some large difference between cameras' FOV, and second, when PTZ parameters change, the rectification parameters should be calculated over again. That will be time consuming and unstable for real applications.

In this paper, we propose a spherical rectification approach to deal with dual-PTZ-camera system. We first use the camera model to map the image plane to a virtual spherical surface according to current PTZ parameters, and then the rectification is applied directly from the sphere to rectified plane. The sphere is independent of PTZ parameters, so when PTZ parameters change, we only need to find the corresponding region on the spherical surface, and use the pre-calculated rectification parameters to wrap the sphere to the goal plane. As this method could avoid the recomputation of rectification parameters, it is convenient for PTZ image pairs' stereo rectification. Following the idea of preserving the distance along epipolar line in [6, 7], we improve the basic spherical rectification algorithm so that it could maintain the disparities between two images, and so, traditional stereo matching procedure can be followed to estimate the depth of the scene.

2. PTZ CAMERA MODEL AND CALIBRATION

Calibration of PTZ camera plays an important role in the vision computing by using PTZ cameras. Our method is similar to [8] which is inherently feature-based, but our method

This work was supported by Natural Science Foundation of China under grant 60673106 and 60573062, and the Specialized Research Fund for the Doctoral Program of Higher Education.

combines the parameters inquired from active camera so that our method much more simple, and the precision is related to the precision of parameters inquired from active camera. For simplicity, we do not consider either focal length or radial distortion.

2.1. PTZ camera model

We chose to use a common camera model as below (In our study, we use SONY EVI D70 camera which can be well represented by this model.),

$$\tilde{x} = \kappa K[\ R \quad -Rt\ ]\tilde{X}, K = \begin{bmatrix} \alpha f & s & u_0 \\ & f & v_0 \\ & & 1 \end{bmatrix}, \qquad (1)$$

where x and X are image coordinates and world coordinates, respectively; symbol ' ˜ ' means homogeneous coordinate. α and s are the camera's pixel aspect ratio and skew respectively; f is the equivalent focal length measured in pixel; (u_0, v_0) is the principal point in the image. In order to simplify the PTZ camera model, we assume:

(1) The center of rotation of the camera is fixed;

(2) For PTZ camera, $t = 0$;

(3) Aspect ratio $\alpha = 1$, and skew $s = 0$;

(4) Principal point (u_0, v_0) is replaced by the zoom center [9] approximately.

According to the assumptions, the camera model can be written as

$$\tilde{x} = \kappa KRX, \qquad (2)$$

where the intrinsic K could write in diagonal form by translating the origin of image to (u_0, v_0).

2.2. PTZ camera Calibration

2.2.1. Zoom center estimation

Simply tracking the feature points in an image sequence with varying zoom level (from z_{min} to z_{max}) with fixed pan and tilt parameters. In our experiment, $(u_0, v_0) = (150.4,\ 127.0)$, while the image coordinates is $[1, 320]$ for x-axis and $[1, 240]$ for y-axis.

2.2.2. R and K estimation

The rotation matrix R can be directly calculated given pan and tilt value which can be inquired from the camera for SONY EVI D70. This formula can be found in many literatures about PTZ cameras. As the intrinsic $K = diag\{k_z, k_z, 1\}$ has only one degree of freedom which related to the zoom value, we use [8]'s approach to estimate K at several discrete zoom levels, and then choose a proper model to fit these samples. We choose the Equ.3 as the approximate model, and the four unknown parameters can be solved by using curve fitting tools. This model works well in our experiment.

$$k_z = a \exp(bz) + c \exp(dz). \qquad (3)$$

3. SPHERICAL RECTIFICATION

3.1. Basic Notation

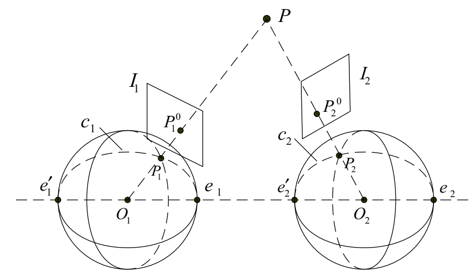

Fig. 1. The sketch map of spherical rectification.

Each pixel on the image plane can be mapped onto a unit spherical surface with the center coincide with the camera optic center, see Fig.1. Actually, the mapped point is the intersection between the sphere and the line which pass the optic center and the given pixel on the image plane. We can define following concepts:

Epipolar plane: $\Pi(O_1, O_2, P)$. All the epipolar planes pass the line O_1O_2.
Epipolar circle: c_1 and c_2, the intersection curve between the epipolar plane $\Pi(O_1, O_2, P)$ and the unit sphere ϕ_i.
Epipole: the intersection between the line O_1O_2 and the sphere. So there exits two epipoles, $e_i, e'_i, i = 1, 2$, for each sphere.

The spherical rectification mainly constitutes 2 steps: first, use the camera model to map the image plane to the unit spherical surface; second, warp the valid sphere region to image plane which is the rectified image.

3.2. Image Plane to Sphere

Let $x = [u\ v]^T$ and $X = [\alpha_x\ \beta_x]^T$ be the image coordinate and corresponding sphere coordinate respectively. α_x and β_x could be seemed as the longitude and latitude, while e and e' are the two poles, see Fig.2. We denote $\Pi(O, \perp ee')$ as the plane passing the sphere center and perpendicular to the line ee', and it intersects the sphere ϕ at a unit circle $c_\perp$. Given an arbitrary point M on $c_\perp$, and plane $\Pi(e, e', X)$ intersects $c_\perp$ at two points. We choose the one which is closer to X, and we denote it as X'. Then α_x is defined as the angle from OM to OX', where $\alpha_x \in [-\pi, \pi)$, and β_x is defined as the angle from Oe to OX, where $\beta_x \in [0, \pi]$. We call OM as the reference vector.

In order to calculate X, we first calculate the world coordinate Y_x by the camera model $Y_x = \kappa R^{-1}K^{-1}\tilde{x}$, where κ is a scale factor and can be simply set to be 1. R and K can be obtain from the calibration result. We normalize Y_x s.t $|OY_x| = 1$, and then Y_x is located on the unit spherical surface. If the world coordinate of e, e' and OM are known, $X = [\alpha_x\ \beta_x]^T$ can be easy calculated from definition.

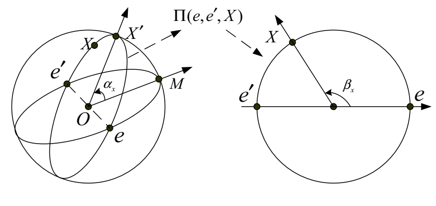

Fig. 2. The definition of α_x and β_x.

3.3. Sphere to Rectified Image

There are several ways to warp the sphere to a plane, i.e. $X_r = [u' \; v']^T = [f_\alpha(\alpha_x) \; f_\beta(\beta_x)]^T$, and we simply choose the linear transformation. If the valid range of α_x and β_x, and the unit length $\Delta\alpha_u$ and $\Delta\beta_u$ associated to one pixel in rectified image are known, $f_\alpha()$ and $f_\beta()$ can be settled. In order to reduce computation, we only examine the four corners of the original image to decide the valid range of α_x and β_x. It is more complicated to decide $\Delta\alpha_u$ and $\Delta\beta_u$, and our target is to minimize the loss of pixel information [7], i.e. every displacement with one unit length on the sphere surface will cause a displacement whose length is less than 1 pixel in the original image. From the definition of the coordinates, $\Delta\alpha_u$ should be estimated at the location with a maximal value of $|\sin\beta|$, and $\Delta\beta_u$ can be estimated anywhere. For the sake of simplicity, we also only examine the four corners, see Fig. 3.

Assume $C1$ gets the maximum of $|\sin\beta|$, where $x_{C1} = (u_1, v_1)$, $X_{C1} = (\alpha_1, \beta_1)$, and $C1'$ is the point after applying a small displacement on $C1$ s.t. $X_{C1'} = (\alpha_1', \beta_1)(\alpha_1' \neq \alpha)$ and $x_{C1'} = (u_1', v_1')$. Let $d_0(C1, C1')$ be the distance measurement between $C1$ and $C1'$ in the original image, then we have

$$\Delta\alpha_u \doteq |\alpha_1 - \alpha_1'| / d_0(C1, C1').$$

Similarly, $\Delta\beta_u$ can be estimated at arbitrary location, such as $C3$ in Fig. 3.

Then the new coordinate in rectified image is

$$X_p = [u', v']^T = [(\alpha - \alpha_{\min})/\Delta\alpha_u, (\beta - \beta_{\min})/\Delta\beta_u]^T \quad .$$

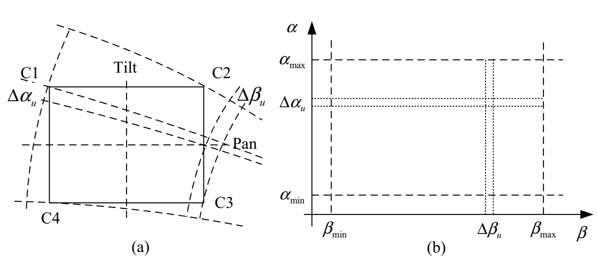

Fig. 3. The mapping from spherical coordinate to rectified image coordinate.

In order to make the same scan line in two rectified images corresponding to the same epipolar plane, we have to, first, construct the relationship between two sphere coordinates; second, choose the same $\Delta\alpha_u$ and $\Delta\beta_u$, i.e. $\Delta\alpha_u = \max(\Delta\alpha_{u1}, \Delta\alpha_{u2})$, $\Delta\beta_u = \max(\Delta\beta_{u1}, \Delta\beta_{u2})$.

3.4. Disparity Preserved Rectification

The (α, β) rectification mentioned in previous section has achieved the basic goal of stereo rectification, but the disparity obtained by traditional stereo matching approach could not reflect the depth of the scene, i.e. the (α, β) rectification could not preserve disparity. In this section we proposed an improved method called the (α, γ) rectification to solve this problem.

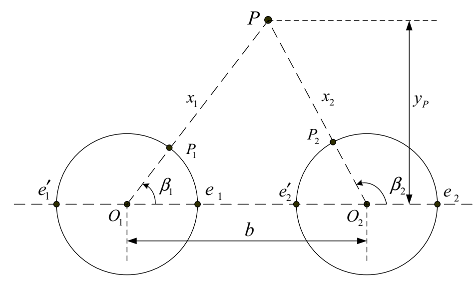

Fig. 4. The sketch map of depth in spherical rectification.

Let P be a point in the scene, and the sphere coordinates of two cameras are (α_1, β_1) and (α_2, β_2), respectively, see Fig. 4. Obviously, (α_1, β_1) and (α_2, β_2) are located on the epipolar plane. Assume we have found a proper reference vector O_2M_2 so that $\alpha_1 = \alpha_2$. Let y_P be the distance between P to the line O_1O_2, and $O_1P = x_1$, $O_2P = x_2$, $O_1O_2 = b$. Then we have

$$\begin{cases} x_1 \sin\beta_1 = x_2 \sin\beta_2 = y_P \\ x_1 \cos\beta_1 - x_2 \cos\beta_2 = b \end{cases} . \tag{4}$$

So, y_P can be solved as

$$y_P = \frac{b}{\cot\beta_1 - \cot\beta_2} = \frac{-b}{\gamma_1 - \gamma_2}, \tag{5}$$

where $\gamma_i = -\cot\beta_i, i = 1, 2$, and $\gamma_1 - \gamma_2$ is the disparity. Equ.5 is similar to the classical expression, which shows the relationship between disparity and depth. As $|\cot\beta| \to \infty(\beta \to 0 \;\; or \;\; \pi)$, the distortion of rectified image might be obvious, so this rectification method could not deal with the case that one of the epipoles located in the image.

For (α, γ) rectification, we use similar method to estimate $\alpha_{\max}$, $\alpha_{\min}$, $\gamma_{\max}$, $\gamma_{\min}$, $\Delta\alpha_u$ and $\Delta\gamma_u$, where $\Delta\gamma_u$ should be estimated at the location when $|\sin\beta|$ reaches its minimum value.

3.5. Construction of Sphere Coordinates

Given two similar images I_1 and I_2 from two PTZ cameras, we can calculate the fundamental matrix F (Harris Corner and RANSAC method) and traditional epipoles e_1 and e_2. The fundamental matrix F is a 3×3 rank-2 matrix that maps points in I_1 to lines in I_2, and F satisfies $Fe_1 = F^T e_2 = 0$, where $e_1 \in I_1$ and $e_2 \in I_2$. e_1, e_2, O_1 and O_2 are collinear.

For the ith camera, in O_i's coordinate system, we can calculate the world coordinate of e_i with a scale factor, while e_i' is the symmetrical point of e_i about O_i. In order to enhance the accuracy and robustness, the mean value of several results is more reasonable.

According to the definition of reference vector, it is not unique to choose O_iM_i. Assume α_1 and α_2 correspond to the same epipolar plane, and $\alpha_1 = \alpha_2 + \alpha_0$, where α_0 is a constant. For convenience, we can adjust one of the reference vector to make $\alpha_0 = 0$ so that the same scan line in two rectified images corresponding to the same epipolar plane. Because the feature points are located on the same epipolar plane, this could help the adjustment through a simple optimization approach.

4. EXPERIMENTAL RESULTS

In our experiment, we utilize the Sony EVI D70 camera as PTZ camera. We take the (α, β) rectification for example. For two images $I_1(PTZ_1)$ and $I_2(PTZ_2)$, we first estimate $\alpha_{\max}, \alpha_{\min}, \beta_{\max}, \alpha_{\min}, \Delta\alpha_u$ and $\Delta\beta_u$ to construct the $\alpha - \beta$ coordinate system, and let I_1^r and I_2^r be the rectified images; second, traverse all points $S_p(i, j) \in I_1^r$, and following a series of coordinate transformation, we can get the sphere coordinate (α_p, β_p), the camera coordinate X_p and the original image coordinate $x_p(u_p, v_p)$. Choose a proper interpolation method to estimate the gray level at $x_p(u_p, v_p)$. This gray level is assigned to (i, j) in I_1^r. Perform the same procedure for I_2^r.

We list two sets of results in Fig.5. The original image size is 320×240, and the two rectified images are normalized to the same size. From the results, we can see that both the (α, β) and (α, γ) rectification method achieve the goal of stereo rectification. Comparing the two rectification method, we found that the density along the width of image is different. As $\gamma = -\cot\beta$, whenβ is closed to $\pi/2$, the difference between the two methods is tiny, like Fig.5(b); otherwise, the difference become bigger, like Fig.5(a). This difference reflects the way to preserve the disparity-depth relationship. For space limitation, we omit the introduction of disparity-depth validation experiment for (α, γ) rectification.

5. REFERENCES

[1] Myron Z. Brown, Darius Burschka, and Gregory D. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, 2003.

[2] D. Papadimitriou and T. Dennis., "Epipolar line estimation and rectification for stereo images pairs," *IEEE Trans. Image Processing*, vol. 3, no. 4, pp. 672–676, 1996.

(a)

(b)

Fig. 5. (a) and (b) are two experiment results. For each result, the left column shows the original image pair, and the (α, β) and (α, γ) rectification results are listed at the right-top and right-bottom, respectively.

[3] Marc Pollefeys and Sudipta N. Sinha, "Iso-disparity surfaces for general stereo configurations.," in *ECCV (3)*, 2004, pp. 509–520.

[4] Charles T. Loop and Zhengyou Zhang, "Computing rectifying homographies for stereo vision.," in *CVPR*, 1999, pp. 1125–1131.

[5] Richard I. Hartley, "Theory and practice of projective rectification," *Int. J. Comput. Vision*, vol. 35, no. 2, pp. 115–127, 1999.

[6] Sébastien Roy, Jean Meunier, and Ingemar J. Cox, "Cylindrical rectification to minimize epipolar distortion.," in *CVPR*, 1997, pp. 393–399.

[7] Marc Pollefeys, Reinhard Koch, and Luc J. Van Gool, "A simple and efficient rectification method for general motion.," in *ICCV*, 1999, pp. 496–501.

[8] S. Sinha and M. Pollefeys, "Towards calibrating a pan-tilt-zoom cameras network," in *The fifth Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras*, 2004.

[9] Mengxiang Li and Jean-Marc Lavest, "Some aspects of zoom lens camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 11, pp. 1105–1110, 1996.