Proceedings of the 2017 IEEE
International Conference on Information and Automation (ICIA)
Macau SAR, China, July 2017

Dense 3D Reconstruction with an Active Binocular Panoramic Vision System

Honglong Zhang, Danrong Li and Zhan Song
Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology
Shenzhen Institutes of Advanced and Technology, Chinese Academy of Science
Shenzhen, Guangdong, China, 518055
{hl.zhang1; dr.li; zhan.song}@siat.ac.cn

Abstract –This paper presents a novel 3D panoramic system which is composed with two panoramic cameras and a laser line projector. By rotating the laser line, a serial of panoramic images with laser strip projections can be obtained. By calibrating the panoramic cameras, epipolar constrain can be used to precisely solve the corresponding points on the laser strips in the panoramic images. As a result, a dense 3D reconstruction result can be obtained based on the binocular panoramic cameras. In the experiment, a complex indoor scene is used. The experimental results show that, the scenario can be well reconstructed by our method in comparison with classical stereo matching approaches. With proposed method, not only panoramic images can be generated, but also dense 3D information can be reconstructed, which has great potentials in various applications.

Index Terms – Panoramic image, 3D reconstruction, spherical stereo, calibration.

I. Introduction

Panoramic cameras have been widely used in the virtual reality related applications. Conventional panoramic systems, usually adopt multiple wide-angle cameras to capture the scenario from different view directions. By correcting the image distortion and stitching multiple images, a panoramic image can be generated. The panoramic image can provide broader view angles than conventional cameras. However, there still few work to address the problem of 3D reconstruction based on panoramic cameras [1].

3D reconstruction has been a classical research topic in the computer vision domain. Classical 3D reconstruction method include photometric stereo methods [2, 3], stereo vision method [4, 5], structure from motion [6], and structured light methods [7, 8] etc. A lot of stereo vision methods have been proposed in the past decades, but most of them are based on conventional cameras which are equipped with narrow view-angle lens. In comparison, panoramic cameras usually have much larger view-angles, even more than 180°. Therefore the pin-hole model that adopted by conventional stereo vision systems cannot be used subject to the serious distortion of the panoramic images. In [9], two panoramic image sensors were used to obtain the full-view image around the vehicle. With the image rectification and object detection methods, useful information can be extracted from the panoramic images to realize aided or automatic driving. A panoramic camera-based SLAM (Simultaneous Localization and Mapping) method was proposed in [10]. Based on the structure from motion method, sparse 3D information of the environment around the robot can be calculated and then used for the guiding of mobile robots. In [11], a spherical stereo vision method was proposed. The spherical stereo disparity image was generated to describe depth of the scene. In [12-14], some panoramic methods were also studied for the immersive virtual reality applications.

Panoramic camera lens has a wide field of view, which brings broader viewing field as well as larger image distortions. To realize precise panoramic image rectification, calibration is a critical issue. In [15], a general fish-eye camera model and calibration method for conventional wide-angel and fish-eye lenses was introduced. An accurate and easy-to-use calibration method for spherical stereoscopic lens was proposed in [16]. Although a lot of panoramic camera-based 3D vision systems have been investigated, but dense and precise 3D reconstruction based on panoramic cameras is still an open issue. The major difficulty in binocular 3D vision methods is the corresponding problem. Accuracy of the 3D reconstruction results is mostly related with the computed disparity map. For the panoramic camera based stereo vision systems, the corresponding problem become more seriously. Large distortions of the panoramic images make the corresponding problem more difficult. As a result, existing panoramic stereo approaches can only provide very sparse 3D information.

In this paper, a structured light device, i.e. a laser line projector is incorporated with a binocular panoramic stereo vision system. By rotating the laser line, a serial of panoramic images with laser strip projections can be obtained. By calibrating the panoramic cameras, epipolar constrain can be used to precisely solve the corresponding points on the laser strips in the panoramic images. As a result, a dense 3D reconstruction result can be obtained based on the binocular panoramic cameras. Some real experiments are implemented based on the proposed 3D panoramic system, and the results are compared with existing methods to demonstrate its feasibility and 3D reconstruction quality.

This work was supported in part by the National Natural Science Foundation of China (Nos. 61375041, U1613213), Shenzhen Science Plan (JCYJ20150401150223645, JSGG20150925164740726) and The CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems.

978-1-5386-3154-6/17/$31.00 ©2017 IEEE

II. Methodology

The panoramic camera model is based on the unified fish-eye lens model as proposed in [17-18]. This model is used for catadioptric system and has been proved effective to approximate fisheye projection model [19]. The unified fish-eye lens model is as illustrated by Fig. 1. Suppose P indicate a world point in 3D space, the coordinate of P is expressed as (x, y, z) in the spherical coordinate system. The point C is the projection center of spherical model, C_1 is the new projection center shifted along z axis by $-\xi$. The projection procedure can be simply described as two steps: 1) the world point P projects onto a camera-centered unit sphere at the point P_1 where C is the projection center; 2) the point P_1 projects onto an image plane at the point P_2 by pinhole projection where C_1 is the projection center.

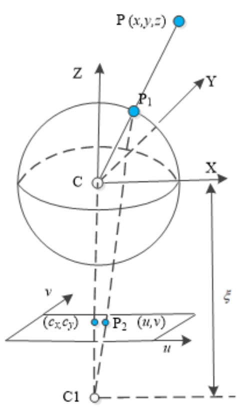

Fig. 1. The unified spherical camera model proposed in [17-18].

With above camera model, the image point in pixel coordinate system can be computed as:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \pi_u(\mathrm{x}) = \begin{bmatrix} f_x \dfrac{x}{z + \|\mathrm{x}\|\xi} \\ f_y \dfrac{y}{z + \|\mathrm{x}\|\xi} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \tag{1}$$

where $\|\mathrm{x}\|$ is the Euclidean norm of x, (fx, fy) is the camera focal length, (cx, cy) is the sensor principal point. Therefore, corresponding un-projection function can be formulated as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \pi_u^{-1}(u', v') = \left(\frac{\xi + \sqrt{1 + (1-\xi^2)(u'^2 + v'^2)}}{u'^2 + v'^2 + 1} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \xi \end{bmatrix} \right) \tag{2}$$

where,

$$\begin{bmatrix} \mathrm{u}' \\ v' \end{bmatrix} = \begin{bmatrix} (u - c_x)/f_x \\ (v - c_y)/f_y \end{bmatrix}. \tag{3}$$

A. Spherical Stereo Vision Model

The binocular panoramic camera model can be illustrated as Fig. 2. For the 3D point P, P_1, P_2 are its projection points on the left and right unit sphere coordinates respectively, P_1', P_2' are the corresponding projection points of P in the camera coordinates, P_1'', P_2'' are the image pixels of P without distortion, P_1''', $P2'''$ are the image pixels with distortion.

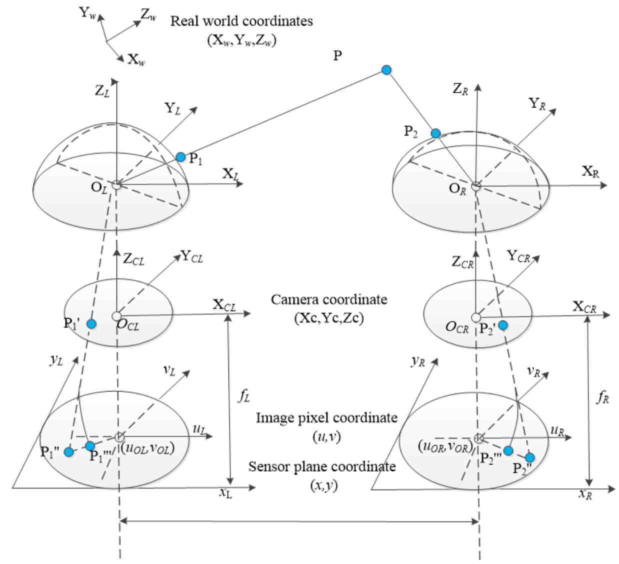

Fig. 2. Diagram of the binocular panoramic stereo cameras.

In order to describe the relationship between two cameras of spherical stereo vision system, we need to formulate the information as shown in Fig. 2. For a world point P, if its non-homogeneous coordinates in the real world coordinates is X_w, X_{SL}, X_{SR} are the left and right camera coordinates of P respectively, we can obtain following equations:

$$X_{SL} = R_{w\to SL} X_w + t_{w\to SL} \tag{4}$$

$$X_{SR} = R_{w\to SR} X_w + t_{w\to SR} \tag{5}$$

where $R_{w\to SL}$ and $t_{w\to SL}$ are the rotation matrix and translation vector from world coordinates to left camera coordinates, $R_{\mathrm{w}\to\mathrm{SR}}$ and $t_{\mathrm{w}\to\mathrm{SR}}$ are the corresponding transformation matrix from world coordinates to right camera coordinates. By eliminating X_w, the relationship between two cameras can be expressed as:

$$X_{SL} = R_{w\to SL} R_{w\to SR}^{-1} X_{SR} - R_{w\to SR}^{-1} t_{w\to SR} + t_{w\to SL} \tag{6}$$

Thus, we can describe the relationship between two cameras using $R_{L\to R}$ and translation vector $t_{L\to R}$ as:

$$R_{L\to R} = R_{w\to SL} R_{w\to SR}^{-1} \tag{7}$$

$$t_{L\to R} = t_{w\to SL} - R_{w\to SR}^{-1} t_{w\to SR} \tag{8}$$

The ideal spherical stereo is similar to conventional planar stereo where x axis of the rectified camera coordinates is aligned with each other and the rest axis of y and z are parallel respectively as shown in Fig. 3.