

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



A Novel Omnidirectional Stereo Vision System via a Single Camera¹

Chuanjiang Luo, Liancheng Su & Feng Zhu
*Shenyang Institute of Automation, Chinese Academy of Sciences
P.R. China*

1. Introduction

Obtaining panoramic 3D map information for mobile robots is essential for navigation and action planning. Although there are other ways to fulfill this task, such as ultrasonic sensors or laser range finders, stereo vision system excels them in its precision and real-time speed without energy emission.

But the conventional stereo vision systems are limited in their fields of view (FOV). An effective way to enhance FOV is to construct an omnidirectional vision system using mirrors in conjunction with perspective cameras. These systems are normally referred to as catadioptric and have been applied to robot localization and navigation by several researchers (Bunschoten & Krose, 2002; Menegatti et al., 2004). A common constraint upon the omnidirectional sensors modeling requires that all the imaged rays pass through a unique point called single viewpoint (SVP) (Baker & Nayar, 1999). The reason a single viewpoint is so desirable is that it is a requirement for the generation of pure perspective images from the sensed images. These perspective images can subsequently be processed using the vast array of techniques developed in the field of computer vision that assume perspective projection. The mirrors popularly used to construct wide FOV catadioptric are hyperbolic or parabolic. But the latter must be coupled with expensive telecentric optics which restricts them to limited applications in panoramic vision.

Mobile robot navigation using binocular omnidirectional stereo vision has been reported in (Menegatti et al., 2004; Yagi, 2002; Zhu, 2001). Such two-camera stereo systems can be classified as horizontal stereo systems and vertical stereo systems according to their cameras' configuration. In (Ma, 2003), the cameras are configured horizontally and the baseline of triangulation is in the horizontal plane. This configuration brings two problems: one is that the epipolar line becomes curved line leading to increasing computational cost; the other is that the accuracy of the 3D measurement depends on the direction of a landmark. In the omnidirectional stereo vision system (Gluckman et al., 1998; Koyasu et al., 2002; Koyasu et al., 2003), two omnidirectional cameras are vertically arranged. Such configuration escapes the shortcomings brought by horizontal stereo system, but the power cable and data bus introduce occlusion to the images captured by this configuration. In

¹ This work is supported by National Science Foundation of P.R. China (60575024).

addition, two-camera stereo systems are costly and complicated besides having the problem of requiring precise positioning of the cameras.

Single camera stereo has several advantages over two-camera stereo. Because only a single camera and digitizer are used, system parameters such as spectral response, gain, and offset are identical for the stereo pair. In addition, only a single set of intrinsic parameters needs to be determined. The prominent advantage of single camera stereo over two-camera configuration is that it does not need data synchronization. Omnidirectional stereo based on a double lobed mirror and a single camera was developed in (Southwell et al., 1996; Conroy & Moore, 1999; Cabral et al., 2004). A double lobed mirror is a coaxial mirror pair, where the centers of both mirrors are collinear with the camera axis, and the mirrors have a profile radially symmetric around this axis. This arrangement has the advantage to produce two panoramic views of the scene in a single image. But the disadvantage of this method is the relatively small baseline it provides. Since the two mirrors are so close together, the effective baseline for stereo calculation is quite small.

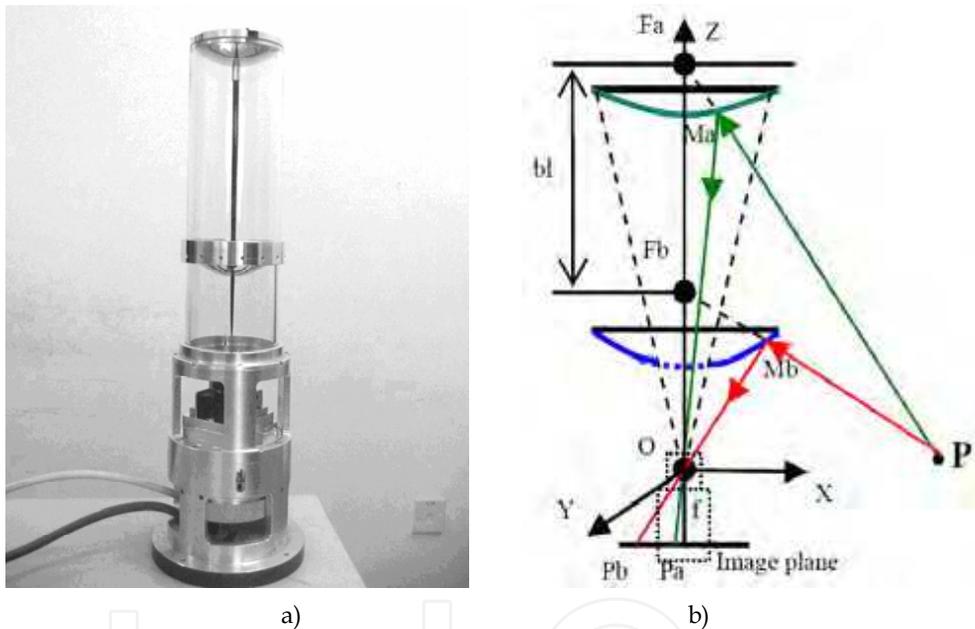


Figure 1. a) The appearance of the stereo vision system. b) The configuration of the system

To overcome this drawback, we have proposed a novel large baseline panoramic vision system in this chapter. We will describe in detail how to use this vision system to obtain reliable 3D depth maps of surrounding environment. In the subsequent arrangement of this chapter, Section 2 is dedicated to describe the principle of our catadioptric stereo vision system. Following that, a full model of calibrating the system including the rotation and translation between the camera and mirrors is presented in Section 3. In Section 4, a three-step method that combines the merit of feature matching and dense global matching is proposed to get a fast and reliable matching result and eventually the 3D depth map. Finally, we will give a brief evaluation of our system and some ideas for our future work in the summary.

2. Principle of Our Vision System

The system we have developed (Su & Zhu, 2005) is based on a common perspective camera coupled with two hyperbolic mirrors, which are separately fixed inside a glass cylinder (Fig.1a). The two hyperbolic mirrors share one focus which coincides with the camera center. A hole in the below mirror permits imaging via the mirror above. As the separation between the two mirrors provides much enlarged baseline, the precision of the system has been improved correspondingly. The coaxial configuration of the camera and the two hyperbolic mirrors makes the epipolar line radially collinear, thus making the system free of the search process for complex epipolar curve in stereo matching (Fig. 3).

To describe the triangulation for computing 3D coordinates of space points, we define the focal point O as the origin of our reference frame, z-axis parallel to the optical axis pointing above. Then mirrors can be represented as:

$$\frac{(z-c_i)^2}{a^2} - \frac{(x^2+y^2)}{b^2} = 1, \quad (i=1,2) \quad (1)$$

Only the incident rays pointing to the focus $F_a(0,0,2c_a)$, $F_b(0,0,2c_b)$ will be reflected by the mirrors to pass through the focal point of the camera. The incident ray passing the space point $P(x,y,z)$ reaches the mirrors at points M_a and M_b , being projected onto the image at points $P_a(u_a,v_a,-f)$ and $P_b(u_b,v_b,-f)$ respectively. As P_a and P_b are known, M_a and M_b can be represented by:

$$\frac{x_{M_i}}{u_i} = \frac{y_{M_i}}{v_i} = \frac{z_{M_i}}{-f}, \quad (i=1,2) \quad (2)$$

Since point M_a and M_b are on the mirrors, they satisfy the equation of the mirrors. Their coordinates can be solved from equation group (1) and (2). Then the equation of rays F_aP and F_bP are:

$$\frac{x_p}{x_i} = \frac{y_p}{y_i} = \frac{z_p - 2c_i}{z_i - 2c_i}, \quad (i=1,2) \quad (3)$$

We can finally figure out coordinate of the space point P by solving the equation (3).

3. System Calibration

3.1 Overview

In using the omnidirectional stereo vision system, its calibration is important, as in the case of conventional stereo systems (Luong & Faugeras, 1996; Zhang & Faugeras, 1997). We present a full model of the imaging process, which includes the rotation and translation between the camera and mirror, and an algorithm to determine this relative position from observations of known points in a single image.

There have been many works on the calibration of omnidirectional cameras. Some of them are for estimating intrinsic parameters (Ying & Hu, 2004; Geyer & Daniilidis, 1999; Geyer Daniilidis, 2000; Kang, 2000). In (Geyer & Daniilidis, 1999; Geyer Daniilidis, 2000), Geyer & Daniilidis presented a geometric method using two or more sets of parallel lines in one

image to determine the camera aspect ratio, a scale factor that is the product of the camera and mirror focal lengths, and the principal point. Kang (Kang, 2000) describes two methods. The first recovers the image center and mirror parabolic parameter from the image of the mirror's circular boundary in one image; of course, this method requires that the mirror's boundary be visible in the image. The second method uses minimization to recover skew in addition to Geyer's parameters. In this method the image measurements are point correspondences in multiple image pairs. Miousik & Pajdla developed methods of calibrating both intrinsic and extrinsic parameters (Miousik & Pajdla, 2003a; Miousik & Pajdla, 2003b). In (Geyer & Daniilidis, 2003), Geyer & Daniilidis developed a method for rectifying omnidirectional image pairs, generating a rectified pair of normal perspective images.

Because the advantages of single viewpoint cameras are only achieved if the mirror axis is aligned with the camera axis, these methods mentioned above all assume that these axes are parallel rather than determining the relative rotation between the mirror and camera. A more complete calibration procedure for a catadioptric camera which estimates the intrinsic camera parameters and the pose of the mirror related to the camera appeared at (Fabrizio et al., 2002), the author used the images of two known radius circles at two different planes in an omnidirectional camera structure to calibrate the intrinsic camera parameters and the camera pose with respect to the mirror. But this proposed technique cannot be easily generalized to all kinds of catadioptric sensors for it requires the two circles be visible on the mirror. Meanwhile, this technique calibrated the intrinsic parameters combined to extrinsic parameters, so there are eleven parameters (five intrinsic parameters and six extrinsic parameters) need to be determined. As the model of projection is nonlinear the computation of the system is so complex that the parameters cannot be determined with good precision. Our calibration is performed within a general minimization framework, and easily accommodates any combination of mirror and camera. For single viewpoint combinations, the advantages of the single viewpoint can be exploited only if the camera and mirror are assumed to be properly aligned. So for these combinations, the simpler single viewpoint projection model, rather than the full model described here, should be adopted only if the misalignment between the mirror and camera is sufficiently small. In this case, the calibration algorithm that we describe is useful as a software verification of the alignment accuracy.

Our projection model and calibration algorithm separate the conventional camera intrinsics (e.g., focal length, principal point) from the relative position between the mirrors and the camera (i.e., the camera-to-mirrors coordinate transformation) to reduce computational complexity and improve the calibration precision. The conventional camera intrinsics can be determined using any existing method. For the experiments described here, we have used the method implemented in http://www.vision.caltech.edu/bouguetj/calib_doc/. Once the camera intrinsics are known, the camera-to-mirrors transformation can be determined by obtaining an image of calibration targets whose three-dimensional positions are known, and then minimizing the difference between coordinates of the targets and the locations calculated from the targets' images through the projection model. Fig. 3 shows one example of calibration image used in our experiments. The locations of the three dimensional points have been surveyed with an accuracy of about one millimeter. If the inaccuracy of image point due to discrete distribution of pixels is taken into account, the total measuring error is about five millimeters.

3.2 Projection Model

Fig. 2 depicts the full imaging model of a perspective camera with two hyperbolic mirrors. There are three essentially coordinate systems.

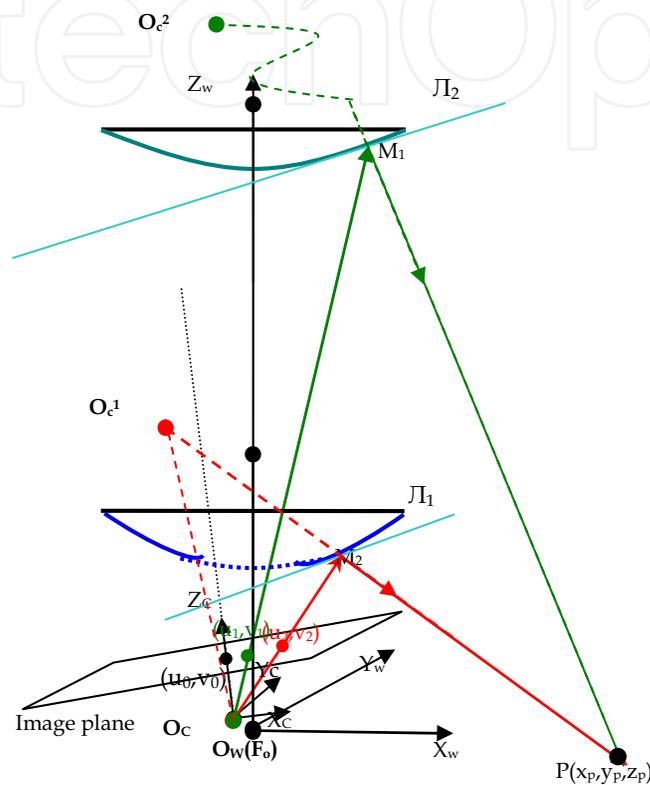


Figure 2. The projection model of the omnidirectional stereo vision system. There are transformations between the camera coordinate system and the mirror (or world) coordinate system

1. The camera coordinate system centered at the camera center O_c , the optical axis is aligned with the z-axis of the camera coordinate system;
2. The mirror system centered at common foci of the hyperbolic mirrors F_o , the mirrors axes is aligned with the z-axis of the mirror coordinate system (We assume that the axes of the mirrors are aligned well, and the common foci are coincident, from the mirrors manufacturing sheet we know it is reasonable);
3. The world system centered at O_w . The omnidirectional stereo vision system was placed on a plane desk. As both the base of vision system and desk surface are plane, the axis of the mirror is perpendicular to the base of the system and the surface of the desk feckly. We make the mirror system coincide with the world system to simplify the model and computations.

So the equations of hyperboloid of two sheets in the system centered at O_w are the same as equation (1). For a known world point $P(x_w, y_w, z_w)$ in the world (or mirror) coordinate system whose projected points in the image plane are also known, $q_1(u_1, v_1)$ and $q_2(u_2, v_2)$ are respectively projected by the upper mirror and bellow mirror. Then we get their coordinates in the camera coordinate system:

$$\begin{bmatrix} x_i^c \\ y_i^c \\ z_i^c \end{bmatrix} = \begin{bmatrix} (u_i - u_0)k_u \\ (v_0 - v_i)k_v \\ f \end{bmatrix}, \quad (i=1,2) \quad (4)$$

Where f is the focal length; k_u and k_v are the pixel scale factors; u_0 and v_0 are the coordinates of the principal point, where the optical axis intersects the projection plane. They are intrinsic parameters of the perspective camera.

So the image points $P_c(x_i^c, y_i^c, z_i^c)$ of the camera coordinate system can be expressed relative to the mirror coordinate system as:

$$\begin{bmatrix} x_i^m \\ y_i^m \\ z_i^m \end{bmatrix} = R \begin{bmatrix} x_i^c \\ y_i^c \\ z_i^c \end{bmatrix} + t, \quad (i=1,2) \quad (5)$$

Where R is a 3×3 rotation matrix with three rotation angles around the x-axis (pitch α), y-axis (yaw β) and z-axis (title χ) of the mirror coordinate system respectively; $t=[t_x, t_y, t_z]^T$ is the translation vector. So the origin $O_c=[0,0,0]^T$ of the camera coordinate system can be expressed in the world coordinate system $O_m=[t_x, t_y, t_z]^T$, so the equations of lines O_cM_1 and O_cM_2 which intersect with the upper mirror and bellow mirror respectively at points M_1 and M_2 , can be determined by solving simultaneous equations of the line O_cM_1 or O_cM_2 and the hyperboloid. Once the coordinates of the point M_1 and M_2 have been worked out, we can write out the equations of the tangent plane π_1 and π_2 which passes the upper and the bellow mirror at point M_1 and M_2 respectively. Then the symmetric points O_c^1 and O_c^2 of the origin of the camera coordinate system O_c relative to tangent plane π_1 and π_2 in the world coordinate system can be solved from the following simultaneous equations:

$$\begin{cases} \frac{x_{o_c^i} - tx}{a_i^2 x_{M_i}} = \frac{y_{o_c^i} - ty}{a_i^2 y_{M_i}} = \frac{z_{o_c^i} - tz}{-b_i^2 z_{M_i} + b_i^2 c_i} \\ a_i^2 x_{M_i} (tx + x_{o_c^i}) + a_i^2 y_{M_i} (ty + y_{o_c^i}) - (-b_i^2 z_{M_i} + b_i^2 c_i)(tz + z_{o_c^i}), \quad (i=1,2) \\ + 2[-a_i^2 x_{M_i}^2 - a_i^2 y_{M_i}^2 - z_{M_i}(-b_i^2 z_{M_i} + b_i^2 c_i)] = 0 \end{cases} \quad (6)$$

Hitherto the incident ray $O_c^1M_2$ and $O_c^2M_1$ can be written out to determine the world point $P(x_w, y_w, z_w)$. Generally, the two lines are non-co-plane due to various parameter errors and

measuring errors, we solve out the midpoint $G = (\hat{x}_w, \hat{y}_w, \hat{z}_w)^T$ of the common perpendicular of the two lines by

$$\begin{cases} [O_c^1 M_2 \times (O_c^1 M_2 \times O_c^2 M_1)] \bullet G_1 M_2 = 0 \Rightarrow \overrightarrow{OG_1} \\ G_1 M_1 = t G_1 O_c^2 \\ [O_c^2 M_1 \times (O_c^1 M_2 \times O_c^2 M_1)] \bullet G_2 M_1 = 0 \Rightarrow \overrightarrow{OG_2} \\ G_2 M_2 = t G_2 O_c^1 \end{cases} \quad \overrightarrow{OG} = (\overrightarrow{OG_1} + \overrightarrow{OG_2}) / 2 \quad (7)$$

From all of them above, we finally come to the total expression to figure out the world point $G = (\hat{x}_w, \hat{y}_w, \hat{z}_w)^T$ from two image points respectively projected by the upper mirror and bellow mirror and six camera pose parameters left to be determined.

$$G(\alpha, \beta, \chi, t_x, t_y, t_z, u_1, v_1, u_2, v_2) = \begin{bmatrix} \hat{x}_w \\ \hat{y}_w \\ \hat{z}_w \end{bmatrix} \quad (8)$$

Equation (8) is a very complex nonlinear equation with high power and six unknown parameters to determine. The artificial neural network trained with sets of image points of the calibration targets is used to estimate the camera-to-mirror transformation.

Taking advantage of the ANN capability, which adjusts the initial input camera-to-mirror transformations step by step to minimize the error function, the real transformations parameters of the camera-to-mirror can be identified precisely.

3.3 Error Function

Considering the world points with known coordinates, placed onto a calibration pattern, at the same time, their coordinates can be calculated using the equation (8) from back-projection of their image points. The difference between the positions of the real world coordinates and the calculated coordinates is the calibration error of the model. Minimizing the above error by means of an iterative algorithm such as Levenberg-Marquardt BP algorithm, the camera-to-mirror transformation is calibrated. The initial values for such algorithm are of consequence. In our system, we could assume the transformation between cameras and mirrors is quite small, as the calculation error without considering the camera-to-mirror transformation is not significant thus using $R=I$ and $T=0$ as the initial values is a reasonable choice.

We minimize the following squared error ε^2 :

$$\varepsilon^2 = \sum_{i=1}^n \|P_i - G_i(\alpha, \beta, \chi, t_x, t_y, t_z, u_1^i, v_1^i, u_2^i, v_2^i)\|^2 \quad (9)$$

Where n is the number of the calibration points.

Because $G_i(\alpha, \beta, \chi, t_x, t_y, t_z, u_1^i, v_1^i, u_2^i, v_2^i)$ depends on the camera-to-mirror transformation, (9) is optimized with respect to the six camera-to-mirror parameters.

3.4 Calibration Result

The calibration was performed using a set of 81 points equally distributed on a desk with different heights from 0 to 122mm around the vision system.

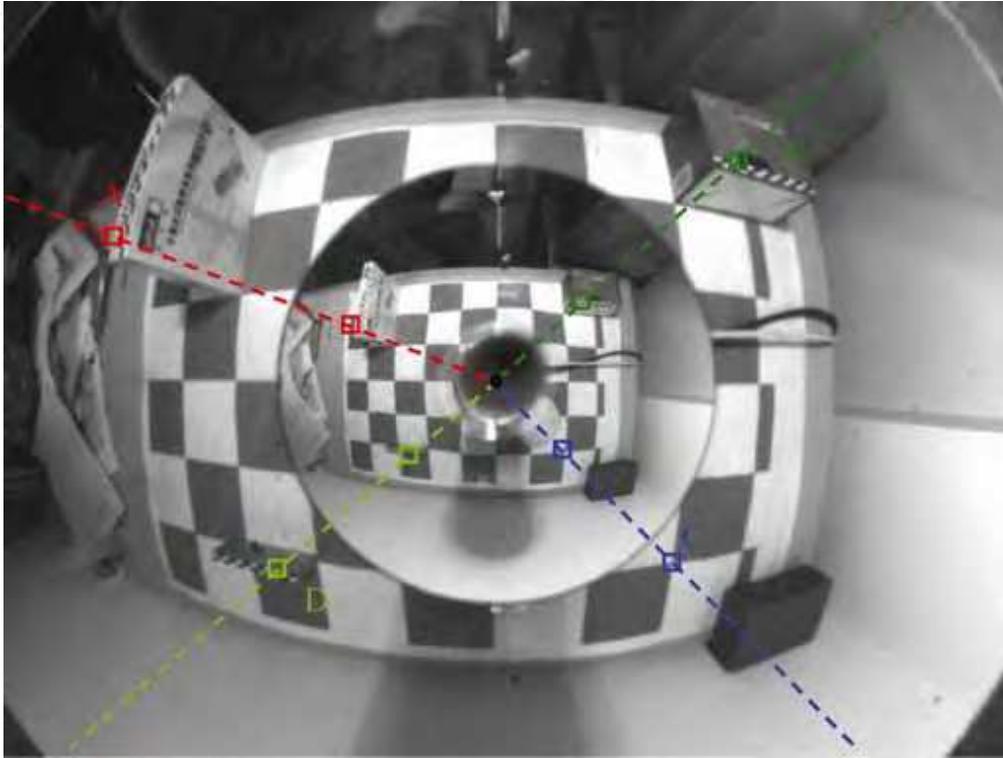


Figure 3. A calibration image used in our experiments. The coaxial configuration of the camera and the two hyperbolic mirrors makes the epipolar line radially collinear, which makes the system free of the search process for complex epipolar curve in stereo matching

The calibration results with real data are listed in Table 1.

| | α | β | χ | t_x | t_y | t_z |
|-------|----------|---------|---------|-----------|-----------|----------|
| value | -0.9539° | 0.1366° | 0.1436° | -0.0553mm | -0.1993mm | 1.8717mm |

Table 1. Calibration result with real data

The calibration error was estimated using a new set of 40 untrained points, the average square error of the set points is 34.24mm without considering the camera-to-mirror transformation. Then we calculate the error with the transformation values listed in Table 1, the average square error decrease to 12.57mm.

4. Stereo Matching

4.1 Overview

To build a depth map for mobile robot navigation, the most important and difficult process is omnidirectional stereo matching. Once two image points respectively projected by upper mirror and bellow mirror are matched, the 3D coordinate of the corresponding space point can be obtained by triangulation. State of the art algorithms for dense stereo matching can be divided into two categories:



Figure 4. Real indoor scene captured by our vision system for depth map generation

1. Local method: These algorithms calculate some kind of similarity measure over an area (Devernay & Faugeras, 1994). They work well in relatively textured areas in a very fast speed, while they cannot gain correct disparity map in textureless areas and areas with repetitive textures, which is an unavoidable problem in most situations. In (Sara, 2002) a method of finding the largest unambiguous component has been proposed, but the density of the disparity map varies greatly depend on the discriminability of the similarity measure in a given situation.
2. Global method: These methods make explicit smoothness assumptions and try to find a global optimized solution of a predefined energy function that take into account both the matching similarities and smoothness assumptions. The energy function is always in the form of $E(d) = E_{data}(d) + \lambda \bullet E_{smooth}(d)$, where λ is a parameter controlling the proportion of smoothness and image data. Most recent

algorithms belong to this category (Lee et al., 2004; Bobick, 1999; Sun & Peleg, 2004; Felzenszwalb & Huttenlocher, 2006). Among them belief propagation (Felzenszwalb & Huttenlocher, 2006) ranked high in the evaluation methodology of Middlebury College. It is based on three coupled Markov Random Fields that model smoothness, depth discontinuities and occlusions respectively and produces good result. The biggest problem of global method is that the data term and the smoothness term represent two processes competing against each other, resulting in incorrect matches in areas of weak texture and areas where prior model is violated.

Although numerous methods exist for stereo matching, they are designed towards ordinary stereo vision purpose. The images acquired by our system (Fig. 4) have some particularities in contrast to normal stereo pairs as follows, which may lead to poor result using traditional stereo matching methods:

1. The upper mirror and bellow mirror have different focal length that the camera focal length has to compromise with the two, thus causing defocusing effect, resulting in much less discriminable similarity measures. A partial solution is to reduce the aperture size at the cost of decreasing the intensity and contrast of the image.
2. Indoor scene has much more weak textured and textureless areas than outdoor scene. There are more distortions in our images, including spherical distortions and perspective distortions due to close quarters of the target areas and the large baseline.
3. The resolution gets lower when moving away from the image center. The result is the farther off the center, the more unreliable the matching result is.

To solve problem (1), we propose a three-step method that allows matching distinctive feature points first and breaks down the matching task into smaller and separate subproblems. For (2) we design a specific energy function used in the third step DTW, in which different weights and penalty items are assigned to points of different texture level and matching confidence; and throw away the matching result of the most indiscriminable points, replacing it with interpolation. For (3), we regard points farther than the most farthest matched feature point off the center as unreliable, leaving them as unknown areas. This is also required by DTW.

Epipolar geometry makes the stereo matching easier by reducing the 2D search to a 1D search along the same epipolar line in both images. To handle epipolar property conveniently, we unwrapped the raw image to two panoramic images which corresponding to images via bellow and upper mirrors respectively (Fig. 9, a, b). The matching process is done on every epipolar pair respectively. The red lines labeled in the two panoramic images are the same epipolar line for the subsequent illustration of our proposed method, of which the one above has 190 pixels and the one below 275 pixels.

4.2 Similarity Measure and Defined Texture Level

The similarity measure we choose here is zero-mean normalized cross correlation (ZNCC), since it is invariant to intensity and contrast between two images. But directly using this measure would result in low discriminability as two templates with great difference in average gray-level or standard deviation which cannot be deemed as matched pair may

have high ZNCC value. To avoid this possibility, we modified ZNCC (called MZNCC) by multiplying a window function as follows:

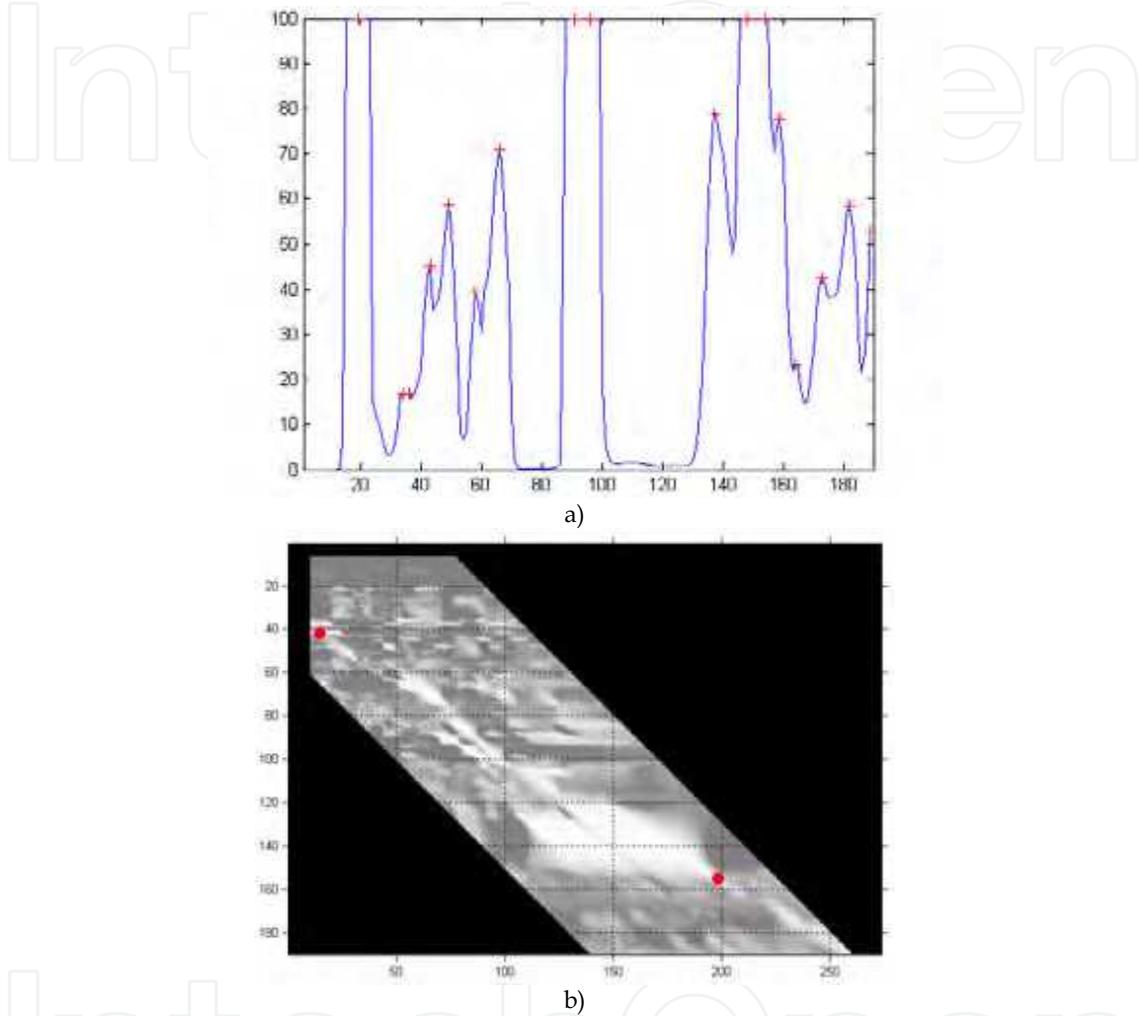
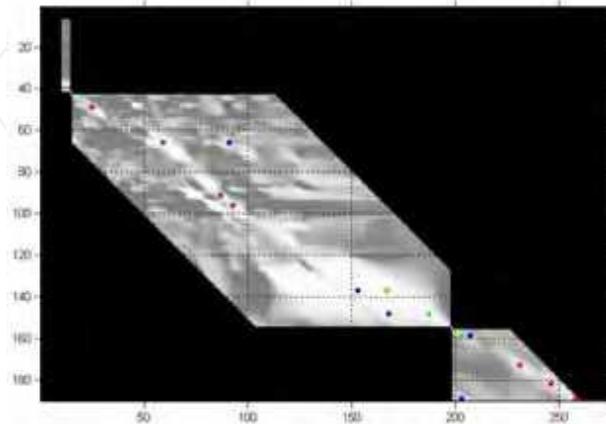


Figure 5. a) The curve of texture intensity along the epipolar line. To show it clearly, we removed the part where the curve is higher than 100 and only labeled the detected feature points. b) Detected reliable FX-dominant matching result in the MZNCC matrix. The black region of the matrix is formed since it is out of the disparity search range

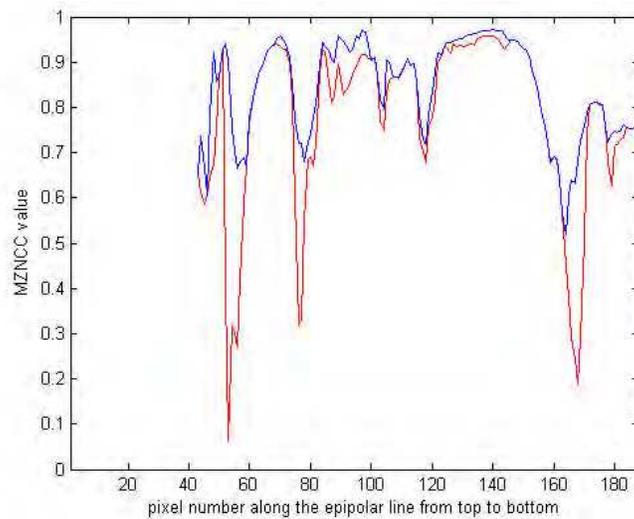
$$MZNCC(p, d) = \frac{\sum (I_a(i, j + d) - \mu_a) \cdot (I_b(i, j) - \mu_b)}{\sigma_a \cdot \sigma_b} \cdot w(|\mu_a - \mu_b|) \cdot w\left(\frac{\max(\sigma_a, \sigma_b)}{\min(\sigma_a, \sigma_b)} - 1\right) \quad (10)$$

where $w(x) = \begin{cases} 1, & x < \lambda \\ 1 - (x - \lambda), & x \geq \lambda \end{cases}$, μ_a and μ_b are the average grey-level of matching window, σ_a and σ_b are the standard deviation. For every epipolar line, all MZNCC values are stored

as a matrix (Fig. 5b) to be used in the next step. The y-axis represents the pixel number in the epipolar of Fig. 9a, while x-axis represents the number in Fig. 9b.



a)



b)

Figure 6. a) Result of feature matching. All points labeled in the graph mean candidate match for detected features, of which red and green are the result chosen by maximization of sum of MZNCC and then green are removed for being ambiguous. b) the global maximum MZNCC value for each point in this epipolar (blue) and MZNCC value along the matching route chosen by our algorithm (red)

We define our texture level of each point following the notion of bandwidth of the bandpass filter. For a given pixel and a given template centred in the pixel, we slide the template one pixel at a time in the two opposite directions along the epipolar line and stop at the location

the MZNCC value of the shifted template with the primary one decrease below a certain threshold for the first time. Let l be the distance between the two stop points, which is inverse proportional the texture level. The definition of texture intensity can be formalized as:

$$T(u,v) = \sum_{-r \leq (i,j) \leq r} (I(u+i,v+j) - \bar{I})^2 / l^2 \quad (11)$$

where r is the radius of the template. The texture intensity curve of the red labelled epipolar line is shown in Fig. 5a. With the use of this defined texture intensity and two thresholds, the whole image can be divided into three regions: strong textured, weak textured and textureless regions.

4.3 Reliable FX-dominant Matching

This step follows the notion of FX-dominant defined by Sara (Sara, 2001). The key of this notion is the uniqueness constraint which means each point may be matched with at most one point in another image, and the ordering constraint which states the order of the matched points in the two epipolar line is the same. The latter one is not always true, but it is reasonable for most cases, especially indoor scene. The FX-region of a certain matched pair (i, j) in the MZNCC matrix is defined as the set of pairs that cannot coexist with (i, j) without violating these two constraints:

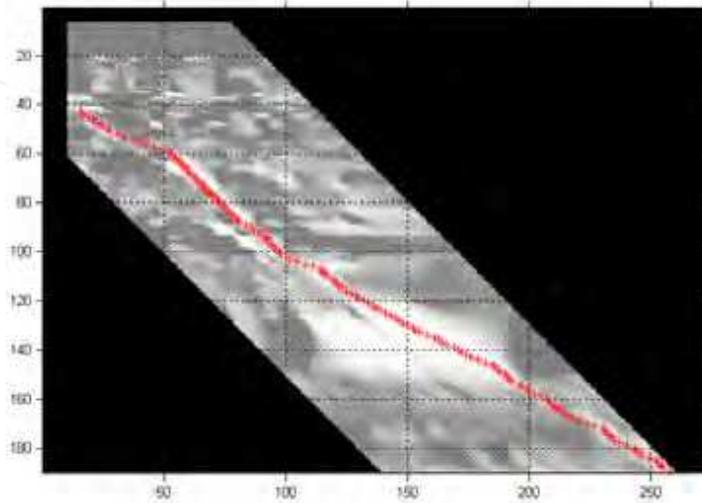
$$FX(p) = \{q = (k, l) | (k \geq i \wedge l \leq j) \vee (k \leq i \wedge l \geq j) \wedge q \neq p\} \quad (12)$$

It is formed by two opposite quadrants around (i, j) in the MZNCC matrix. And FX-dominant matching is to find pairs that have higher value than any pair in the FX-region. However, due to noise and distortion, the selected FX-dominant pairs still can not ensure its reliability. We only choose pairs from the FX-dominant results which satisfy the condition that the difference of the MZNCC value of the pair and the second local maximum MZNCC of $FX(p) \cup p$ is higher than a threshold (we choose 0.15). The number of pairs chosen by such strategy is quite small (2--8 in our case), but it does make sense because FX-region of these pairs can be removed that the matching problem is divided into subproblems. Without this step, the next step of feature matching will find much less number of reliable matched features. The result of reliable FX-dominant matching is shown in Fig. 5b, and the matrix with FX-region cut is shown in Fig. 6a.

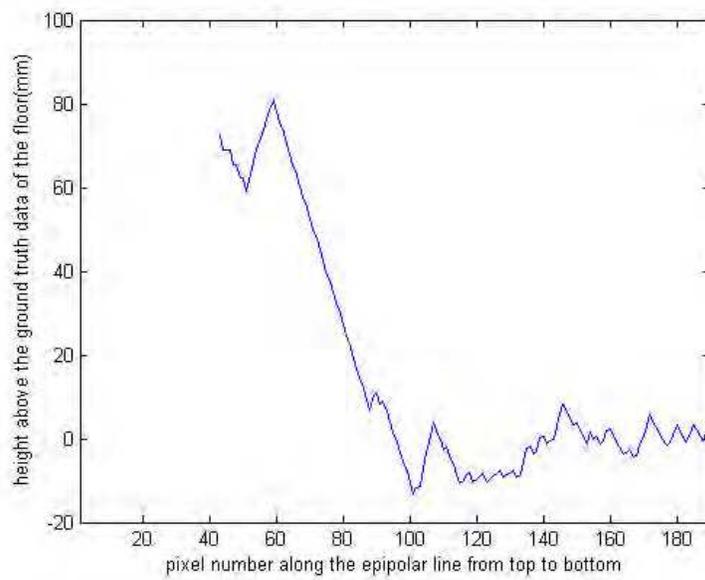
4.4 Feature Matching and Ambiguous Removal

In this step, firstly we plot the curve of the texture intensity for a given epipolar line and choose all local maximum as feature points (Fig. 5a all points labelled red cross). For every feature point every matching pair with local maximum MZNCC higher than 0.7 is labelled as a candidate match (Fig. 6a all labelled points). Then we select a combination of candidate matching pairs that obeys uniqueness constraint and ordering constraint and has the highest sum of MZNCC (A feature point can be left unmatched with a zero contribution to the sum of MZNCC). The selected combination of illustrating epipolar shown in Fig. 6a is labelled red and green. In this selected combination, still some ambiguous match candidates exist. We mean a selected candidate is unambiguous if it is the only choice without altering other matched feature points under uniqueness and ordering constraint, otherwise it is

ambiguous. We will then remove all ambiguous feature points until no matched feature is ambiguous. In Fig. 6a, the ambiguous match candidates are labelled green and they are to be removed from the feature matching result.



a)



b)

Figure 7. a) Matching route in the MZNCC matrix via DTW. b) computed depth curve for this epipolar line

4.5 Dense Matching via DTW

The remaining correspondences can be determined by dynamic time warping (DTW) (Lee et al., 2004). A starting and an ending point should be known at first to use DTW. The matched feature points in the last step can naturally perform this role. Therefore, DTW can be applied to every range between adjacent matched feature points. This objective of DTW is achieved by finding a path with optimized energy function in a search space defined by the search range and restricted by the starting and ending point as well as the uniqueness and ordering constraint, using dynamic programming technique. The most important part is the definition of the energy function. Unlike others straightforwardly use sum of intensity difference (Lee et al., 2004) or define the energy function with smoothness item (this can hardly be implemented in our case as the assumption that flat areas correspond to constant disparity route usually does not hold due to the large baseline), we define our energy function in the form of sum of MZNCC value plus a penalty item aim to assign different weights to different points based on the texture level and matching confidence:

$$E = \sum MZNCC(i, j) + \sum penalty(i, j) \quad (13)$$

where (i, j) is in the matching route. To define the penalty item, we make another classification of all points. A point is belong to Class A (high confidence) if the global maximum MZNCC value is higher than 0.7, Class B (low confidence) if the global maximum MZNCC is between 0.5 and 0.7, otherwise Class C (noise). Then the penalty item is defined as Table 2,

| | strong textured | weak textured | textureless |
|---|--|--|-------------|
| A | $-\lambda \cdot \mu \cdot \max((0.7 - MZNCC), 0)$ | $-\mu \cdot \max((0.7 - MZNCC), 0)$ | 0 |
| B | $-\sigma \cdot \lambda \cdot \mu \cdot \max((0.5 - MZNCC), 0)$ | $-\sigma \cdot \mu \cdot \max((0.5 - MZNCC), 0)$ | $-MZNCC$ |
| C | $-MZNCC$ | $-MZNCC$ | $-MZNCC$ |

Table 2. The penalty item

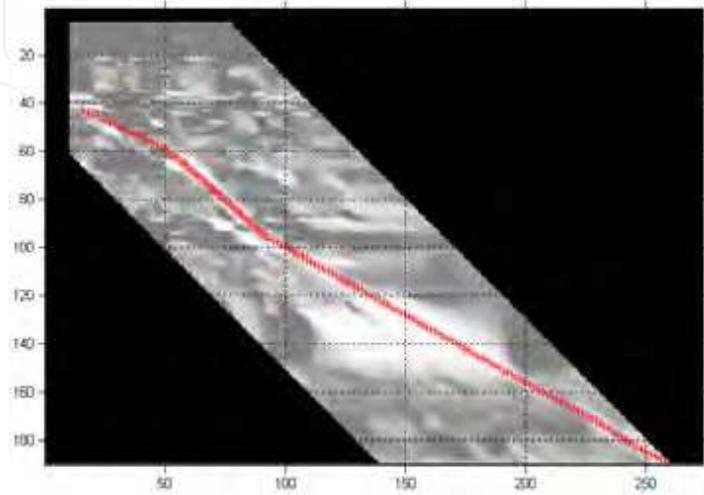
where λ is the strong texture weight, μ is penalty level and σ low confidence weight (in our case, $\lambda=4$, $\mu=4$, $\sigma=0.4$).

The result of DTW performed in the red labelled epipolar is shown in Fig. 7a, the computed depth curve in Fig. 7b. Fig. 6b shows the MZNCC curve along the matching route and the global maximum MZNCC curve for the epipolar. From Fig. 5a and Fig. 6b, we can see that the result route only deflects the global maximum MZNCC curve in textureless points and points belonging to low confidence or noise, which is an ideal result.

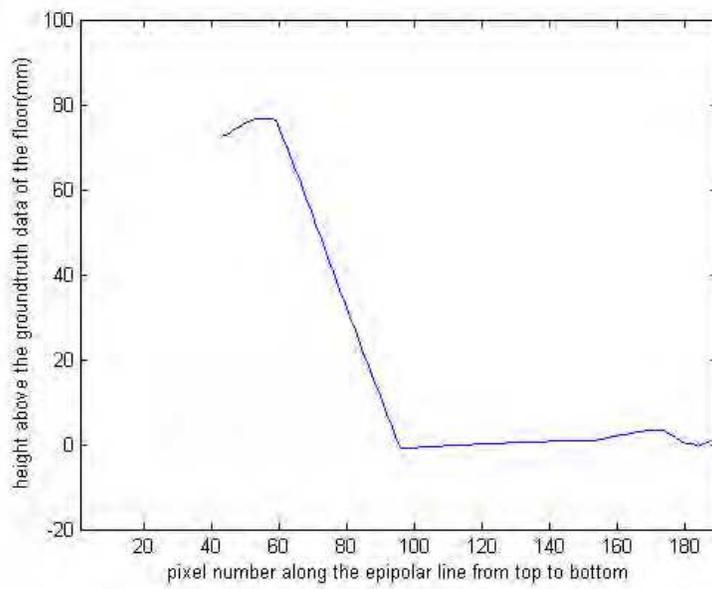
4.6 Postprocessing

A postprocessing step replacing textureless match with interpolation is applied to get smooth surfaces. As in the textureless areas, the similarity value is very ambiguous that the matching route can vary greatly with very small energy variation. The result is that the maximization of energy function does not necessarily correspond to the correct match, causing jagged depth map (Fig. 7b) Easily observed, the textureless areas almost correspond to a plane (as the threshold of textureless area is set so low that uneven areas will be

categorized as weak textured areas due to slightly different illumination), we use two nearest textured (strong or weak) match to interpolate the textureless point. After that, a medium filter is applied to ensure the smoothness of depth map. Fig. 8 shows the result of matching route and the depth curve after postprocessing.



a)



b)

Figure 8. a) Final matching route in the MZNCC matrix after post-processing. b) computed depth curve for this epipolar line

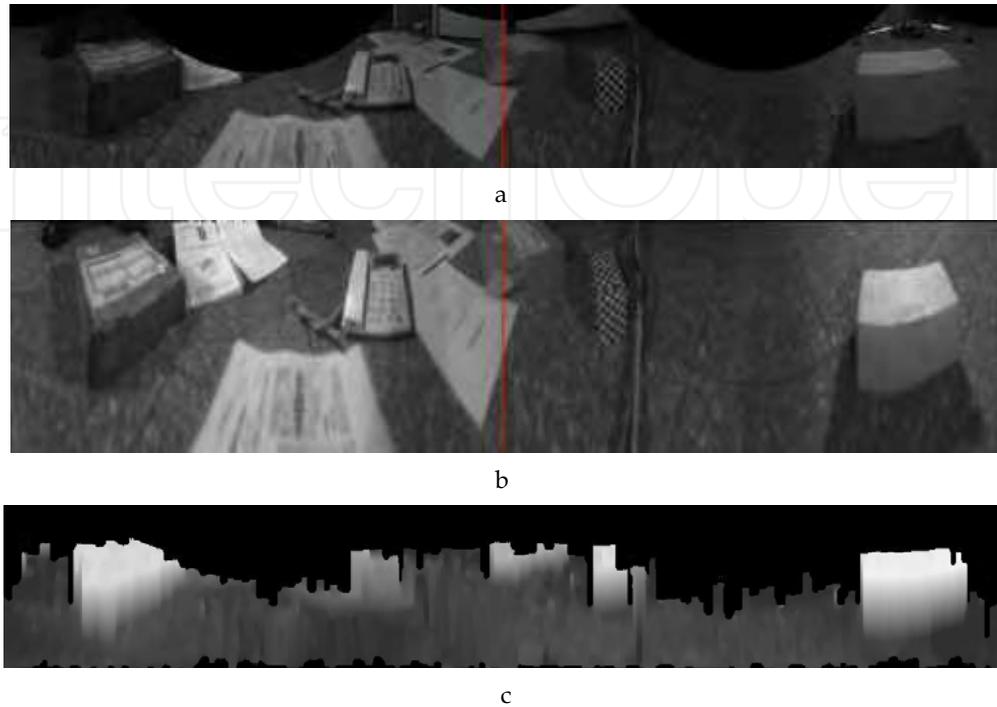


Figure 9. Panoramic images unwrapped from the raw images of Fig.4 (a and b are converted by images via the lower and upper mirrors respectively) and the detected depth map (c) corresponding to a

4.7 Result

Fig. 9c shows the result of the generated depth map via the proposed method. This depth map measures the height above the floor. The brightness of the map is proportional to the height, while black represents unknown areas. Although the ground truth map is unavailable, the real height can be surveyed accurately for most points. We randomly selected hundreds of points and checked the error, finding that most are smaller than 15mm, only slightly higher than the calibration error. We set the threshold of navigable areas as 25mm above the floor and get the navigable map in Fig. 10.

5. Summary

We have developed a complete framework of automatically generating omnidirectional depth maps around a mobile robot using a novel designed panoramic vision sensor. Compared to previous vision systems, our system has such significant advantages as its geometry calculating easy and fast and simultaneous acquisition of precise range information without high cost or system complexity. And as the separation between the two hyperbolic mirrors provides a large baseline, the range information obtained from this method has much improved precision. We have proposed an imaging model for

omnidirectional cameras that accounts for the full rotation and translation between the camera and mirrors, and a LMBP method for recovering the relative position from back-projection the images points. The method is general in that any combination of camera and mirrors can be calibrated, including non-single-viewpoint combinations. For SVP cameras, where the merit of a single viewpoint can be exploited only if the camera and mirrors are assumed to be perfectly aligned, this algorithm can be used to verify the alignment accuracy. We also have presented a three-step method for stereo matching of our vision system, which combines the advantage of feature matching and global matching. This method basically solved the three major difficulties faced by our vision system. The experimental result is quite convincing. Although this method uses some thresholds and parameters, the matching result does change in case of small variation of parameters.

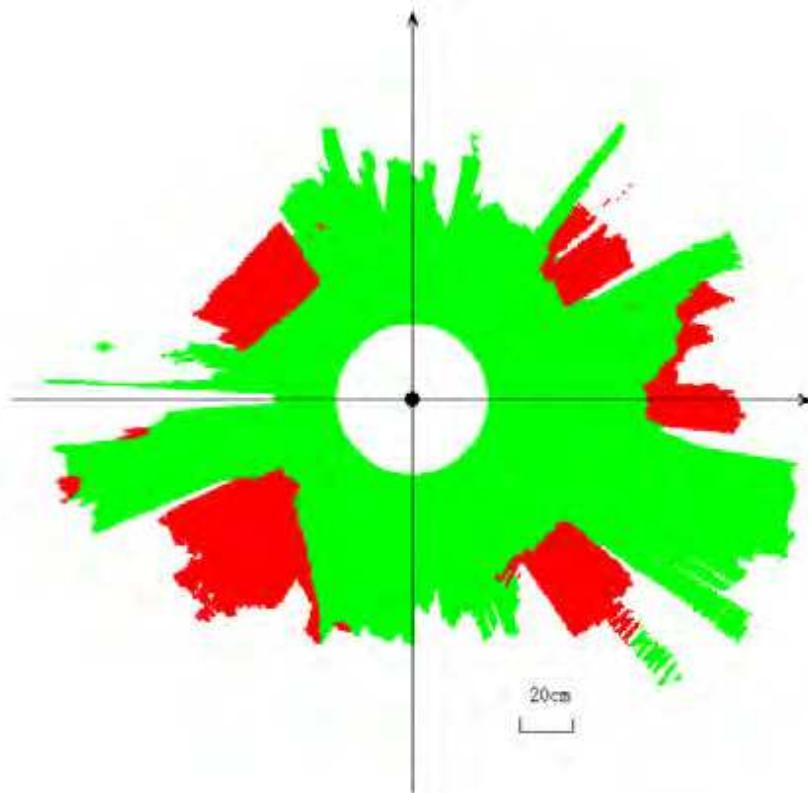


Figure 10. Omnidirectional scene information obtained by our system, of which green represents navigable areas, red detected obstacle areas

However, the proportion of detected areas is a bit small in a few epipolar lines. To solve this problem, one way is to impose inter-epipolar consistency in the stereo matching method. Another is to use multibaseline stereo, that is, we can first estimate relative pose between different shooting positions, and then depth map can be generated more reliably with more

virtual cameras. Our future work will focus on these aspects. Also, we found that some techniques to deal with the defocusing effect have been proposed to improve the image quality. We will also investigate the possibility to get a better method with some pre-processing techniques.

6. References

- Baker, S. & Nayar, S.K. (1999). A Theory of Single-Viewpoint Catadioptric Image Formation, *International Journal of Computer Vision*, Vol. 35, No. 2, pp. 1 - 22
- Bobick, A. (1999). Large Occlusion Stereo, *International Journal of Computer Vision*, Vol. 33, No. 3, pp. 181-200
- Bunschoten, R. & Kröse, B. (2002). Robust Scene Reconstruction from an Omnidirectional Vision System, *IEEE Transaction on Robotics and Automation*, Vol. 19, No. 2, pp. 351-357
- Cabral, E.L.L.; de Souza Jr., J.C. & Hunold, M.C. (2004). Omnidirectional Stereo Vision with a Hiperbolic Double Lobed Mirror, *Proceedings of International Conference on Pattern Recognition*, pp. 1-4, Vol. 1, Cambridge, England, Aug 2004
- Conroy, T.L. & Moore, J.B. (1999). Resolution Invariant Surfaces for Panoramic Vision Systems, *Proceedings of International Conference on Computer Vision*, pp. 392-397, Kerkyra, Greece, Sep 1999
- Devernay, F. & Faugeras, O. (1994). Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models, *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 208-213, Seattle, Washington, Jun 1994
- Fabrizio, J.; Tarel, J. & Benosman, R. (2002). Calibration of Panoramic Catadioptric Sensors Made Easier, *Proceedings of Workshop on Omnidirectional Vision*, pp. 45-52, Copenhagen, Denmark, Jun 2002
- Felzenszwalb, P.F. & Huttenlocher, D.P. (2006). Efficient Belief Propagation for Early Vision, *International Journal of Computer Vision*, Vol. 70, No. 1, pp. 41-54
- Geyer, C. & Daniilidis, K. (1999). Catadioptric Camera Calibration, *Proceedings of International Conference on Computer Vision*, Vol. 1, pp. 398-404, Kerkyra, Greece, Sep 1999
- Geyer, C. & Daniilidis, K. (2000). Paracatadioptric Camera Calibration, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 687-695
- Geyer, C. & Daniilidis, K. (2003). Conformal Rectification of Omnidirectional Stereo Pairs, *Proceedings of IEEE Workshop on Omnidirectional Vision and Camera Networks*, pp. St. Louis, USA, Jun 2003
- Gluckman, J.; Nayar, S.K. & Thoresz, K.J. (1998). Real-time Omnidirectional and Panoramic Stereo, *Proceedings of the DARPA Image Understanding Workshop*, pp. 299-303, Monterey, California, Nov 1998
- Kang, S.B. (2000). Catadioptric Self-calibration, *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 201-207, Hilton Head Island, South Carolina, Jun 2000
- Koyasu, H.; Miura, J. & Shirai, Y. (2002). Recognizing Moving Obstacles for Robot Navigation Using Real-time Omnidirectional Stereo Vision, *Journal of Robotics and Mechatronics*, Vol. 14, No. 2, pp. 147-156
- Koyasu, H.; Miura, J. & Shirai, Y. (2003). Mobile Robot Navigation in Dynamic Environments Using Omnidirectional Stereo, *Proceedings of IEEE Int. Conf. on Robotics and Automation*, pp. 893-898, Taipei, May 2003

- Lee, Y.; Kim, D. & Chung, M. (2004). Feature Matching in Omnidirectional Images with a Large Sensor Motion for Map Generation of a Mobile Robot, *Pattern Recognition Letters*, Vol. 25, No. 4, pp. 413-427
- Luong, Q.T. & Faugeras, O.D. (1996). The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *Int. J. of Computer Vision*, Vol. 17, No. 1, pp. 43-76
- Ma, J. (2003). Omnidirectional Vision Based Localization and Path Planning for Mobile Robots, *Doctor Thesis of Beijing Institute of Technology*, 2003
- Menegatti, E.; Maeda, T. & Ishiguro, H. (2004). Image-Based Memory for Robot Navigation Using Properties of Omnidirectional Images, *Robotics and Autonomous Systems*, 47 (4), pp. 251-267
- Miousik, B. & Pajdla, T. (2003a). Estimation of Omnidirectional Camera Model from Epipolar Geometry, *Proceedings of 2003 IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 485-490, Madison, USA, Jun 2003
- Miousik, B. & Pajdla, T. (2003b). Omnidirectional Camera Model and Epipolar Geometry Estimation by RANSAC with Bucketing, *Proceedings of Scandinavian Conf. on Image Analysis*, pp. 83-90, Goteborg, Sweden, Jun 2003
- Nayar, S.K. (1997). Catadioptric Omnidirectional Camera, *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 482-488, San Juan, Puerto Rico, Jun 1997
- Sara, R. (2001). The Class of Stable Matchings for Computational Stereo, *Research Report, CTU-CMP-1999-22*, Center for Machine Perception, Czech Technical University, 2001
- Sara, R. (2002). Finding the Largest Unambiguous Component of Stereo Matching, *Proceedings of European Conf. on Computer Vision*, Vol. 2, pp. 900-914, Copenhagen, Denmark, May 2002
- Southwell, M.F.D.; Basu, A. & Reyda, J. (1996). Panoramic Stereo, *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 378-382, Vienna, Austria, Aug 1996
- Su, L. & Zhu, F. (2005). Design of a Novel Stereo Vision Navigation System for Mobile Robots, *Proceedings of IEEE Conference on Robotics and Biomimetics*, No. 1, pp. 611-614, Hong Kong, Jun 2005
- Sun, C. & Peleg, S. (2004). Fast Panormaic Stereo Matching Using Cylindrical Maximum Surfaces, *IEEE Trans. on Systems, Man and Cybernetics, Part B*, Vol. 34 No. 1, pp. 760-765
- Yagi, Y. (2002). Real-time Omnidirectional Image Sensor for Mobile Robot Navigation, *Proceedings of IEEE International Symposium on Intelligent Control*, pp. 702-708, Vancouver, Canada, Oct 2002
- Ying, X. & Hu, Z. (2004). Catadioptric Camera Calibration Using Geometric Invariants, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 10, pp. 1260-1271
- Zhang, Z. & Faugeras, O. (1997). An Effective Technique for Calibrating a Bincular Stereo Through Projective Reconstruction Using Both a Calibration Object and the Environment, *Journal of Computer Vision Research*, Vol. 1, No. 1, pp. 58-68
- Zhu, Z. (2001). Omnidirectional Stereo Vision, *Proceedings of Workshop on Omnidirectional Vision in the 10th IEEE ICAR*, pp. 1-12, Budapest, Hungary, Aug 2001



Scene Reconstruction Pose Estimation and Tracking

Edited by Rustam Stolkin

ISBN 978-3-902613-06-6

Hard cover, 530 pages

Publisher I-Tech Education and Publishing

Published online 01, June, 2007

Published in print edition June, 2007

This book reports recent advances in the use of pattern recognition techniques for computer and robot vision. The sciences of pattern recognition and computational vision have been inextricably intertwined since their early days, some four decades ago with the emergence of fast digital computing. All computer vision techniques could be regarded as a form of pattern recognition, in the broadest sense of the term. Conversely, if one looks through the contents of a typical international pattern recognition conference proceedings, it appears that the large majority (perhaps 70-80%) of all pattern recognition papers are concerned with the analysis of images. In particular, these sciences overlap in areas of low level vision such as segmentation, edge detection and other kinds of feature extraction and region identification, which are the focus of this book.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Chuanjiang Luo, Liancheng Su and Feng Zhu (2007). A Novel Omnidirectional Stereo Vision System via a Single Camera, Scene Reconstruction Pose Estimation and Tracking, Rustam Stolkin (Ed.), ISBN: 978-3-902613-06-6, InTech, Available from:

http://www.intechopen.com/books/scene_reconstruction_pose_estimation_and_tracking/a_novel_omnidirectional_stereo_vision_system_via_a_single_camera

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2007 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen