Omnidirectional Stereo Vision[*]

Zhigang Zhu
University of Massachusetts, Amherst, MA 01003, USA
Email: zhu@cs.umass.edu

Abstract

This paper discusses several interesting configurations of omnidirectional stereo (omnistereo) - binocular omnistereo, N-ocular omnistereo, circular projection omnistereo and dynamic omni-stereo. An omnidirectional image can be either obtained by an omnidirectional camera or generated by image mosaicing. Usually an omnidirectional image has a 360-degree view around a viewpoint, and in its most common form, can be presented in a cylindrical surface around the viewpoint. This paper shows that an omnidirectional image for stereo vision can either have a single viewpoint or multiple viewpoints, and can be either viewer-centered or object-centered. With these generalizations, omnidirectional stereo vision can be extended from a viewer-centered binocular/ N-ocular omnistereo with a few fixed viewpoints to more interesting omnistereo configurations – circular projection omnistereo with many viewpoints in a small region, dynamic omnistereo with a few reconfigurable viewpoints in a large region, and object-centered omnistereo with many viewpoints distributed in a large region. Important issues on omnidirectional stereo imaging, epipolar geometry, and depth accuracy are discussed and compared.

1. Introduction

Usually an omnidirectional image has a 360-degree view around a viewpoint, and in its most common form, can be presented in a cylindrical surface around the viewpoint. There are at least two ways to obtain omnidirectional images. The first approach is to use an omnidirectional camera. Different kinds of omni-directional (or panoramic) imaging sensors have been designed [1-6], and a systematic geometric analysis of omni-directional sensors has been given by Baker & Nayar [1]. Real-time omnidirectional vision has become quite popular with many vision approaches for robot navigation, 3D reconstruction and video surveillance. In this class, a single effective viewpoint is desired, so that the annular projection on the sensing target can be easily re-projected to a cylinder or a plane. Omnistereo approach has been proposed with two vertically-aligned omnidirectional cameras [7,8], where epipolar lines are constrained to vertical lines, and 3D depth of the entire 360-degree FOV can be obtained. Sogo et al [9] proposed a "N-ocular stereo" approach to compensate observation errors using redundant information from multiple omnidirectional sensors. Panoramic virtual stereo approach [10, 27] has been proposed to form the optimal triangulation by two omnidirectional cameras mounted on two mobile robots to track and localize moving objects.

The second approach is to generate omnidirectional images is by image mosaicing techniques. Plenoptic modeling approach [11] was proposed to use two cylindrical panoramas from rotating cameras in two viewpoints to estimate the disparity map. However, the emphasis of this work is the rendering from this representation rather than the modeling of this representation from an image sequence. Kang and Szeliski [12] proposed a multibaseline omnidirectional stereo approach to recover the 3D scene data. Shum et al. [13] made an effort to construct 3D models from two cylindrical panoramic mosaics from rotating cameras interactively, using environmental constraints such as parallel lines, lines with known orientations, etc. Creating stereo mosaics from two rotating cameras was proposed by Huang & Hung [14]. More interesting approaches are the generation of stereo mosaics with *circular projection* from a single off-center rotating camera proposed by Peleg & Ben-Ezra [15,16] and Shum & Szeliski [17,18]. Each mosaic thus generated has multiple viewpoints on a viewing circle, and a pair of concentric mosaics has better stereo geometry than a pair of omnidirectional images with two separate viewpoints. In fact, the idea of generating stereo panoramas for either an off-center rotating camera or a translating camera can be traced back to the earlier work in robot vision applications by Ishiguro, et al [19] and Zheng & Tsuji [20]. The attraction of the recent studies on off-center rotating cameras lies in how to make stereo mosaics with nicer epipolar geometry and higher image qualities, and how to use them in image-based rendering.

In stereo mosaics with an off-center rotating camera, the viewpoints -- therefore the parallax -- are limited to images taken from a very small area, and the viewers are constrained to rotationally viewing the stereo representations. If we extend the concept of omnidirectional vision from viewer-centered imaging to object-centered imaging, thus allowing the viewpoints distributing on a large region, omnidirectional vision and omnidirectional stereo vision can cover a much wider range of panoramic stereo imaging with much more general motion of the camera. Two of the interesting cases are the stereo mosaics with a camera turning around a large object and an airborne camera flying around the earth. In both cases, the motion of the camera can be well approximated by a dominant translation. Note that a rotating camera can be easily controlled to achieve the desired motion, but the translation of a camera over a large distance is much hard to control. Nevertheless, translational motion is the typical prevalent sensor motion during ground vehicle navigation [20-22] or aerial surveys [23-26]. In this paper a generalized omnistereo vision is proposed by allowing different kinds of omnidirectional imaging configurations.

2. Binocular and N-Ocular Omnistereo

In this section, we discuss several configurations of *omnistereo with only a few viewpoints*. In this context, an omnidirectional image is a 360-degree cylindrical projection from a single viewpoint where the projection in the vertical direction is perspective. The projection (ϕ, v) in such an image of a 3D point (X, Y, Z) can be represented as

$$(\phi, v) = (\tan^{-1}(Y, X), F \frac{Z}{D}) \qquad (1)$$

where $0° \leq \phi < 360°$ is the direction of rays in the 360-degree horizontal field of view (FOV), $D = \sqrt{X^2 + Y^2}$ is the perpendicular distance from the point to the image cylinder's vertical axis passing through the focal point, and F is the focal length of the "virtual" cylindrical omni camera (refer to Fig. 1a).
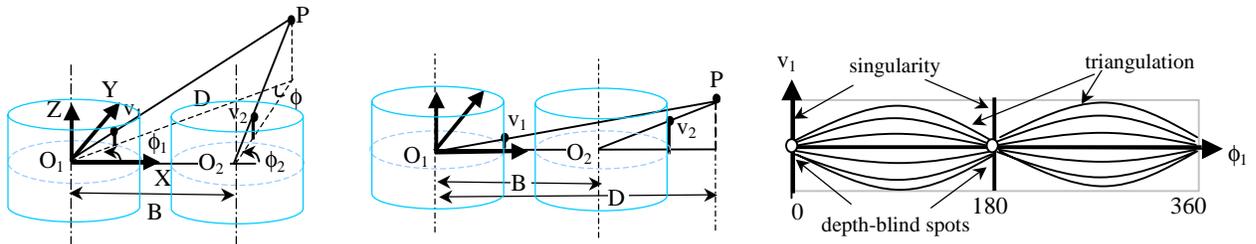

Fig. 1. Horizontally-aligned omnistereo (a). triangulation (b) singularity (2) epipolar curves

2.1. Horizontally-aligned omnistereo

To start with, we analyze the simplest case where two omnidirectional cameras are horizontally aligned and have a fixed baseline B (Fig. 1a). Given the corresponding points in the stereo pair, (ϕ_1, v_1) and (ϕ_2, v_2) of a 3D point P, the distance ("depth") of the point to the first camera can be calculated by *triangulation*:

$$D = B \frac{\sin \phi_2}{\sin(\phi_2 - \phi_1)} = B \frac{\sin \phi_2}{\sin \phi} \qquad (2)$$

where $\phi = \phi_2 - \phi_1$ is the horizontal disparity (and also the vergent angle between the two rays). Different from planar perspective stereo, this triangulation equation is invalid in the *singularity* cases where the point is aligned with the baseline, i.e. $\phi_1 = \phi_2 = 0°$ or $180°$. More generally, the depth error can be estimated as

$$\partial D = \frac{D \sqrt{D^2 - B^2 \sin^2 \phi_2}}{B \sin \phi_2} \partial \phi \approx \frac{D^2}{B \sin \phi_2} \partial \phi \ , \ D \gg B \qquad (3)$$

where $\partial \phi$ is the error in estimating disparity (for simplicity, we assume that there is no error in the angle ϕ_2, which implies that we find the corresponding point (ϕ_1, v_1) of a given point (ϕ_2, v_2)). From Eq. (3) we have the following conclusions for the error characteristics of the horizontally aligned binocular omnistereo (H-binocular omnistereo).

Conclusion 1.1. The depth accuracy of the H-binocular omnistereo is non-isotropic. For a given baseline and a given distance of points, the best distance estimation is achieved when the vergent angle between two rays from the two cameras to the point is maximum (where $\phi_2 = 90°$ or $270°$), and maximum error (∞) when the vergent angle is zero (where $\phi_1 = \phi_2 = 0°$ or $180°$).

Conclusion 1.2. The depth accuracy of the H-binocular omnistereo is proportional to the square of the depth, and inversely proportional to the baseline length.

Conclusion 1.2 shows the same depth error characteristic as in the traditional perspective stereo, but unfortunately conclusion 1.1 shows decreasing accuracy of H-binocular omnistereo when the FOV of the stereo is extended to 0° and 180°. It is true that we cannot estimate any distance when $\phi_1 = \phi_2 = 0^\circ$ or 180° (singularity cases) and the point is on the same height as the optical centers of the two cameras (i.e., $v_1 = v_2 = 0$). However, it is possible to estimate the depth in the singularity cases otherwise [10, 27]. For example, when $\phi_1 = \phi_2 = 0^\circ$ we have (Fig. 1b)

$$D = B \frac{v_2}{v_2 - v_1} = B \frac{v_2}{v} \tag{4}$$

where $v = v_2 - v_1$ is the vertical disparity. In this sense, theoretically, there is only two "depth blind" points in the omnidirectional image pair. However, in practice, the two cameras will occlude each other in the direction of the baseline, so no information is available in the *mutual occluding* regions.

Another drawback of the H-binocular omnistereo is its epipolar geometry. Given point (ϕ_2, v_2) in the 2nd image, its corresponding point (ϕ_1, v_1) in the 1st image lies in a sine curve (Fig. 1c)

$$v_1 = \frac{\sin \phi_1}{\sin \phi_2} v_2 \tag{5}$$

This equation holds iff $\sin \phi_2 \neq 0$. In the singularity cases, the epipolar lines are along the v directions (Fig. 2):

Conclusion 1.3. The epipolar curves are sine curves in the non-singularity cases and the epipolar lines are along the v directions in the singularity cases.

Previously work in H-binocular omnistereo includes plenoptic modeling[11], omnidirectional multibaseline stereo[12], interactive 3D reconstruction from panoramic mosaics [13] and panoramic virtual stereo [10]. The panoramic virtual stereo approach will be further discussed in Section 4 (dynamic omnistereo).
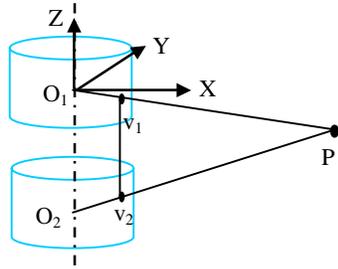

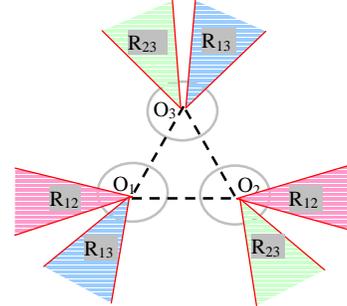Fig. 2. Vertically-aligned omnistereo

Fig. 3. Trinocular omnistereo

2.2. Vertically-aligned omnistereo

Some of the drawbacks in H-binocular omnistereo can be overcome by a vertically-aligned binocular (V-binocular) omnistereo configuration (Fig. 2). The depth equation of the V-binocular stereo is simply the same as a traditional perspective stereo

$$D = F \frac{B_v}{v_2 - v_1} = F \frac{B_v}{v} \tag{6}$$

where B_v is the vertical baseline length, and v is the vertical disparity. The similar depth accuracy properties as in conclusion 1.2 hold for the V-binocular omnistereo, but we have nicer properties here:

Conclusion 2. The depth accuracy of the V-binocular omnistereo is isotropic in all directions, and the epipolar lines are simply vertical lines in omnidirectional image.

In addition, there are no mutual occlusions by the sensors in both images. This configuration cannot be used for stereo viewing by human eyes that horizontally aligned. Previously work in V-binocular omnistereo includes the approaches using two optical omnidirectional sensors [7 ,8] and using a pair of 1D scanning cameras [29]. In the later approach, high-resolution stereo pair is captured by a pair of vertically aligned 1D scan cameras. Because of the simple epipolar geometry, the depth map could be recovered during rotation in the scanning approach. Although the

scanning approach cannot be applied in a dynamic scene, it is a simple and practical solution for modeling static scenes in high resolution.

2.3. N-ocular omnistereo

Naturally, depth accuracy in the H-binocular omnistereo can be improved by using more than two omnidirectional cameras. This is the general case of *omnistereo with a few viewpoints*. For example, Sogo et al [9] proposed a "N-ocular stereo" approach that verifies the correspondences of multiple targets of binocular stereo by a third omnidirectional camera, without visual features. In addition, they have developed simple methods to compensate observation errors using redundant information from multiple sensors. They have shown that the uncertainty in estimating 3D locations is reduced by using the best estimations of pairs of four fixed panoramic cameras put in the vertices of a square region, and. In fact, as shown in Fig. 3, this goal can be achieve with three omnidirectional cameras. Further, every point of the 360 FOV from the center of the sensor-triangle can be covered by at least two pairs of rays from different cameras with good triangulations, so that one pair of stereo match can be verified using the second pair. In Fig. 3, the shaded region R_{ij} indicates the region that does not have good triangulations by camera O_i and O_j. For example, regions R_{12} do not have good triangulations by camera O_1 and O_2; however, good triangulations in the regions can be formed by both camera pairs O_1-O_3 and O_2-O_3. In addition, there is less mutual occlusion problem by using the redundant information from the three (or four) cameras. This approach gives better 3D estimation than the H-binocular omnistereo at the expense of more cameras. Although the depth accuracy is still not isotropic, it is more uniform in directions. Note that the error of localizing a target is still proportional to the square of the target's distance from the cameras of fixed baseline distances. The next two sections describe two approaches that overcome these two drawbacks respectively.

3. Circular Projection Omnistereo

A further extension to the N-ocular omnistereo is the omnivergent stereo with two nice properties[18]: (1) every point in the scene is imaged from two cameras that are vergent on that point with maximum vergence angle; and (2) stereo recovery yields isotropic depth resolution in all directions. One practical configuration is the circular projection omnistereo [15-18] where the a virtual camera moving in a viewing circle captures two set of rays on a plane tangent to the viewing circle: the left-eye omnidirectional image uses the rays in the clockwise direction of the circle, and the right-eye omnidirectional image uses the rays in the counterclockwise direction (Fig. 4). In each of such an omnidirectional image $I(\phi, v)$, the horizontal coordinates are the ray directions, and the vertical coordinates are the 1D perspective projections.
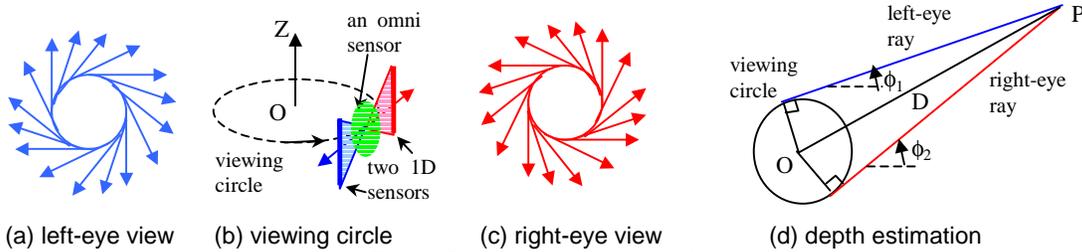

(a) left-eye view   (b) viewing circle   (c) right-eye view   (d) depth estimation

Fig. 4. Circular projection omnistereo

Given a point (ϕ_2, v_2) in the 2nd image, its corresponding point (ϕ_1, v_1) in the 1st image lies in the horizontal line that contains the point (ϕ_2, v_2), i.e., we always have $v_1 = v_2$. The reason for the horizontal epipolar lines in the circular projection omnistereo is that both the viewing planes that contain point (ϕ_1, v_1) and (ϕ_2, v_2) are tangent to the viewing circle [18]. Assuming that the radius of the viewing circle is r, then the distance of the scene point from the center of the viewing circle is

$$D = r / \sin \frac{\phi}{2} \tag{7}$$

where $\phi = \phi_2 - \phi_1$ is the horizontal disparity (vergent angle). The depth error can be estimated as

$$\partial D = \frac{D\sqrt{D^2 - r^2}}{2r} \partial \phi \approx \frac{D^2}{2r} \partial \phi, \ D \gg r \tag{8}$$

where $\partial\phi$ is the error in disparity, *2r* is the "baseline" length between the left and right viewpoints of the corresponding rays. Circular projection omnistereo has the following properties.

Conclusion 3.1. The depth estimation in the circular projection omnistereo is independent of directions and hence is isotropic, since every point at the same distance has the same (maximum) vergent angle between rays from the left- and right- eye views.

Conclusion 3.2. The depth accuracy of the circular projection omnistereo is proportional to the square of the depth, and is inversely proportional to the baseline length (diameter of the viewing circle).

Conclusion 3.3. Circular projection omnistereo has horizontal epipolar lines.

Compared with the H-binocular omnistereo, circular projection omnistereo has the same depth error properties in terms of distance and baseline (conclusion 3.2). However, circular projection omnistereo has isotropic depth estimation and horizontal epipolar lines instead of the non-isotropic depth estimation and sine epipolar curves.

In the basic configuration of the circular projection omnistereo, if we have a camera that can collect rays of all directions from each viewpoint in the corresponding tangent plane of the viewing circle, we will obtain an omnistereo image pair in a full coverage of 360x360 FOV of the scene. This could be realized by using an omnidirectional camera with such imaging property. For example, we can use an omnidirectional sensor such as the panoramic annular lens (PAL) camera [3] that can view the circle whose center is the effective viewpoint of the sensor (Fig. 4b). Possible real-time full optical solutions (without any camera rotation) are also proposed in [16]. Here we compare two practical approaches using a conventional video camera.
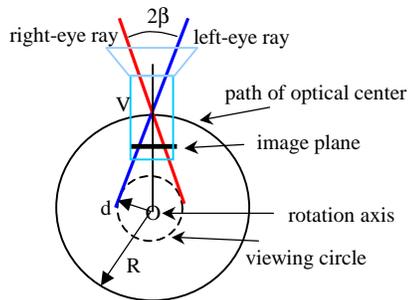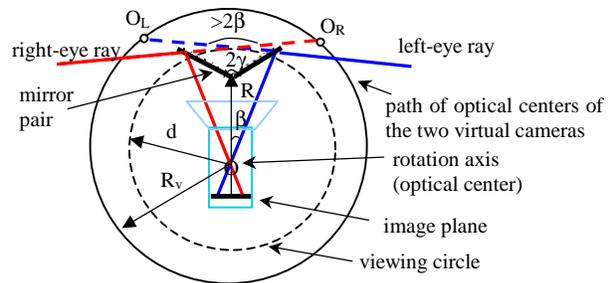

Fig. 5. A single-camera approach   Fig. 6. A simple catadioptric approach

3.1. A single camera approach

The first simple approach is using an off-center rotating camera, looking outward [15, 18]). When the optical center (V) of the camera is moving on a circle around a rotation axis O, two column of angular distance β from the center of the video frame are taken, one contributing to the left-eye mosaic and the other to the right-eye mosaic. The omnistereo images thus generated has a smaller viewing circle than the circular path of the optical center (Fig. 5). Given the radius R of the camera circular path and the vertical slit separation 2β, the radius of the viewing circle is

$$r = R\sin\beta \qquad (9)$$

For example, if we use a handheld camcorder with $\beta = 10°$ and R = 50 cm (the arm length), we have $r = 0.174R = 8.68$cm. A stereo vision with such a baseline (*2r*) should be enough for human stereo viewing; however it could be inconvenient to view the scene if you stretch your camera away from your eyes when capturing the video. For accurate 3D reconstruction application, we may need larger baselines therefore the acquisition rig might be much bigger than a handy camcorder.

3.2. A simple catadioptric approach

Here a simple catadioptric approach is proposed that overcome these two drawbacks. The idea is to use planar mirrors to enlarge the viewing circle given the same angular distance 2β between the two columns for the omnistereo. In the catadioptric stereo rig, two planar mirrors with angle 2γ are mounted at distance R from the optical center of the camera (Fig. 6). When the entire catadioptric system of the camera and the mirrors is rotating

around the optical center of the camera, the optical centers of the two virtual cameras (by mirror reflections) form a circle of radius R_v:

$$R_v = 2R \sin \gamma \tag{10}$$

The radius of the viewing circle for the catadioptric circular projection omnistereo with the two virtual cameras is

$$r = 2R \sin \gamma \sin(90° - \gamma + \beta), \quad \beta \le \gamma < 90° \tag{11}$$

The maximum radius can be obtained when the angle between the two mirrors is

$$\gamma = (90° + \beta)/2 \tag{12}$$

The radius of the viewing circle of the optimal catadioptric omnistereo under this configuration is

$$r_{max} = 2R \sin \frac{90° + \beta}{2} \cos(\frac{90° - \beta}{2}) \tag{13}$$

As a comparison, if we use the same parameters as in the single-camera case ($\beta = 10°$ and $R = 50$ cm), the optimal catadioptric omnistereo has two mirrors with angle $2\gamma = 100°$. The reflections of the two off-center rays of angular distance 2β form much larger viewing angle (140°) than 2β (=20°), and a much larger viewing circle with radius $r_{max} = 1.17R = 58.5$cm, almost 7 times of the single camera case (Fig. 6). On the other hand, if we want to achieve the same radius of viewing circle, the rotating radius is only R = 7.4 cm, and hence small sizes of the mirrors. In summary, the catadioptric circular projection omnistereo rig has exactly the same geometry as the basic one, with two additional advantages: (1). The video is much easier to capture; the camera viewer is right in front of your eyes instead of being stretched out. (2). The system has better baseline/size configuration; a compact omnistereo rig can capture omnistereo pair with large viewing circle.

4. Dynamic Omnistereo

The binocular/ N-ocular omnistereo systems have a few (2 or more) viewpoints, and fixed baselines. In the circular projection omnistereo, the viewpoints of cameras move on a small circular path. Even though the circular projection omnistereo has many viewpoints, it still has a fixed baseline - the diameter of the viewing circle. With a fixed baseline configuration, the depth accuracy is proportional to the square of the depth, in both the binocular / N-ocular and circular projection omnistereo systems. In this section, we explore the configurations that the viewpoints of omnidirectional cameras can move freely in a large space. In the dynamic omnistereo, or *panoramic virtual stereo*[10] where the viewpoint and baseline relation can change, it is interesting to find the best configuration for estimating the distance of a target. We have studied the potentials of using only two panoramic cameras to best detect and localize multiple moving humans in a full 360-degree view [10, 27,28]. A detailed numerical analysis of the triangulation error by the panoramic virtual stereo has been given in [27], which leads to useful results for view planning between the two mobile platforms with omnidirectional cameras. Here a simplified version of the analysis is given, following the discussions in Section 2.1. Given the baseline B, the best triangulation is formed when the two rays from the two camera centers to the target has maximum vergent angle (i.e. $\sin \phi_2 = 1$). In this best viewing angle configuration, the depth error can be represented by

$$\partial D = \frac{D^2}{B} \partial \phi \tag{14}$$

It seems that larger baseline will give better depth accuracy. However, since we are dealing with the omnidirectional stereo vision on mobile platforms, the baseline itself should be also determined by a dynamic calibration procedure, and hence is subject to errors. So the error equation for the dynamic omnistereo should be re-written as

$$\partial D = \frac{D}{B} \partial B + \frac{D^2}{B \sin \phi_2} \partial \phi \tag{15}$$

where ∂B is the baseline estimation error. Using a calibration target in the scene, the error of the omnistereo baseline is roughly determined by the target's distance to each camera. The problem of this approach is that we need to have a known target in the scene, which is often impossible in real applications.

4.1. Mutual Calibration and Optimal Configuration

We have proposed a special dynamic calibration procedure called *mutual calibration* based on the visible epipole property in omnistereo. Mutual calibration neither needs to setup any additional calibration targets nor requires using a third object in the environment. Instead each of the omnidirectional cameras can use the other as the calibration

target. In other words, we are making use of the "mutual-occlusion" by the sensors themselves. The advantage of "sensor as the target" in mutual calibration is that the geometric structures and the photometric properties of the sensors as well as their platforms can be well designed and are known *a prior.*

Several practical approaches have been proposed for this purpose by using special structures, such as cylinders, vertical lines and rectangular planar surfaces [28]. The basic idea is to make the detection and calculation robust and fast. One of the approaches is to design the body of each robot as a cylinder with some vivid colors (e.g. white in the intensity images of our current implementation), which can be easily seen and extracted in the image of the other robot's camera (Fig. 7a). We assume that the rotation axis of each panoramic camera is coincident with the rotation axis of the cylindrical body of the corresponding robot, therefore the baseline between the two panoramic cameras can be estimated using the occluding boundary of either of the two cylinders. For example from the image of camera 2 we have (Fig. 7b)

$$B = R_c / \sin(\frac{\alpha}{2}) \tag{16}$$

where α is the angle between two occluding projection rays measured in the image of camera 2, and R_c is the radius of the 1st cylindrical body. Fig. 7c and 7d show a real example of mutual calibration, where the omnidirectional image pair was transformed from the panoramic annular lens (PAL) camera designed by Pal Greguss [3]. The error in estimating the baseline by Eq. (16) can be derived as

$$\partial B = \frac{B\sqrt{B^2 - R_c^2}}{2R_c} \partial\alpha \approx \frac{B^2}{2R_c} \partial\alpha \tag{17}$$

where $R_c \ll B$, and $\partial\alpha$ is the error in estimating the angle α in an image. Assuming the same angular errors in both mutual calibration and stereo matching, i.e. $\partial\alpha = \partial\phi$, and inserting Eq. (17) into Eq. (15), we can derive that the minimum depth error is
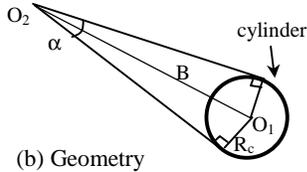
$$\partial D = \frac{2D^{1.5}}{\sqrt{2R_c}} \partial\phi \tag{18}$$

with the following optimal omnistereo configuration (refer to Fig. 1a):

$$\text{Baseline: } B = \sqrt{2R_c D} \text{ ; Vergent angle: } \phi = \sin^{-1}(\frac{B}{D}), \text{ (i.e. } \sin\phi_2 = 1) \tag{19}$$

(a) Setup

(b) Geometry

(image of the cylindrical body of the second robot)

(c) Pano 1: α=11.52° (32 pixels), R_c = 18 cm, B = 180 cm

(image of the cylindrical body of the first robot)

(d) Pano 2: α=11.52° (32 pixels), R_c = 18 cm, B = 180 cm

Fig. 7. Mutual calibration for dynamic omnistereo

In summary, we give the following conclusions:

Conclusion 4.1. The depth accuracy of the dynamic omnistereo (using the specific mutual calibration method) is proportional to the 1.5 power of the subject's distance, and inversely proportional to the square root of the calibration target size.

Conclusion 4.2. The best baseline for a given distance of the subject is proportional to the square root of the subject's distance and the mutual calibration target size.

These two conclusions simply states that better depth accuracy can be achieved by a dynamic omnistereo system than fixed-baseline omnistereo systems, since the viewpoints of the two omnidirectional cameras are freely re-configurable in both vergent angle and baseline. Suppose the baseline length of a fixed-baseline omnistereo system (B in binocular and $2r$ in circular projection) equal to the diameter of the cylindrical calibration body, $2R_c$, then the depth error ratio (Eq. (18) to Eq. (14) or Eq. (8)) is

$$\rho = 2\sqrt{2R_c / D} \tag{20}$$

Obviously dynamic omnistereo has better depth accuracy than fixed omnistereo (i.e., $\rho < 1$) if the distance of the object from the cameras is larger than four times of the baseline in the fixed omnistereo, or the calibration cylinder's size in the dynamic omnistereo (i.e., $D > 8R_c$).

4.2. View Planning with Omnistereo Vision

From the above error analysis, view planning strategy for a single moving object can be derived (Fig. 8a). While the object is moving (from $T^{(1)}$ to $T^{(2)}$), one of the cameras changes its location (from $O_2^{(1)}$ to $O_2^{(2)}$) in order to always form best triangulations to localize the moving object. This implements a *rotating* binocular head that can also *change the distance* between the two eyes. The view planning strategy can be extended to deal with multiple moving objects. There are three interesting cases.

(1). In general, N+1 robots can construct optimal configurations for N moving objects (N >2), i.e. a main robot can cooperate with each of the N robots for the detection and localization of each of the N objects (Fig. 8b). However, this method is inefficient and needs N moving robots.

(2). As a special case (Fig. 8c), two moving robots with panoramic cameras (O_1 and O_2) can construct optimal configurations for estimating the distances of two moving objects($T^{(1)}$ and $T^{(2)}$), which are mirrored to each other by the alignment of the two viewpoints of the cameras.

(3). As an approximation method, two moving robots with panoramic cameras can construct near optimal configurations for estimating the distances of multiple moving objects. This can be done by clustering the targets into two groups, and the two cameras then configure two best triangulations for the centers of the two groups (Fig. 8c). Apparently, more other two robots can do a better job in view planning.
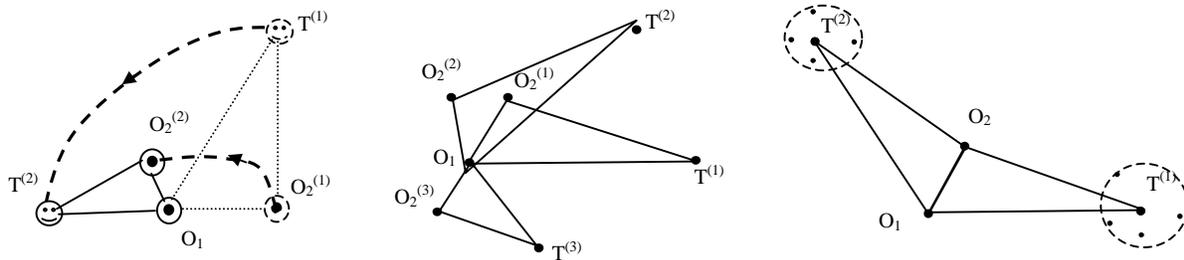
(a). Tracking a single object   (b) N objects, N+1 robots   (c) 2 (groups of) objects, 2 robots
Fig. 8. Optimal view planning for multiple robots and multiple objects

5. Generalized Omnistereo

In the above discussion, an omnidirectional image is represented in a cylindrical projection, either from a single viewpoint or viewpoint along a viewing circle. Shum et al [18] proposed the spherical omnistereo in which a camera moves over a spherical surface (Fig. 9a). In principle, we can consider a point camera moving along a viewing circle, collecting only two tangent rays in the plane of the circle, one forward and one backward, at each viewpoint. After a complete scan of the entire circle (denoted by ϕ), the viewing circle rotates a small angle (denoted by φ) around a pole passing through the center of the viewing circle. The same scan procedure collects another full circle of rays. In this way, a pair of omnidirectional stereo images can be generated , both of which are indexed by (ϕ,φ). The omnistereo pair has the full coverage of the 360x360 FOV, and exhibits horizontal epipolar lines (along the ϕ direction), and has the same depth equation as in Eq. (7). Unfortunately, this configuration is not practical since it is difficult to move a camera precisely over a spherical surface. However, if we turn our attention from the viewer-centered omnidirectional imaging to an object-centered one, it could become an attractive approach.
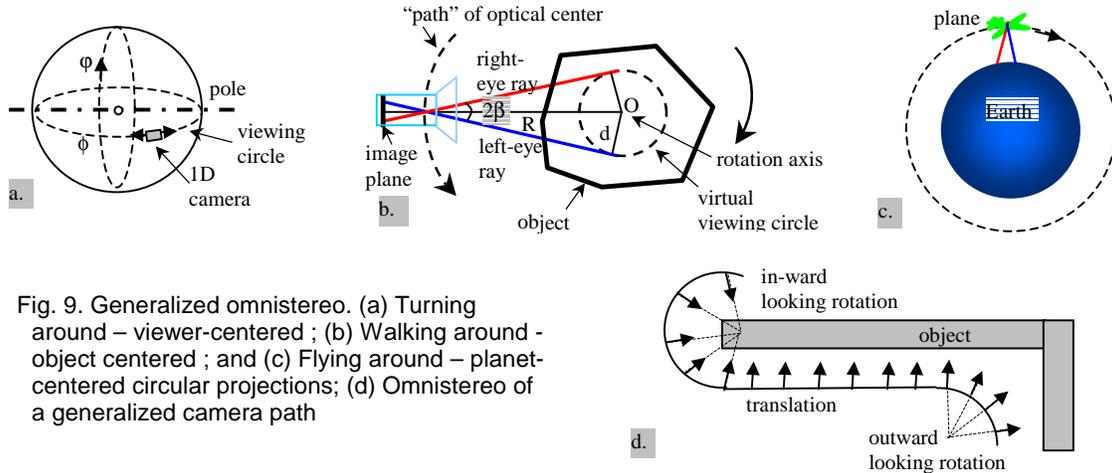
Fig. 9. Generalized omnistereo. (a) Turning around – viewer-centered ; (b) Walking around - object centered ; and (c) Flying around – planet-centered circular projections; (d) Omnistereo of a generalized camera path

5.1. From Viewer-Centered to Object Centered Rotation

Needless to say that it is desirable to fully model a 3D object from all directions. For this purpose we extend the concept of omnidirectional imaging to a full 360-degree FOV coverage of the object. In this case, a camera pointing to the center of the object moves on a spherical surface embracing the entire object. While it is also difficult to control the motion of the camera, we can instead rotate the object if it is small. It becomes very simple since we only need to rotate the object around two orthogonal axes, and keep the camera pointing to the center of the rotation (Fig. 9b). To understand the omnistereo geometry, Fig. 9b shows the 1D case: an object is rotating around its center O, while the camera is pointing to this center. Two rays are collected in the 1D camera – left-eye and the right eye rays, with angular distance 2β. The object-centered omnistereo geometry is essentially equivalent to the "viewer-centered" circular projection omnistereo from a virtual viewing circle inside the object. Again, assume that the distance from the optical center of the camera to the rotation center is R, then the radius of the viewing circle can exactly be computed by Eq. (9), and the distance of a point on the surface of the object to the rotation center can calculated from Eq. (7), too! It has almost the same properties as the viewer-centered circular projection omnistereo in Fig. 5, except that the omnidirectional images and the depth are of the outside surfaces of the object in the object-centered representation. Additional constraint on the data collection is that the viewing circle should be completely inside the object and the circular path of the real camera should be outside the object. Assume the depth range of the object is $[D_{min}, D_{max}]$ in the object-centered omnidirectional representation, we have

$$R > D_{max} , r < D_{min} , \text{ and } \beta < \sin^{-1}(D_{min} / D_{max}) \tag{21}$$

For example, with a range ratio $D_{min}/D_{max} =1/4$, the angular distance of the left and right slit windows should be smaller than 14.5° otherwise the object will be "out of the views" of the omnistereo images.

5.2. Walking Around an Object or Flying around the Earth

We can further loose the motion constrain over a spherical surface to a rather arbitrary smooth motion along a 3D path with known camera poses, e.g. from GPS/INS instrumentation. Then we can model a large building or even our planet by using a single video camera! For example, we can fly an airplane over the earth along successive great circles to create a pair of huge object-centered omnivergent stereo images. Then we can view the earth in 3D using a pair of stereo glasses or recover the omnidirectional 3D map of the earth.

For large objects such as roadside buildings and the earth, the motion along a circular path can be well approximated by translational motion. In fact, an arbitrary camera path can be divided into connecting combinations of outward looking rotation ("viewer-centered" circular projection), inward-looking rotation ("object-centered" circular projection) and translation("parallel projection") (Fig. 9d)[21]. As the first step, we have generated pairs of stereo mosaic strips using the image mosaicing techniques [23-26]. The omnistereo in the translational part can be modeled by the parallel-perspective stereo [23-25,30], having the following properties:

Conclusion 5.1: Parallel-perspective stereo mosaics provide a stereo geometry with adaptive baselines for all the points of different depths.

Conclusion 5.2: The depth accuracy of parallel-perspective stereo is independent of absolute depths in theory [23-25, 30], and the depth error only linearly increases with depths in stereo mosaics from real video of perspective images[26].

Conclusion 5.3:In the ideal case where the viewpoints of stereo mosaics lie in a 1D straight line, the epipolar curves in the general case of 3D motion in stereo mosaics will turn out to be horizontal lines.

Generally, parallel-perspective is superior to all the fixed baseline stereo configurations, due to its adaptive baseline stereo configuration. Fig. 10 shows a pair of stereo mosaics from a video sequence taken in a ground vehicle driving on the road. Depth map was recovered by an efficient epipolar plane analysis method [21,22]. Fig. 11 and Fig. 12 show stereo mosaics from video sequences taken from an airplane of 1000 ft above the ground, for a forest and an urban scene respectively. Depth map in Fig. 11 was recovered by an efficient hierarchical sub-pixel stereo match program[31]. In Fig. 12, stereo viewing image (right-view in the red channel and left view in the blue and green channels) shows vivid 3D effect of the buildings by using a pair of simple red/blue glasses, without any computation of 3D information. These results show the possibilities of modeling and rendering large objects (e.g. a building) and our planet using the object-centered omnistereo vision techniques.
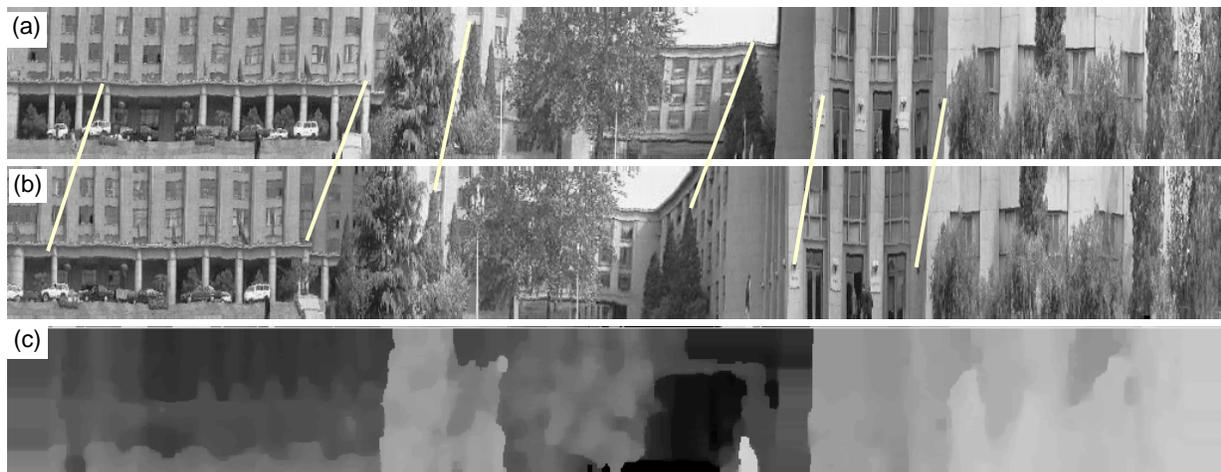
Fig. 10. (a), (b) Omnistereo pair and (c) the depth map of the Main Building at Tsinghua University from a ground vehicle video sequence (1024 frames of 128*128 gray scale images).
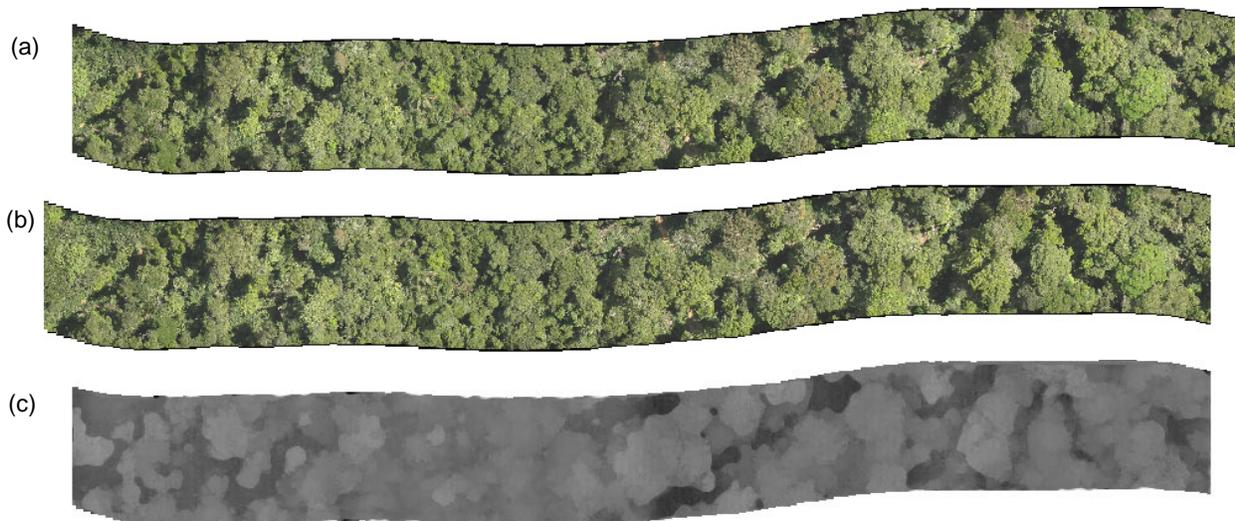
Fig. 11. (a), (b) Omnistereo pair and (c) the depth map of the Amazon rain forest scene from an aerial video camera (166 frames of 720*480 color images)
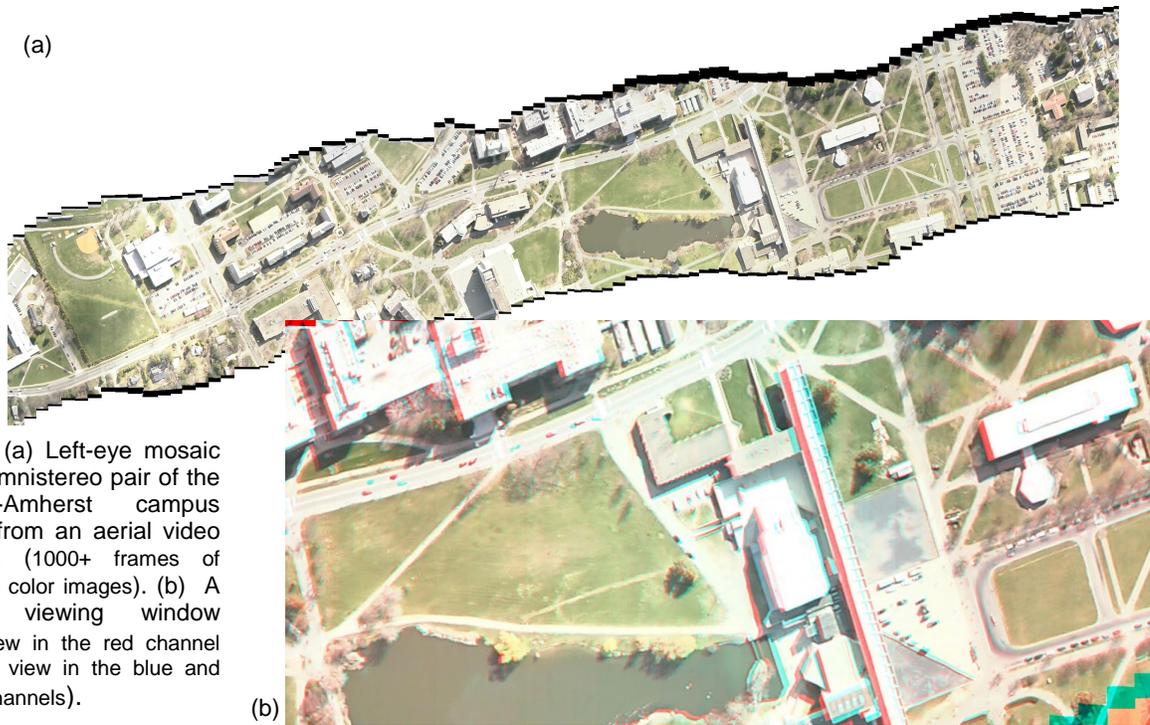
(a)

Fig. 12. (a) Left-eye mosaic of an omnistereo pair of the Umass-Amherst campus scene from an aerial video camera (1000+ frames of 720*480 color images). (b) A stereo viewing window (right-view in the red channel and left view in the blue and green channels).

(b)

Table 1 Comparison of different omnistereo configurations

Omnistereo Configuration	viewpoints	baselines	vergence	epipolar geometry	depth error in directions	depth error in distance	mutual-occlusion	stereo viewing
H-Binocular	2, fixed	fixed	Largely un-uniform	sine curve	non-isotropic	$\propto D^2$	yes	okay
N-Ocular	N (a few), fixed	fixed	Select from C_N^2 pairs	sine curve	roughly "isotropic"	$\propto D^2$	maybe no	okay
V-Binocular	2, fixed	fixed	Similar to perspective	vertical lines	isotropic	$\propto D^2$	no	No
Dynamic	2, freely movable	optimal for given points	max for a few points (pts)	sine curve	optimal for given points	$\propto D^{1.5}$	mutual awareness	N/A
VCP[1]	many, on a small circle	fixed	max for all pts	horizontal lines	isotropic	$\propto D^2$	no	good
OCP[2]	many, on a large circle	fixed	max for all pts	horizontal lines	isotropic	$\propto D^2$	no	good
PPP[3]	many, on a camera path	optimal for all pts	max for all pts	horizontal lines	uniform everywhere	uniform or $\propto D$	no	good

1. VCP : Viewer-centered Circular Projection; 2. OCP: Object-centered Circular projection; 3. PPP: Parallel-Perspective Projection

6. Summaries

This paper discusses a class of interesting configurations of omnidirectional stereo (omnistereo) - binocular omnistereo, N-ocular omnistereo, circular projection omnistereo and dynamic omni-stereo based on how many viewpoints of the omnistereo configurations and how they are configured. A new catadioptric circular projection omnistereo rig is designed that is more compact, is easier for capture, and has better stereo configuration. A novel dynamic omnistereo approach is presented in which viewpoints of two omnidirectional cameras can form optimal stereo configurations for localizing moving objects. Further extension is made to the concept of omnidirectional imaging from viewer-centered to object centered representation, thus allowing building omnidirectional models of large objects or even our planet. Numerical analysis is given on omnidirectional representation, epipolar geometry

and depth error characteristics (see Table 1 for a summary), which could be very useful for the research and applications of omnidirectional stereo vision.

References

[1]. Baker S and Nayar S K, 1998, A theory of catadioptric image formation, In *ICCV'98*: pp 35-42

[2]. Nalwa V, 1996. A true omnidirectional viewer. *Technical Report*, Bell Lab, Holmdel, NJ.

[3]. Greguss P, 1986. Panoramic imaging block for three-dimensional space, *U.S. Patent 4,566,763* (28 Jan)

[4]. Yagi Y, Kawato S, 1990. Panoramic scene analysis with conic projection, In *Proc. IROS*.

[5]. Yamazawa K, Yagi Y and Yachida M, 1993. Omnidirectional imaging with hyperboloidal projections, *Proc. IROS*.

[6]. Powell I., 1994. Panoramic lens, *Applied Optics*, 33 (31):7356-7361

[7]. Gluckman J, Nayar S K, and Thorek K, 1998. Real-time omnidirectional and panoramic stereo, In *Proceedings of DARPA Image Understanding Workshop*, volume 1, pages 299- 303.

[8]. Konolige K G, Bolles R C, 1998, Extra set of eyes. In *Proceedings of DARPA Image Understanding Workshop*, volume 1, pages 25- 32.

[9]. Sogo T, Ishiguro H, Trivedi M M, 2000, N-Ocular Stereo for Real-time Human Tracking. *Panoramic Vision: Sensors, Theory and Applications*, (R. Benosman and S. B. Kang, eds.), Springer Verlag.

[10]. Zhu Z, Rajasekar K D, Riseman E M, Hanson A R, 2000. Panoramic virtual stereo vision of cooperative mobile robots for localizing 3D moving objects. *Proceedings of IEEE Workshop on Omnidirectional Vision – OMNIVIS'00*, Hilton Head Island, 29-36.

[11]. McMillan L and Bishop G, 1995. Plenoptic modeling: an image-based rendering system. In *ACM SIGGRAPH 95*: pp 39-46.

[12]. Kang S B and Szeliski R, 1996. 3D scene data recovery using omnidirectional multibaseline stereo , *IEEE CVPR'96*: pp 364-370.

[13]. Shum H-Y, Han M and Szeliski R, 1998. Interactive construction of 3D models from panoramic mosaics. *IEEE CVPR'98:* pp. 427-433.

[14]. Huang H-C and Hung Y-P, 1998, Panoramic stereo imaging system with automatic disparity warping and seaming. *Graphical Models and Image Process.*, 60(3): 196-208.

[15]. Peleg S and Ben-Ezra M, 1999, Stereo panorama with a single camera. *CVPR'99*: 395-401

[16]. Peleg S, Ben-Ezra M and Pritch Y, 2001. Omnistereo: Panoramic Stereo Imaging, IEEE Trans. PAMI, 23(3): pp 279-290.

[17]. Shum H -Y and Szeliski R, 1999, Stereo reconstruction from multiperspective panoramas. *ICCV99*, 14-21.

[18]. Shum H-Y, Kalai A, Seitz S M, 1999, Omnivergent stereo. In *ICCV'99*: pp 22 – 29

[19]. Ishiguro H, Yamamoto M, and Tsuji S, 1990, Omni-directional stereo for making global map. *ICCV'90*, 540-547.

[20]. Zheng J Y and Tsuji S. 1992. Panoramic representation for route recognition by a mobile robot. *IJCV*, 9(1): 55-76.

[21]. Zhu Z, 1997. *Environment Modeling for Visual Navigation*, Ph.D. Dissertation, Tsinghua University, May 1997 (in Chinese). A revised version with title "*Full View Spatio-Temporal Visual Navigation -Imaging, Modeling and Representation of Real Scenes*" will be published by China Higher Education Press, 2001.

[22]. Zhu Z, Xu G, Lin X, 1999. Panoramic EPI Generation and Analysis of Video from a Moving Platform with Vibration, *Proc. IEEE CVPR'99*: 531-537

[23]. Zhu Z, Hanson A R, Schultz H, Stolle F, Riseman E M, 1999, Stereo Mosaics from a Moving Video Camera for Environmental Monitoring, *First International Workshop on Digital and Computational Video*, December 10, 1999, Tampa, Florida, USA, pp. 45-54.

[24]. Zhu Z, Riseman E M, Hanson A R, 2001. Theory and practice in making seamless stereo mosaics from airborne video. CS *TR #01-01*, Umass-Amherst, January.

[25]. Zhu Z, Riseman E M, Hanson A R, 2001. Parallel-perspective stereo mosaics, In *ICCV'01*, Vancouver, Canada.

[26]. Zhu Z, Hanson A R, Schultz H, Riseman E M, 2001. Error characteristics of parallel-perspective stereo mosaics, *IEEE Workshop on Video Registration* (with ICCV'01), Vancouver, Canada, July 13.

[27]. Zhu, Z., K. D. Rajasekar, E. M. Riseman and A. R. Hanson, 3D localization of multiple moving people by an omnidirectional stereo system of cooperative mobile robots, *Technical Report TR #00-14*, Computer Science Dept., UMASS-Amherst, March, 2000

[28]. Zhu, Z., E. M. Riseman, A. R. Hanson, Geometrical modeling and real-time vision applications of panoramic annular lens (PAL) camera, *Technical Report TR #99-11*, Computer Science Dept., UMASS-Amherst, Feb. 1999

[29]. Benosman R, Devars J, 1998. Panoramic stereovision sensor ", In *ICPR'98*, August 20-25, Brisbane, Australia

[30]. Chai J and Shum H -Y, 2000, Parallel projections for stereo reconstruction. *CVPR'00*: II 493-500.

[31]. Schultz H, 1995. Terrain Reconstruction from Widely Separated Images, In *SPIE*. Orlando, FL.