Technical Report TR #99-11, Computer Science Department, University of Massachusetts at Amherst, February, 1999

Geometrical Modeling and Real-Time Vision Applications of a Panoramic Annular Lens (PAL) Camera System

Zhigang Zhu, Edward. M. Riseman, Allen R. Hanson

Computer Vision Lab

Computer Science Department

University of Massachusetts at Amherst

Amherst, MA 01003

{zhu | riseman | hanson}@cs.umass.edu

http://www.cs.umass.edu/~zhu

Geometrical Modeling and Real-Time Vision Applications of a Panoramic Annular Lens (PAL) Camera System

Zhigang Zhu*, Edward. M. Riseman, Allen R. Hanson

Computer Vision Lab, Computer Science Department

University of Massachusetts at Amherst, Amherst, MA 01003

{zhu | riseman | hanson}@cs.umass.edu

Abstract

Flexible, reconfigurable vision systems can provide the richest sensing modality for sophisticated multiple robot platforms. We propose a cooperative and adaptive approach to vision applied to the problem of finding and protecting humans by a robot team in such emergent circumstances, for example, during a fire in an office building. A panoramic camera system plays an important role in this approach. This report presents the recent progress on the development of vision algorithms based on a panoramic annular lens (PAL) camera system. First, we give a brief survey of existing panoramic camera systems, with an emphasis on geometrical properties of capturing panoramic images from a single-viewpoint. Second, a mathematical model of the panoramic annular lens (PAL) camera that we use is built, and then several issues about camera calibration and image un-warping are addressed. An empirical method is given to roughly 'calibrate" the PAL camera system. Next, a panoramic virtual stereo (PVS) vision approach is proposed, and the problems of self-calibration and real-time 3D estimation are discussed. Finally, we present an experimental system that can detect and track multiple moving objects in real-time using a PAL camera system. The panoramic vision algorithm for moving object tracking also serves the initial step of the PVS approach.

Keywords: *Panoramic imaging, robot team, moving object tracking, cooperative virtual stereo*

---

* The author is currently on a leave from Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P. R. China.

I. Introduction

A multi-objective system requires that the sensory algorithms should have enough flexibility to support adaptation in unknown environments and to facilitate on-line dynamic reorganization of algorithms and sensors in response to a given task. The approach by the Computer Vision Lab, in collaboration with the Perceptual Robotics Lab, in the Computer Science Department at the University of Massachusetts at Amherst, involves the development of behaviors that embody cooperative closely-coupled interactions of the resources across robot teams; this in turn allows virtual sensors to be assembled from multiple, cross modal sensors to make spatial observations, detect key objects and events, and to interpret context.

Our approach can be viewed as a methodology for development of **Cooperative and Adaptive Vision and other Sensing (CAVS)** to provide a set of sensor processing techniques that can fulfill both low-level and high-level objectives in an open environment. The cooperative interaction of the members and elements of the robot team requires system resources, including robot platforms, sensors, computation, and communication resources, to be used effectively by the mission planner.

To achieve the desired robustness, our 3-robot team will be outfitted with a variety of sensors and algorithms. Vision will be the primary sensing modality, but it will be complemented by inexpensive infrared point sensors, sonar, and possibly acoustic sensors. Multiple types of sensors will likely be placed on a single robot platform to allow flexibility by the mission planning and resource scheduling systems in response to changing conditions and goals. The sensors and algorithms are intended to support the following functionality:

1. Detecting and tracking people and other moving objects;
2. Recovery of depth and acquisition of 3D topological models of the environment;
3. Obstacle avoidance and path planning;
4. Recognition of key objects such as entrances and passageways, and selected landmarks to support landmark-based navigation;

5. Recognition of key events such as fires, blocked doorways, etc.

In the current stage we have chosen to focus on vision sensing and algorithms as the primary sensing modality.

1.1. Cooperative and Adaptive Vision (CAV)

Flexible, reconfigurable vision systems can provide the richest sensing modality for sophisticated robot platforms. We propose a cooperative and adaptive approach to vision applied to the problem of finding and protecting humans in such emergent circumstances, for example, during a fire in an office building. The primary vision sensors that we assume during the following discussion are standard CCD cameras and special 360-degree panoramic cameras; in addition, other sensors such as infrared motion sensors and acoustic sensors will be utilized. Thus, we will equip our robotic team with these sensors and develop dynamic and cooperative algorithms for effectively utilizing the sensory data.

Diverse sensory functionality will support the development of robust multi-robot, multi-sensor behaviors involving tracking, 3D modeling, visual cooperation among sensors on different platforms, and visual learning and adaptation. Each of these in turn can be broken down into the structured application of a set of primitive operations. For example, in dealing with moving objects, the processing is decoupled into five stages (motion detection, object extraction, moving object tracking, 3D localization and object identification). Varying amounts of time and resources are needed for different stages, but every stage can provide some useful information to the control system (which may also be sufficient in a given context).

1.2 Cooperative Panoramic Vision and Behaviors Embodying Virtual Stereo Sensing

Real-time processing is essential for the dynamic and unpredictable environments in our application domain, and it is important for visual sensing to rapidly focus attention on important activity in the environment. Any room or corridor should be searched quickly to detect people and fire. Field-of-view issues using standard optics are challenging since

panning the camera takes time, and multiple targets/objectives may require saccades to attend to important visual cues. Thus, we employ a camera with a panoramic lens to detect and track multiple objects in motion in a full 360-degree view in real time.

We note that there is a fairly large body of work on detection and tracking of humans [Bri98, et al; Har98, et al; Lip98, et al; Pap98, et al; Pen98, et al], motivated most recently by the DARPA VSAM effort (part of the DARPA IU program)[DARPA98]. What is truly novel about our approach is the ability to compose cooperative sensing strategies across the robot team to synthesize *robust virtual stereo sensors*. We briefly discuss here a strategy for virtual stereo processing using multiple optical sensors of different types.

Any fixed-baseline stereo vision system has limited depth resolution because of the physical constraints imposed by the separation of cameras, whereas a system that combines multiple views allows the planning system to take advantage of the current context and goals in selecting viewpoints. This strategy can be implemented by a single camera generating sequential viewpoints over time in an active vision paradigm [Alo93] and/or as in traditional motion analysis [Kum92, et al, Saw93, et al; Ren98, et al]. However, there are significant time delays involved in moving the camera to another position in the room. Instead we focus on cooperative behavior involving *mutually aware cameras* on different mobile platforms to compose a virtual stereo sensor with a flexible baseline. The sensor geometry can be controlled to manage the precision of the resulting virtual sensor.

Critical issues in this approach include the *dynamic self-calibration* of the combination of cameras on 2-3 mobile robots, which forms the dynamic virtual stereo sensor. Several different scenarios must be considered in order to achieve the required robustness to sensor availability and types. To perform calibration (a necessary precursor to stereo recovery), a known target in the scene is needed, otherwise the motion between two snapshots can only be determined up to scale factor. We are considering several methods for calibrating the cameras across the team to alleviate accuracy problems inherent in odometry. One method involves robot platforms in each other's field of view; recognition of each other could be facilitated by appearance-based models [Rav97]] of

the robots, special features [Ris97 et al; Col98 et al; Hoe98 et al] on the bodies, active signals such as visual beacons formed from flashing lights, etc. The "*mutual calibration*" strategy may be particularly effective with a pair of mobile panoramic sensors that have the potential of always seeing each other. Once calibrated by looking at each other, they can view the environment to estimate the 3D structure of the scene.

It can be seen that a panoramic camera system plays an important role in our CAVS approach. This report presents our recent progress on the development of vision algorithms based on a panoramic annular lens (PAL) camera system. The organization of this report is as follows. In Section II, we give a brief survey of existing panoramic camera systems, with an emphasis on geometrical properties of having panoramic images with a single-viewpoint. In Section III a mathematical model of our panoramic annular lens (PAL) camera is built, and then several issues about camera calibration and image un-warping are addressed. Section IV gives an empirical method to roughly 'calibrate" the PAL camera system, which is sufficient in our application. A panoramic virtual stereo (PVS) vision approach is proposed in Section V. The algorithms of self-calibration and real-time 3D estimation are discussed in this section. In Section VI we present an experimental system that can detect and track multiple moving objects in real-time using a PAL camera system. The panoramic vision algorithm of real-time moving object detection and extraction also serves as the base of the PVS approach. Finally we give a brief discussion about our future work.

II. Panoramic Imaging System : a Survey

Using a fish eye lens, a set of planar mirrors, a conic mirror, a spherical mirror or a paraboloidal mirror can obtain 360 degree omnidirectional images. The panoramic annular lens (PAL) camera of Optechnology Co., designed by Pal Greguss (1986), can captures its surroundings using a new design based on reflection and refraction of light. The PAL-3802 system, a commercial system of this design, includes a 40 mm diameter PAL and a built-in collector lens with a "C" mount. The virtual image, formed within the PAL, is conveyed to the camera sensor using the collector. The field of view (FOV) is 360-degrees horizontally and -15 ~ +20 degrees vertically. However detailed materials

about the geometrical model of the PAL-3802 are not provided by the manufacturer. In the patent of Greguss (1986), a drawing is included, which depicts an arrangement comprising a panoramic block followed by a tele-centric lens and a video camera. The patent implies that the image is formed at the location where the pupil rays come to a focus, i.e., at the exit pupil. Both reflective surfaces are paraboloidal in shape for achieving nearly "faultless" imagery according to the specifications. Each of these surfaces can be replaced, once again according to the specification, with a surface having a radius of the best-fit sphere for a still acceptable image quality.

Powell (1994) reported the design of a similar panoramic lens, and stated that the Greguss system would work better with spherical reflective surfaces than with paraboloidal ones. In the actual system design, aspheric reflective surfaces are used to further improve the image quality, and a relay lens are added to the Greguss(1986) system. Each of the surfaces is a conicoid revolution, with its asphericity ranging from an oblate spheriod to a hyperboloid.

For vision applications, it is desired that the fixed viewpoint constraint can be satisfied, so that we can generate perfect perspective images from the panoramic image by PAL camera. This property is valuable for robot navigation, virtual environment modeling, teleconferencing, surveillance, where a large body of work in stereo vision or visual motion algorithms under linear perspective projection are applied. However none of the above designs mentioned this problem. In this report, we assume that the PAL-3802 satisfies with single viewpoint constraint and two possible geometrical models will be given. If it is not true for the PAL, we are designing a new single-viewpoint PAL!

Here we briefly list some of the useful conclusions on panoramic imaging. The interesting readers can refer to the litterateurs (Nayar97, Baker98) for detailed explanations.

1. *Panoramic camera system using a pyramid* with four planar mirrors and four cameras (Nalwa96) has a single effective viewpoint (Fig. 2.1).

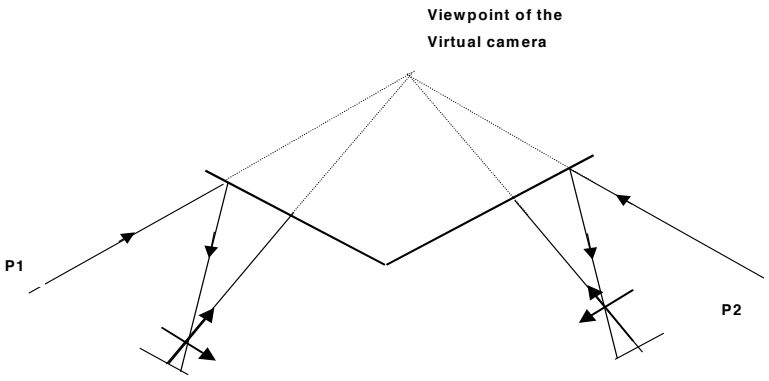


Fig. 2.1 Panoramic camera system using four planar mirrors

2. *Panoramic vision with a conic mirror*: (Yagi90, Zhu96/98): The locus of the effective viewpoint is a circle (Fig. 2.2). The circle will shrink to a point if the pinhole moves to the vertex of the cone, but the camera can only see the mirror in this case.

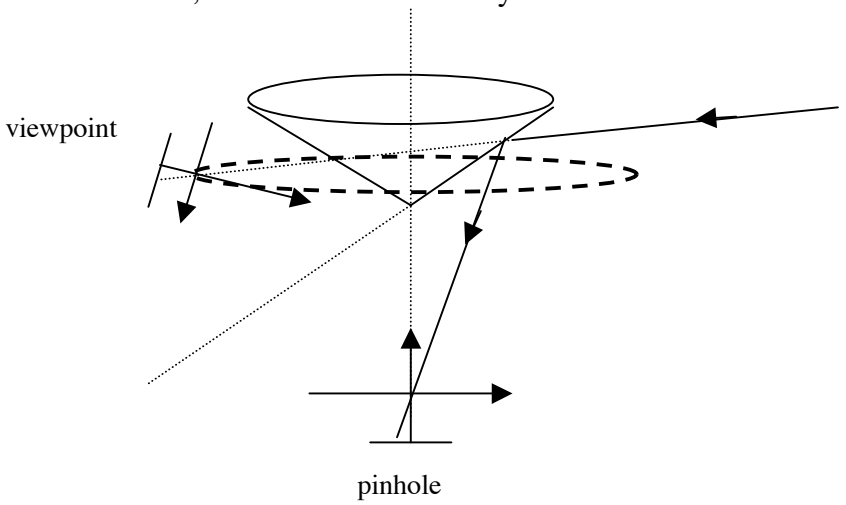


Fig. 2.2. Panoramic vision with a conic mirror

3. *Panoramic vision with a spherical mirror*: (Hong91): The locus of the effective viewpoints lie on a sphere-like surface (Fig. 2.3). Single view point can be satisfied only if the viewpoint and the pinhole coincide at the center of the sphere, when observer can only sees itself.

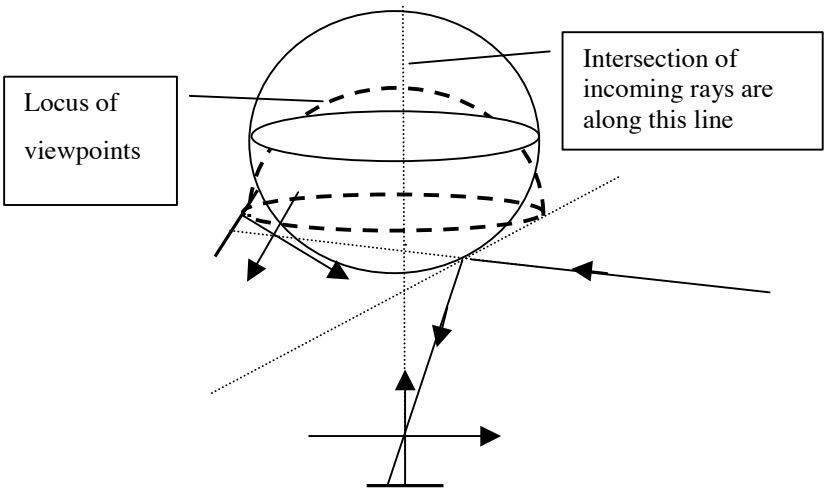


Fig. 2.3. Panoramic vision with a spherical mirror

4. *Panoramic vision with an ellipsoidal mirror* (Nayar97, Baker98): It satisfies the fixed viewpoint constraint if the pinhole of the real camera and the virtual viewpoint are located at the two loci of the ellipsoid respectively (Fig. 2.4).

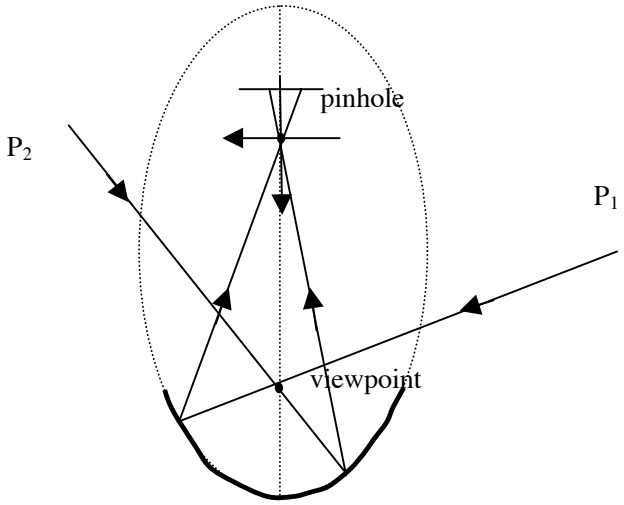


Fig. 2.4. Panoramic vision with an ellipsoidal mirror

5. *Panoramic vision with a hyperboloidal mirror* (Yam 93, Baker98): It satisfies the fixed viewpoint constraint if the pinhole and the viewpoint are located at the two loci of the hyperboloid (Fig. 2.5).

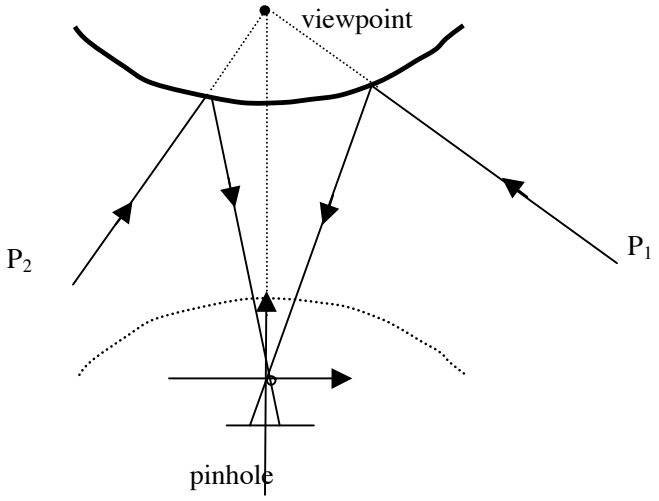


Fig. 2.5. Panoramic vision with a hyperboloidal mirror

6. *Panoramic vision with a paraboloidal mirror* (Nayar97): For orthographic projection, the solution is a paraboloid with the viewpoint located as the locus of the paraboloid (Fig. 2.6). Orthographic projection makes the geometric mapping between the image, the mirror and the world invariant to translation of the mirror. This greatly simplifies calibration and the computation of perspective images from paraboloidal ones.

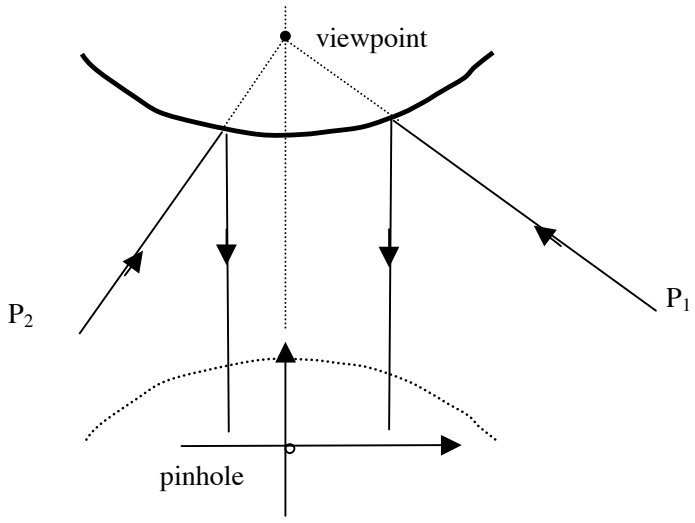


Fig. 2.6. Panoramic vision with a paraboloidal mirror

III. Geometrical Model of PAL

3.1. Perspective model

The geometry of the PAL imaging system is somewhat complex since there are two reflections and two times of refraction (refer to Fig. 3.4). However, if we assume the large circular mirror is an *ellipsoid* and the small top mirror is a *hyperboloid*, and the setup of the two mirrors and a pinhole camera satisfies conditions 4 and 5 in Section II, we can obtain a rather elegant geometry of a single effective viewpoint under *perspective projection*. If that is the case, the real system can be modeled by the single-viewpoint geometry perfectly.

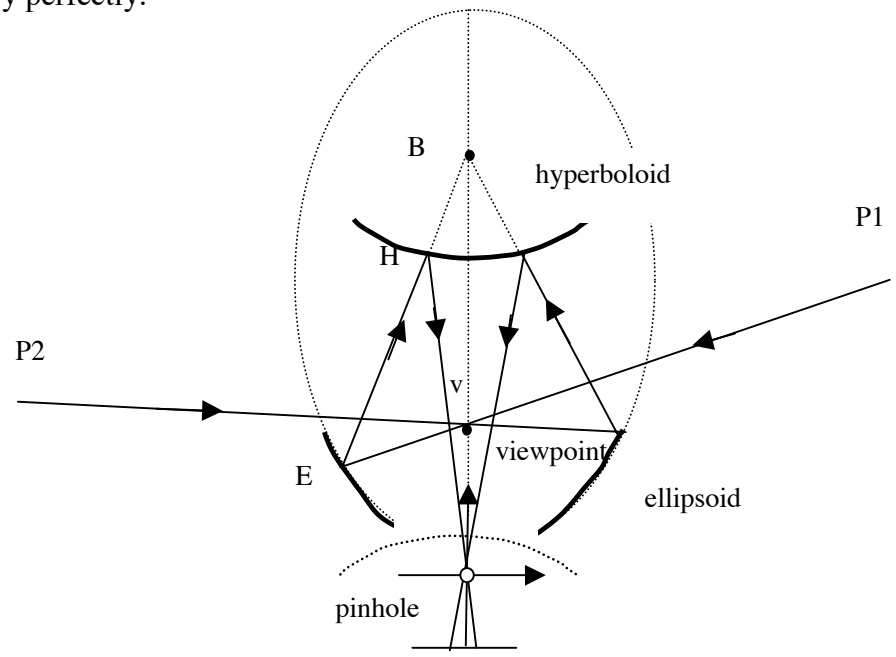


Fig. 3.1 Perspective model of PAL system

At first, we assume that a PAL image is generated by two reflections if we ignore the refraction effects. Suppose the loci of the hyperboloid and the ellipsoid lie on the optical axis of the real camera. One of the loci of the hyperboloid coincides with one locus of the ellipsoid, and the optical center of the pinhole camera is at the other locus of the hyperboloid. Thus, the viewpoint of the virtual camera is right at the second locus of the ellipsoid. The camera can view the entire up-hemisphere scene, except that there is a

circular hole in the center of a PAL image. So the viewing angle is 360 degree horizontally and $0^o \le \alpha < 90^o$ vertically.

3.2. Orthographic model

We still assume that the large circular mirror is an ellipsoid. If the small top mirror is a paraboloid, and the setup among the two mirrors and the pinhole camera satisfies the conditions 4 and 6 in Section II, we can obtain a nice geometry of a single effective viewpoint under *orthographic projection*.

Again, at first, we ignore the refractions, then the PAL image is generated only through two reflections. Suppose the locus of the paraboloid coincides with one locus of the ellipsoid, and the optical axis of the orthographic lens camera is along the long axis of the ellipsoid. Thus the viewpoint of a "virtual camera" is just at the second locus of the ellipsoid. Same as the perspective case, the camera can view the entire up semi-sphere scene except a circular hole in the center. So the viewing angle is also 360 degree horizontally and $0^o \le \alpha < 90^o$ vertically.

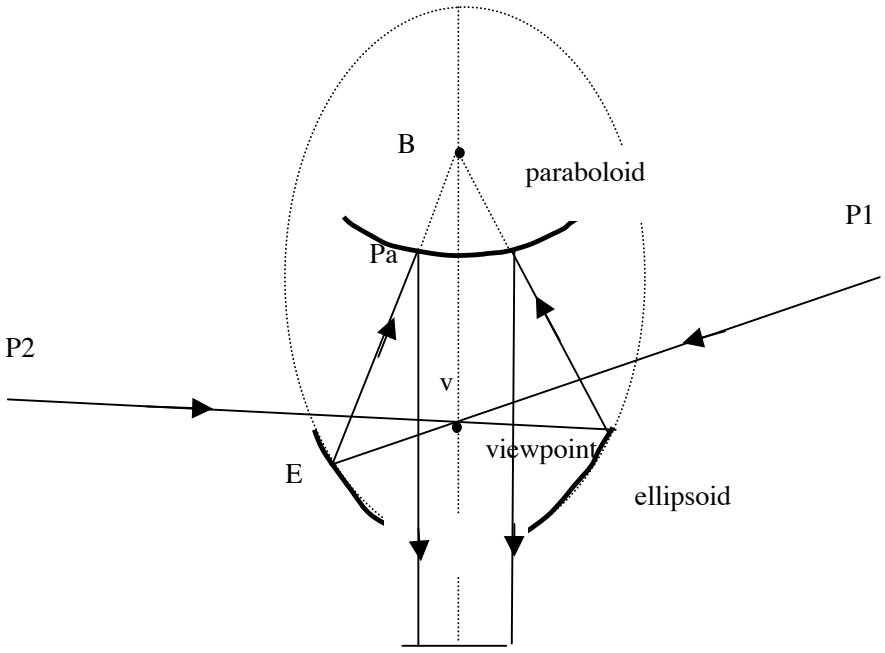


Fig. 3.2 Orthographic model of PAL system

3.3. Two paraboloidal mirrors seem impractical

If the shapes of the two reflective mirrors are paraboloid, as suggested by Greguass (1986), the incoming rays passing through the locus (v) of the first (circular) reflective mirror will be reflected as parallel rays and reach the second mirror. These parallel rays will diverge from the locus (B) of the second mirror.

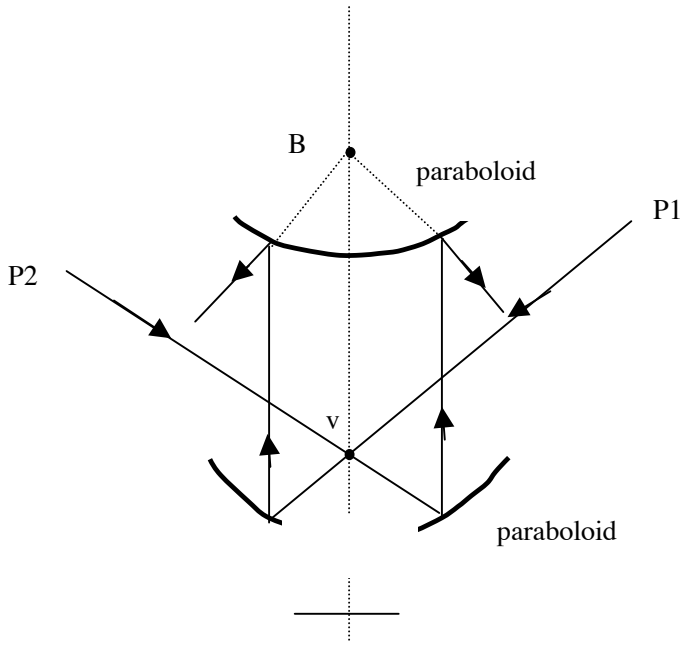


Fig. 3.3 Two paraboloidal mirrors

34. Mathematical models

Considering the refraction does not add too much complex. We will use the pinhole camera model in the following mathematical model development. The first refraction through the ellipsoidal surface changes the vertical viewing range from $[0^o, 90^o)$ to $[-\alpha_1, +\alpha_2]$, where $0^o < \alpha_1 < \alpha_2 < 90^o$, which is often desired for panoramic imagery (Fig. 3.4). The second refraction through the planar surface just moves the converging point of rays from the top mirror up a distance. Since this refraction may be compensated by the collector lens, we will ignore the second refraction.

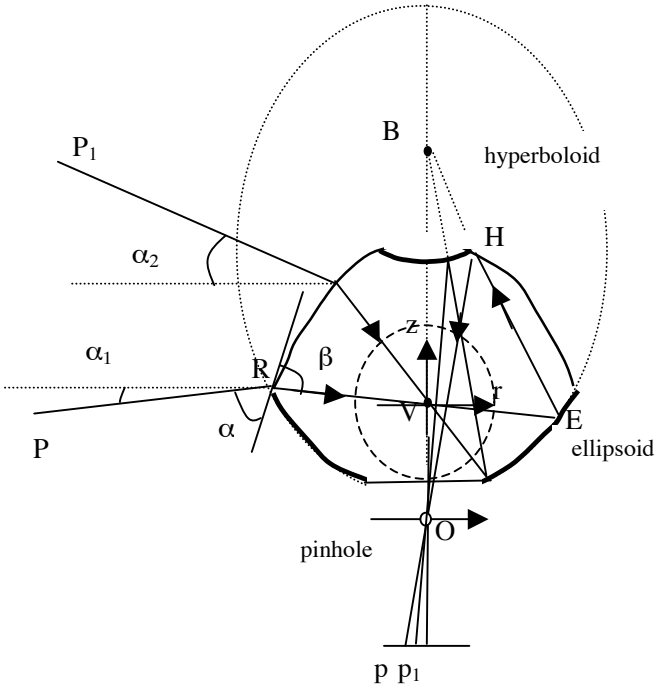


Fig. 3.3. Mathematical models

We only consider the perspective case; the orthographic model has slightly simpler relations. In Fig. 3.4, Define $\overline{VB} = 2c_e, \overline{BO} = 2c_h, \overline{VO} = 2(c_e - c_h)$, and a 3D coordinate system xyz with origin V and z axis as the optical axis of the camera. Denote $r = \sqrt{x^2 + y^2}$. Then we have the following equations to describe the surfaces of the mirrors and the lines of rays.

1. Reflection Hyperboloid (H):

$$\frac{1}{a_h^2}(z - c_h)^2 - \frac{1}{b_h^2}r^2 = 1 \qquad (3\text{-}1)$$

where $2a_h$ and $2b_h$ are the long and short axis of the hyperboloid.

2. Reflection Ellipsoid (E):

$$\frac{1}{a_e^2}(z - c_e)^2 + \frac{1}{b_e^2}r^2 = 1 \qquad (3\text{-}2)$$

where $2a_e$ and $2b_e$ are the long and short axis of the ellipsoid.

3. Ray pO:

$$r = -\frac{r_I}{f}(z + c_h - c_e) \tag{3-3}$$

where f is the focal length of the camera, $r_I = \sqrt{x_I{}^2 + y_I{}^2}$ is the image point.

4. Ray HE

$$r = -\frac{r_h}{z_h - c_e}(z - c_e) \tag{3-4}$$

where (x_h,y_h,z_h) is the intersection of ray HE with the hyperboloid, $r_h = \sqrt{x_h{}^2 + y_h{}^2}$.

5. Ray VR

$$r = -\frac{r_e}{z_e}z \tag{3-5}$$

where (x_e,y_e,z_e) is the intersection of ray VR with the reflection ellipsoid, $r_e = \sqrt{x_e{}^2 + y_e{}^2}$.

6. Refraction Ellipsoid

$$\frac{1}{a_R^2}(z - c_R)^2 + \frac{1}{b_R^2}r^2 = 1 \tag{3-6}$$

where $2a_R$ and $2b_R$ are the long and short axis of the refraction ellipsoid, and $2c_R$ is the distance between two loci. Here we assume that upper locus of the refraction ellipsoid is in the same location of the bottom locus of the reflection ellipsoid. Point (x_R,y_R,z_R) is the intersection of ray VR with the refraction ellipsoid, $r_R = \sqrt{x_R{}^2 + y_R{}^2}$. The tangent line at this point is

$$\frac{1}{a_R^2}(z_R - c_R)(z - c_R) + \frac{1}{b_R^2}r_R r = 1 \tag{3-7}$$

7. Refraction PR

$$\cos a = \lambda \cos\beta \tag{3-8}$$

where λ is the refraction coefficient.

3.5. Camera Calibration Problems

Camera calibration is related to the following mapping.

1. Image warping :

After the camera calibration, the mapping from an original PAL image to a unit sphere is as follows

$$(r_I,\theta)->(r_h,\theta,z_h)->(r_e,\theta,z_e)->(r,\theta,z)$$

where θ *is the orientation angle around z axis,* (r,θ, z) is a point in the unit sphere whose origin is V. Then spherical image can be easily un-warped to a cylinder.

2. Inverse Mapping:

Given a point in a PAL image, we try to define the corresponding ray PR. If two such PAL cameras are used and are calibrated, then the 3D location P(X,Y,Z) can be decided by triangulation. The mapping from image point to the ray is

$$(r_I,\theta)->(r_h,\theta,z_h)->(r_e,\theta,z_e)->(r_R,\theta,z_R)->\beta,(r_R,\theta,z_R)->\alpha,(r_R,\theta,z_R)->PR$$

3. Camera calibration

Given a number of 3D point (X_i,Y_i,Z_i), and its image (r_{Ii},θ), the calibration procedure is try to find the following unknown parameters and points:

$$a_e,b_e,c_e,a_h,b_h,c_h,a_R,b_R,c_R,f,z_e,r_e,z_h,r_h,z_R,r_R$$

The question is to find an algorithm and a numeric solution for this non-linear parameter estimations. Before we go further in this direction, we will stop and look at an empirical solution to the calibration of the PAL camera system. We hope that we can verify our assumption about the model of the PAL camera by carefully designed experiments, while at the same time by-pass the hard problem of non-linear parameter estimations for model-based PAL camera calibration.

IV. Empirical Verification and Image Un-warping

4.1 Center determination

First, we adjust the camera to point vertically upward so that projections of vertical lines in the world remain straight in a PAL image and they intersect at a single point in the center of the PAL image (Fig. 4.1). So if more than two such lines are detected in an original PAL image, this center point can be determined by their intersection.

Once we have the center (x_0,y_0) of the PAL image $I(x,y)$, a cylindrical panoramic image $I(r,\theta)$ can be generated (Fig. 4.2)

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2}, \theta = \tan^{-1}\frac{y - y_0}{x - x_0} \qquad (4\text{-}1)$$

4.2 Radius distortion rectification

Distortion exists in the radial direction due to the non-linear reflections of the 2nd-order mirror surfaces. Notice the unequal widths of the black-white bars on the white broad in Fig. 4.2 without eliminating the radial distortion, where the widths are equal in the real board. In practice, we use an N-order polynomial to approximate the distortion along radius direction:

$$R = v_0 + v_1 r^1 + v_2 r^2 + v_3 r^3 + \ldots \qquad (4\text{-}2)$$

where r is the radius in the original cylindrical image, and R is the radius in the rectified cylindrical image. Fig 4.3 shows the rectification result using a 2nd-order polynomial approximation. Notice the equal intervals of the black bars on the white broad.

4.3 Single viewpoint verification

Using a square broad placed vertically and then with an angle, we are designing an experiment to verify the perspective geometry assumption, taking width-length ratio, straightness, perspective effect, etc. into consideration.

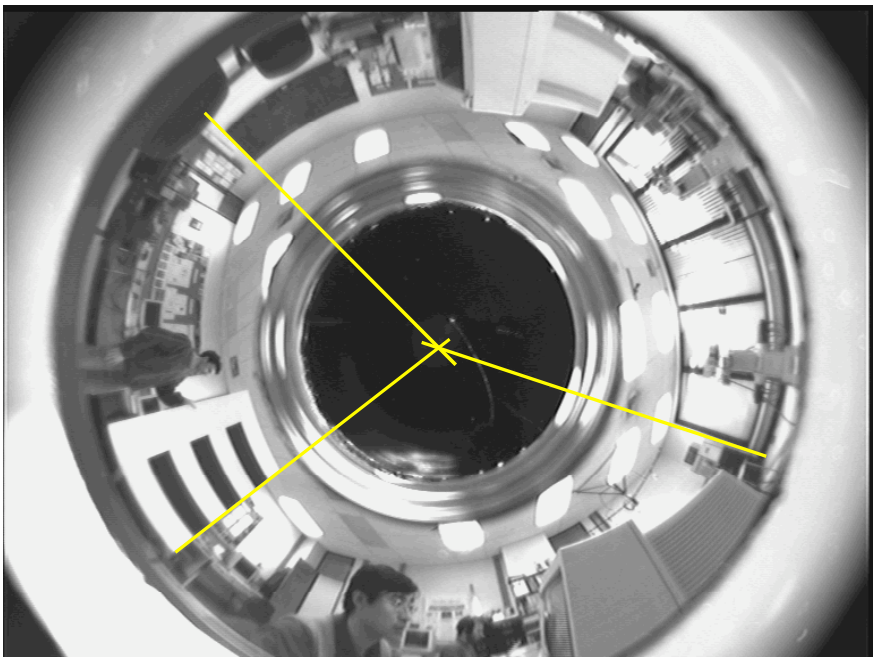


Fig. 4.1. Original PAL (Panoramic Annular Lens) image (768*576)

Fig. 4.2. Cylindrical panoramic image, without eliminating radial distortion

Fig. 4.3. Cylindrical panoramic image, after eliminating radial distortion

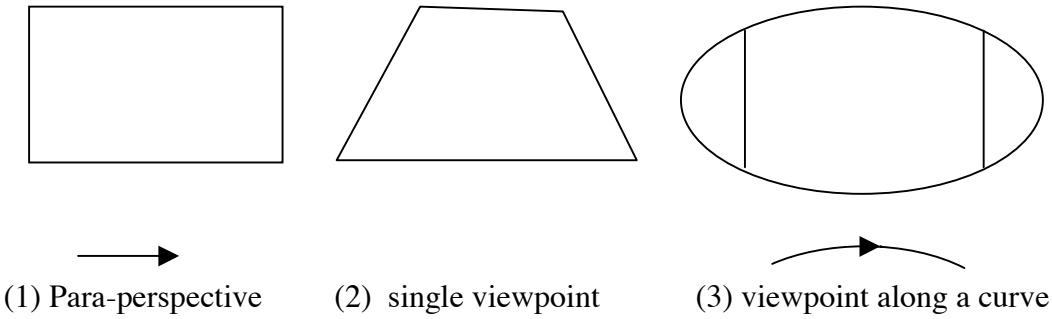


Fig. 4.4. Single viewpoint verification

Fig. 4.4 shows three examples of the possible camera geometry. If the viewpoint is on a straight line or curve, the image of the vertically placed rectangle will be as Fig. 4.4(1) or Fig. 4.4 (3). The relation between the square (rectangle) and its image projected into a plane from a cylindrical image is the planar projective transformation (Fig. 4.4(2)) if the single viewpoint constraint is satisfied. This step needs future work. Fig. 6.1 shows some examples of the perspective images from a PAL image.

V. PAL Stereo and Motion: Self-Calibration and 3D Estimation

In the robot team task, an interesting scenario is that the team leader has a panoramic camera, and one or two other team members are equipped with either panoramic cameras and/or standard cameras. The advantage of *panoramic and cooperative stereo vision* in tracking moving objects is clear: cooperative approaches solve or reduce the severity of some of the difficult problems in calibration, correspondence, and 3D reconstruction. If the team leader can view the member robots, we can determine the baseline and use it as the basis for reconstruction of the environment in the common field of view. This can greatly simplify the recovery of the 3D motion of independently moving objects in the scene (using tracking) as well as the recovery of 3D environmental structure (using point and feature correspondences). The "mutual calibration" strategy is particularly effective with a pair of mobile panoramic sensors that have the potential of always seeing each other.

5.1. Geometry of a Cooperative Panoramic Stereo System

Conceptually, suppose that we have two panoramic cameras and both of them are subject to planar motion on the floor. If they can see each other and in the same time see a target T, Then we can find the bearing and distance of the target without any priori calibration of the two cameras. After the dynamic calibration, the distance of the target can be calculated as

$$D_1 = B\frac{\sin(\beta_{12} - \theta_2)}{\sin(\beta_{12} - \beta_{21} + \theta_1 - \theta_2)} = B\frac{\sin\phi_2}{\sin\phi_0} \qquad (5\text{-}1)$$

where D_1 is the distance between the target and the first camera, B is the distance between the two cameras, θ_1 and θ_2 are the bearings of the target in image 1 and image 2

respectively, and β_{12} and β_{21} is the bearings of the images of camera 1 in image 2 , and of camera 2 in image 1 respectively.

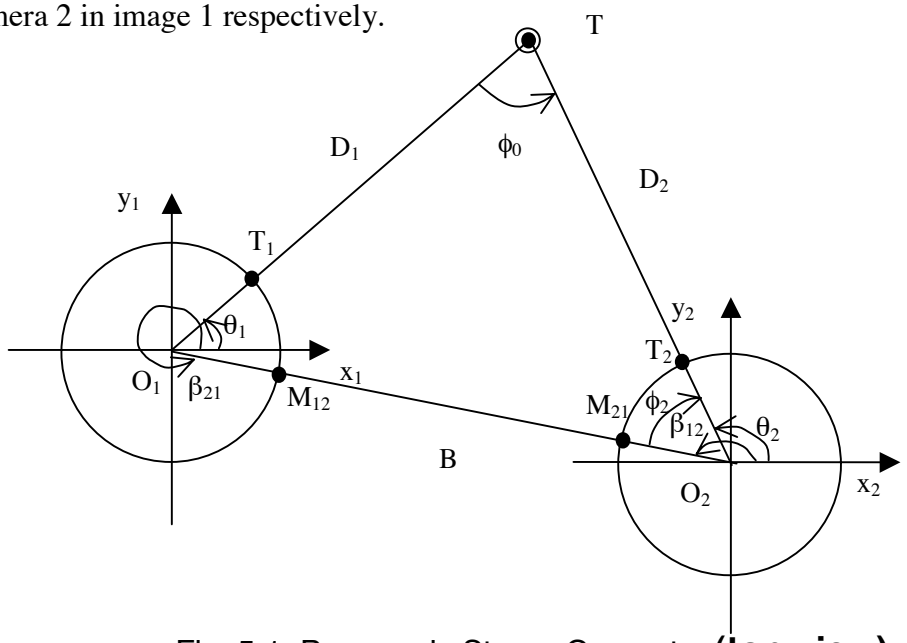


Fig. 5.1. Panoramic Stereo Geometry **(top view)**

5.2. Several practical approaches to estimate distance and angles

The following subsections only discuss the calculation by processing the image of the second robot in the first camera. It is true *vice versa,* and the abundant information can be used for double-check.

5.2.1. Each robot is a cylinder

If the optical axis of each panoramic camera lies in the rotating axis of the cylindrical body of the corresponding robot, the baseline between the two panoramic cameras can be estimated using the occluding boundary of one of the cylinders, i.e.,

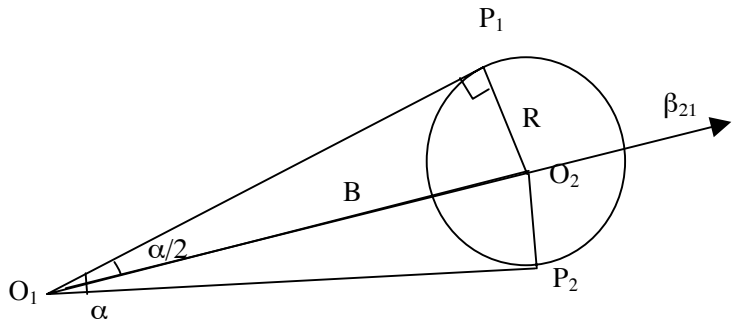


Fig. 5.2. Find the orientation and the distance by a cylinder (top view)

$$B = Rctg(\frac{\alpha}{2}) \tag{5-2}$$

where α is the angle between two occluding projection rays. The orientation angle (β_{21}) of O_1O_2 is simply the average of the bearings of P_1 and P_2.

5.2.2. The target attached to each robot is a set of vertical lines around a circle

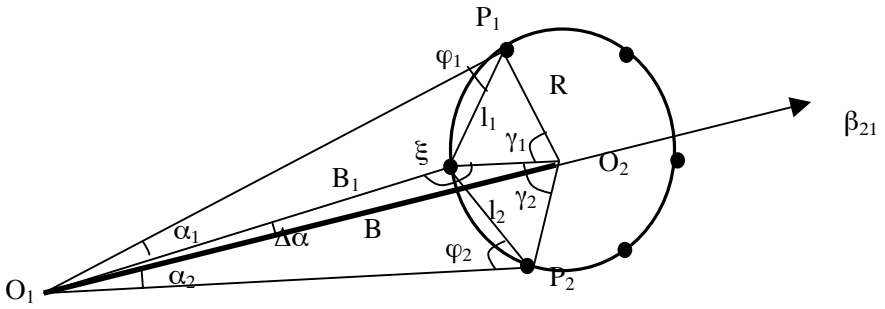


Fig. 5.3. Find the orientation and the distance by vertical bars (top view)

If the target attached to each robot is a set of vertical lines around a circle, we can also determine the bearing and distance of the second robot with at least 3 bars in the view of the first camera. Notice that the calculation is different from the cylinder case, since the bars in the view do not necessarily be the occluding boundary of the cylinder.

From Fig. 5.3 we have

$$B_1 = l_i \frac{\sin\varphi_i}{\sin\alpha_i}, i = 1,2 \tag{5-3}$$

where

$$l_i = R\gamma_i, i = 1,2$$

$$\alpha_1 + \alpha_2 + \varphi_1 + \varphi_2 + 90^o - \frac{\gamma_1}{2} + 90^o - \frac{\gamma_2}{2} + \gamma_1 + \gamma_2 = 360^o$$

so we have

$$\varphi_2 = 180^o - \frac{\gamma_1 + \gamma_2}{2} - (\alpha_1 + \alpha_2) - \varphi_1 \tag{5-4}$$

From equations (5-3) and (5-4) we have

$$A_1 \sin\varphi_1 = A_2 \sin(\alpha + \varphi_1)$$

where

$$A_i = l_i / \sin\alpha_i, i = 1,2$$

$$\alpha = \alpha_1 + \alpha_2 + \frac{\gamma_1 + \gamma_2}{2} \tag{5-5}$$

so we have

$$\varphi_1 = \tan^{-1} \frac{A_2 \sin\alpha}{A_1 - A_2 \cos\alpha} \tag{5-6}$$

Finally the baseline can be calculated by

$$B = \sqrt{B_1^2 + R^2 + 2B_1 R\cos\xi} \tag{5-7}$$

where

$$\xi = 360 - (180^o - \alpha_1 - \varphi_1) - (90^o - \frac{\gamma_1}{2}) = 90^o + \alpha_1 + \varphi_1 + \frac{\gamma_1}{2} \tag{5-8}$$

The orientation angle β_{21} of the baseline is $\angle O_1B_1+\Delta\alpha$, where

$$\Delta\alpha = \sin^{-1} \frac{R\sin\xi}{B} \tag{5-9}$$

5.2.3. The target attached to each robot is a rectangle with known size

If the target is a rectangle with known size and position related to the rotational axis of the camera, then the distance and the orientation of the baseline can also be decided. Suppose that the rectangle is placed vertically. Then the two vertical edges will be vertical in a cylindrical image. Suppose that the heights of them in the image of the first camera are h_1 and h_2 respectively. If the with and the height of the rectangle are W and H, and the distances of these two edges to the rotating axis of the first camera O_1 is B_1 and B_2, then

$$\frac{B_1}{B_2} = \frac{h_1}{h_2}$$

so we have

$$B_2 = \frac{h_2}{h_1} B_1 \tag{5-10}$$

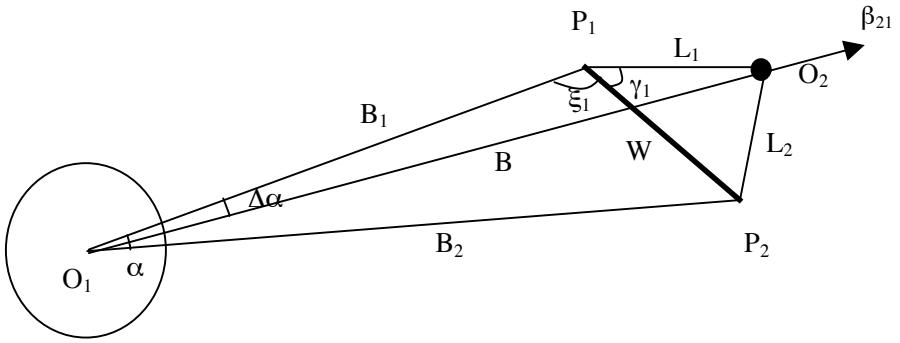


Fig. 5.4. Find the orientation and the distance by a rectangle (top view)

From the cosine relation of

$$W = \sqrt{B_1^2 + B_2^2 + 2B_1 B_2 \cos\alpha}$$

we have

$$B_1 = \frac{Wh_1}{\sqrt{h_1^2 + h_2^2 + 2h_1 h_2 \cos\alpha}} \tag{5-11}$$

and the length of the baseline is

$$B = \sqrt{B_1^2 + L_1^2 + 2B_1 L_1 \cos\xi} \tag{5-12}$$

where

$$\xi = \sin^{-1}\left(\frac{B_2 \sin\alpha}{W}\right) \tag{5-13}$$

The orientation angle β_{21} of the baseline is $\angle O_1P_1+\Delta\alpha$.

5.3. What about rotating camera(s)?

A rotating camera can generate a 360-degree cylindrical panorama, so the 3D geometry is the same as the panoramic stereo system. It does not matter which frame in the image sequence is the reference.

5.4. Error analysis

Here we give an analysis on the error of estimating distance by using equation (5-1). The error of D_1 can be estimated by partial differentials of equation (5-1) as

$$\partial D_1 = \left|\frac{\sin\phi_2}{\sin\phi_0}\right|\partial B + B\left|\frac{\sin(\phi_0+\phi_2)}{\sin^2\phi_0}\right|\partial\phi \qquad (5\text{-}14)$$

where ∂B is the estimating error of the baseline B, and $\partial\phi$ is the average angle error. The smaller the angle ϕ_0 and/or B, the larger is the error. Note that ϕ_0 and B have an inherent dependency. Given D_1 and D_2, the distances between the target and the two cameras (Fig.5.1), the change of ϕ_0 and B are in the same direction (increasing or decreasing).

VI. Fast Moving Object Extraction and Tracking

In the current DARPA-funded Active Software Composition (ASC) project at the University of Massachusetts, initial results have been achieved. A fast moving object extraction and tracking system by using a stationary panoramic vision system has been developed. The system consists of the following steps.

1. Fast Image Un-warping

Look-up Table (LUT) technique is used to map a circular image into a cylindrical image described by equation (4-1). The LUT technique guarantees the real time software image transformation in order to be used in moving object detection. Cylindrical panoramic image also provides a natural way for human perception in a friendly human-computer interface. Fig. 4.3 shows a cylindrical images of the sample image in Fig. 4.1.

2. Change Detection

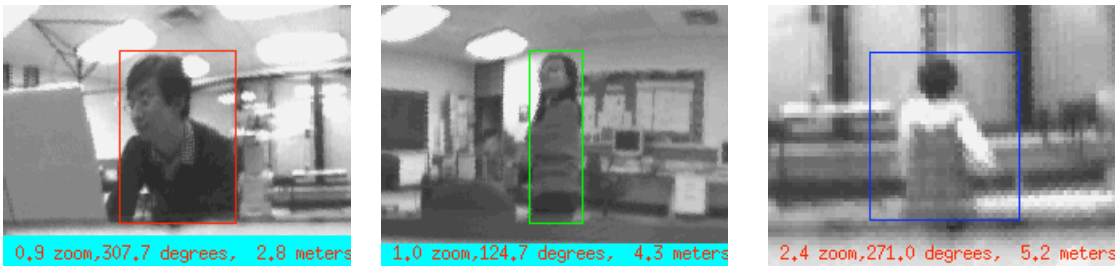
Given a stationary camera, moving objects can be detected and maybe well extracted by subtracting a background image from a current image. However, during a long monitoring period, illumination of the background may change, either gradually or abruptly. So frame difference is calculated to detect any change between the current image and the previous image. Note that the difference of two successive images is basically the difference of the images of an object instead the image of that object itself.

However, the difference image of the successive frames do tell us the motion of the object; there is very small inter-frame change in the background part. So we use this difference image to judge if a region extracted from the subtraction of the current frame from the background image is really a moving object instead of the dynamic change of the background.

(1) 330c.tif: Current Cylindrical image, with bounding rectangles of moving objects superimposed

(2) 330b.tif: Background image

(3) 330o.tif: object image of three people



(4) 330z0.tif, 330z1.tif, 330z2.tif: Zoom perspective images

Fig. 6.1 Multiple object segmentation and virtual zooming

3. Background Updating and Estimation

The background-updating algorithm has two modes: *permanent mode* and *instantaneous mode*. In the *permanent mode*, an initial background is generated at the beginning of the object detection process by estimating the median value of each pixel in many frames (e.g., 30 frame). A good background image can be generate if there is no moving object in a scene during the background initialization, or any pixel in the images are not occupied by objects longer than half of that time period. Then the background image remains unchanged until a "abrupt" change of the illumination (e.g. turn on/off the light) occurs, or the old background image causes many false alarm. Then a new background is generated.

In the *instantaneous mode*, moving object detection begins in the very first frame, and an initial background image is the first frame when the detection starts. Then, the background image is updated pixel-wisely by a weighted average of the existing background pixel and the current image only in those non-object regions. Note that object regions are separated by the following object extraction step. Fig. 6.1(1) shows one frame in a detection example where 3 people walked in a room. The frame rate for multiple object detection is about 5 Hz in a Pentium 300M Hz PC for 1080*162 panoramic images. Fig. 6.1 (2) shows the current background image that was generated for the first 24 frames when all the three peoples were walking around, and then was updated for each processing frame, i.e. 5 times per second.

4. Object Extraction.

This step separates the entire gray-level image of each isolated moving object from the background. As mentioned above an object image $O_i(x,y)$ is a function of the current image $C_i(x,y)$, previous image $P_i(x,y)$, and background image $B_i(x,y)$. First a difference image $D_i(x,y)$ and a subtraction image $S_i(x,y)$ are computed

$$D_i(x,y) = |C_i(x,y) - P_i(x,y)| \quad (6\text{-}1)$$

$$S_i(x,y) = |C_i(x,y) - B_i(x,y)| \quad (6\text{-}2)$$

Then a connected- and close-region grouping algorithm is used to find the region of each objects in an binarized subtraction image

$B_i(x,y) = 1$ if $S_i(x,y) >= minD$ otherwise $B_i(x,y) = 0$ (6-3)

For each Region R_j (j=0,…), the system checks if there is enough change in the difference image $D_i(x,y)$; if not, then delete that region. Then a set of moving object regions MOR_j (j=0,…) is extracted. The object image is the image that copies the original intensity value from the current image only for those regions with moving objects. Fig. 6.1(3) shows the object image of processing the image in Fig. 6.1(1). Note that three people are almost completely extracted from the background; and the boundaries of the people are smoothly along the contours of their bodies.

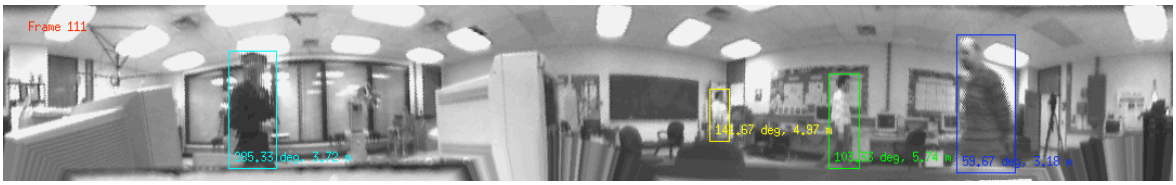
5. **Virtual Gazing of Multiple Moving Objects.**

For each moving object, a perspective view image is generated where the object is kept in the center of the window. All the views have suitable zoom factors to ensure the certain sizes of objects (Fig. 6.1(4)).

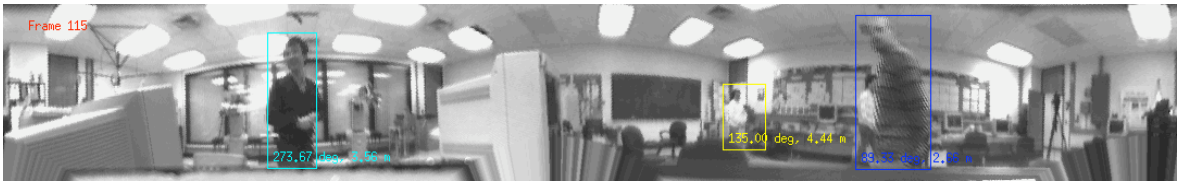
6. Object Tracking

In the current implementation, multiple objects are tracked based on such features as size, aspect ratio and position of each object. Fig. 6.2 and 6.3 depicts multiple moving human object detection and tracking from demonstrations at a DARPA ASC site visit. Multiple moving objects (4 people) were detected in real-time while moving around in the scene in an unconstrained manner and the panoramic sensor is stationary. A background image is generated automatically by tracking dynamic objects across multiple frames. The number of frame depends on the number of moving objects in the scene.

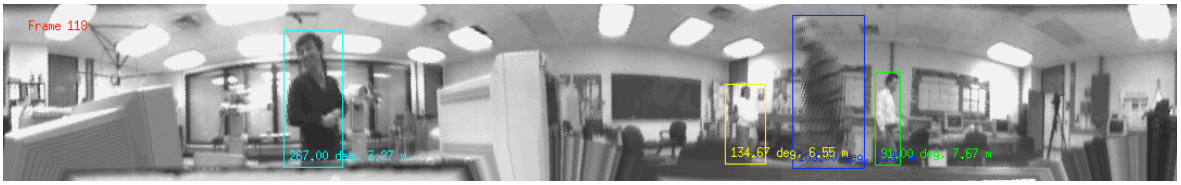
The four moving objects are shown in the un-warped cylindrical image of Fig. 6.2, a more natural panoramic representation for user interpretation. Each of the four people were completely extracted from the complex background as depicted by the bounding rectangle, direction, and distance to each object. The system tracks each object through the image sequence, even if there are overlap and occlusion between two people. The dynamic track, represented as a small circle and icon (elliptic head and body) for the last 30 frames of each person is shown in Fig. 6.3 in different colors. The final object image is depicted at the end of the corresponding track. Notice that the humans reversed directions, and that overlap and occlusion were successfully handled (see the blue and the green tracks). The system can report a alarm signal when more than half of a current image changes due to self-motion, change in the environment, illumination, and sensor failure, while refreshing the background accordingly.



(a) Before meet (s111c.tif) current image, object regions labeled



(b) Meet (s115c.tif) : current image, object regions labeled



(c) After meet (s118c.tif) : current image, object regions labeled

Fig. 6.2. Object tracking: Cylindrical images with bounding rectangles of moving objects superimposed

(a) Before meet (s111t.tif) : track image

(b) Meet (s115t.tif): track image (green merges into blue)

(c) After meet (s118t.tif) : track image (green comes out from blue)

Fig. 6.3. Object tracking: three track images, each track is for the last 32 frames

7. Synopsis

Another way to visually represent the dynamic event of a moving people is a video *synopsis*, which is a single image with the still background and all the images of the moving people. Fig. 6.4 shows one of the examples, where each image of a person is pasted on the background if it is not overlapped with the previous one. A video synopsis gives us a strong feeling of the person's dynamic existence in the environment.

(a) 311c.tif: Current image with the bounding rectangle of a person superimposed

(b) 311b.tif : Background image after updating

(c) 311s.tif : Synopsis of the moving person

Fig. 6.4. Synopsis: showing activity in one image

VII. Future Works

Future work includes the following research topics: cooperative and panoramic vision, appearance-based panoramic vision, and the moving object tracking in a crowded environment.

7.1 **Cooperative and panoramic vision**

In Section 4, we have proposed an approach to construct a virtual stereo vision system by the cooperation of the two platforms. Critical issues include:

1. Detecting the geometrical features in each platform and use them in "mutual calibration".
2. Finding the correspondence of an object in the images of two cameras.
3. The communication burden between the two platforms.
4. Using a zoom camera to identify the moving objects (people, fire, screen, curtain,…)
5. Moving object tracking in a crowded environment where occlusion occurs.

7.2. Appearance-based panoramic vision

Appearance-based panoramic vision (Zhu98) is a combination the appearance-based approach (Murse95) with the panoramic imaging. The basic idea is to build the relation between each sampled panoramic image and the location. A PAL image is represented in the eigen-space by using Principal Component Analysis (PCA) approach. There are several advantages.

1) No image segmentation and feature detection is needed.

2) The panoramic image is rotation-invariant so that we do not need to worry about the view angle problems as in the common appearance-based approach.

3) The representation is the manifold vectors over x-y grid. For simplicity, it is just a Look-up Table with (x,y) indices.

4) Intermediate view can be generated by just interpolation between the sampled images. So the localization problem is transformed as a matching and interpolation problem.

5) Realistic scene image sequences may be synthesized using this representation.

However there are some limitations. In order that the robot works, a learning process should be carried out in a given environment; it may be data extensive; the algorithm requires that the robot know it absolute orientation, otherwise the algorithm should match each rotating version of a panoramic image.

Research Issues include:

1) Image-based environment modeling by PCA learning

2) Optimal view planning by eigen-vector comparison

3) Localization by eigen-space matching

4) Image synthesis for virtual environment

Acknowledgments

This work is supported by DARPA Active Software Composition project under contract No. F30602-97-2-0032.

References

1. Aloimonos, Y. (ed.), "Active perception - Advances in Computer Vision," Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.
2. Baker S and Nayar S K, A theory of catadioptric image formation, In Proceedings of the 6th International Conference on Computer Vision, IEEE, India, January 1998.
3. Brill, F. Z., T. J. Olson and C. Tserng, "Event Recognition and Reliability Improvements for the Autonomous Video Surveillance Systems," In Proceedings of DARPA Image Understanding Workshop, volume 1, pages 267- 284, November 1998.
4. Collins, R., Jaynes, C., Cheng, Y.Q., Wang, X., Stolle, F., Riseman, E., Hanson, A., "The Ascender System Automated Site Modeling from Multiple Aerial Images", Computer Vision and Image Understanding Special Issue on Building Detection and Reconstruction from Aerial Images, 1998.
5. DARPA Image Understanding Workshop Proceedings, VSAM- Video Surveillance and Monitoring Session, Monterey, November 1998
6. Greguss P, Panoramic imaging block for three-dimensional space, U.S. Patent 4,566,763 (28 Jan, 1986)
7. Haritaoglu, I., D. Harwood and L. Davis, "W4S: A Real-time System for Detection and Tracking People in 2.5D," In Proceedings of ECCV, 1998.
8. Hoepfner K., A. R. Hanson, E. M. Riseman, " Recovery of Building Structure from SAR and IFSAR Images," In Proceedings of DARPA Image Understanding Workshop, volume 1, pages 559-563, November 1998.
9. Hong J, et al, Image based homing, Proc. IEEE ICRA, May 1991.

10. Kumar, R., and Hanson, A., "Model Extension and Refinement Using Pose Recovery Techniques", Journal of Robotic Systems, 9(6), 1992, pp.753-771.

11. Lipton, A. J., H. Fujiyoshi, R. S. Patil, "Moving Target Classification and Tracking from real-time Video," In Proceedings of DARPA Image Understanding Workshop, volume 1, pages 129- 136, November 1998.

12. Murase H, Nayar S K, Visual learning and recognition of 3-D objects from appearance. Int. J. Computer Vision, vol. 14, 1995: 5-24.

13. Nayar, S K, Baker S, Catadioptric image formation, Prof. DARPA Image Understanding Workshop, May 1997:1431-1437

14. Nelwa V, A true omnidirectional viewer, Technical Report, Bell Lab, Holmdel, NJ, Feb, 1996

15. Papageorogiou, C., T. Evgeniou, and T. Poggio, " A Trainable Object Detection System," In Proceedings of DARPA Image Understanding Workshop, volume 2, pages 1019-1024, November 1998.

16. Pentland, A., A. Azarbayjani, N. Oliver and M. Brand, " Real-time 3-D Tracking and Classification of Human Behavior," In Proceedings of DARPA Image Understanding Workshop, volume 1, pages 193-200, May 1997.

17. Powell I, Panoramic lens, Applied Optics, vol. 33, no 31, Nov 1994:7356-7361

18. Ravela, S., Manmatha, R., and Riseman, E., "Retrieval From Image Databases Using Scale-Space Matching", ECCV '96, Cambridge, England, April 14-18, 1996.

19. Renault P., O.D. Faugeras and T. Vieville, " Continuos multi-image preprocessing for Euclidean reconstruction," INRIA Tech Report 3482, September, 1998.

20. Riseman, E.A. Hanson, J.R. Beveridge, R. Kumar, H. Sawhney, "Landmark-Based Navigation and the Acquisition of Environmental Models", in Visual Navigation: From Biological Systems to Unmanned Ground Vehicles , Yiannis Aloimonos (Ed.), Lawrence Erlbaum Associates, Inc., Chapter 11, pp. 317-374, 1997.

21. Sawhney, H., and Hanson, A., "Trackability as a Cue for Potential Obstacle Identification and 3D Description", International Journal of Computer Vision, 11:3, pp. 237-265, 1993.

22. Yagi Y, Kawato S, Panoramic scene analysis with conic projection, Prof. IROS, 1990

23. Yamazawa, K, Yagi, Y, & Yachida M, Omnidirectional imaging with hyperboloidal projections, Prof. IROS, 1993.

24. Zhu Z, Xi H, Xu G, Combining rotation-invariance images and neural networks for road scene understanding, *Proc. IEEE International Conference on Neural Network*, pp. 1732-1737, 1996

25. Zhu Z, Yang S, Xu G, Lin X, Shi D, Fast road classification and orientation estimation using omni-view images and neural networks, *IEEE Trans Image Processing (special issue on applications of neural networks to image processing)*, Vol7, No.8, August 1998: pp. 182-1197.