Panoramic Virtual Stereo Vision of Cooperative Mobile Robots for Localizing 3D Moving Objects*

Zhigang Zhu, K. Deepak Rajasekar
Edward M. Riseman, Allen R. Hanson
Computer Vision Lab, Computer Science Department
University of Massachusetts at Amherst, Amherst, MA 01003
{zhu | deepak | riseman | hanson}@cs.umass.edu

Abstract

Flexible, reconfigurable vision systems can provide an extremely rich sensing modality for sophisticated multiple robot platforms. We propose a cooperative and adaptive approach of panoramic vision to the problem of finding and protecting humans by a robot team in an emergency circumstance (e.g. a rescue in an office building). A panoramic virtual stereo vision method is proposed for this cooperative approach, which features omni-directional visual sensors, cooperative mobile platforms, selected 3D matching, and real-time moving object (people) detection and tracking. The problems of dynamic self-calibration and robust 3D estimation of moving objects are discussed. A careful error analysis of the panoramic stereo triangulation is presented in order to derive rules for optimal view planning. Experimental results are given for detecting and localizing multiple moving objects using two cooperative robot platforms.

Keywords: *omnidirectional vision, multiple mobile robots, moving object extraction, cooperative stereo*

1. INTRODUCTION

Flexible, reconfigurable vision systems can provide an extremely rich sensing modality for sophisticated robot platforms. We propose a cooperative and adaptive approach to the problem of finding and protecting humans in emergency circumstances, for example, during a fire in an office building. 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 a 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 [e.g., 1-3] motivated most recently by the DARPA VSAM effort. On the other hand, different kinds of omni-directional imaging sensors have been designed [4-7], and a systematic theoretical analysis of omni-directional sensors has been given [4]. Omnidirectional vision has become quite popular with many vision approaches for robot navigation [6,8], stereo reconstruction [9,10] and video surveillance [11,12]. What is truly novel about our approach is the ability to compose cooperative sensing strategies across the distributed panoramic sensors of a robot team to synthesize robust "virtual" stereo sensors for human detection and tracking.

The idea of distributing sensors and cooperation across different robots stems from the requirements of potentially limited (sensor) resources for a large robot team. Nevertheless, the advantages of cooperative vision arise from more than this compromise. 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. However, there are significant time delays involved in moving the camera to another position in the room as well as the difficulty of dynamic calibration.

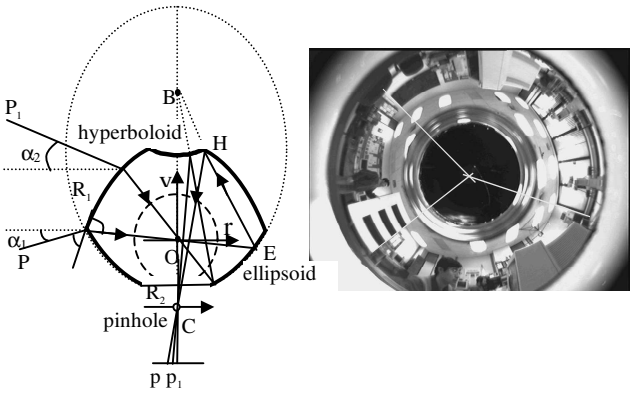
In this paper, we focus on cooperative behavior involving cameras that are aware of each other, residing on different mobile platforms, to compose a virtual stereo sensor with a flexible baseline. In this model, the sensor geometry can be controlled to manage the precision of the resulting virtual sensor. This cooperative stereo vision strategy is particularly effective with a pair of mobile panoramic sensors that have the potential of almost always seeing each other. Once calibrated by "looking" at each other, they can view the environment to estimate the 3D structure of the scene. We will discuss the following issues: 1) modeling and image unwarping of an omnidirectional vision system, 2) dynamic self-calibration among the two cameras on two separate mobile robots, which forms the dynamic "virtual" stereo sensor, 3) view planning by taking advantage of the current context and goals, and 4) detection and 3D localization of moving objects by two cooperative panoramic cameras, and the correspondence of objects between two views, given the possibly large perspective distortion.

* This work was supported by AFRL/IFTD under contract numbers F30602-97-2-0032 (SAFER), and DARPA/ITO DABT63-99-1-0022 (SDR Multi-Robot), and by the Army Research Office under grant number DAAD19-99-1-0016.

2. PANORAMIC IMAGING GEOMETRY

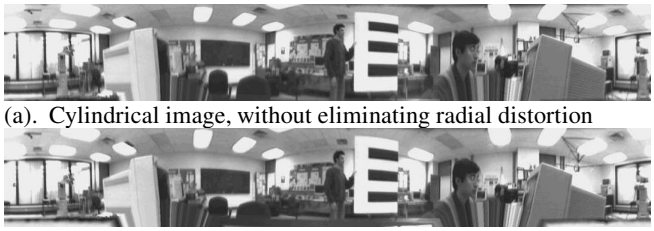
2.1. Modeling the PAL Sensor

In this paper we use the panoramic annular lens (PAL) camera system designed by Pal Greguss [5] for its compactness and view angles. It can capture its surroundings with a field of view (FOV) of 360-degrees horizontally and -15 ~ +20 degrees vertically. We have noticed that many other omnidirectional cameras have their vertical FOV either entirely above or below the horizon. In a robotic application, a vertical viewing angle that spans the horizon is preferred. The PAL-3802 system that we are using includes a compact 40-mm diameter PAL glass block and a built-in 16-mm collector lens with a "C" mount. The geometry of the PAL imaging system is somewhat complex since there are two reflections and two refractions. Fortunately, we can obtain a rather elegant geometry of a single effective viewpoint under perspective projection [5, 13] given that (Fig. 1a): (1) the concave circular mirror (E) is ellipsoidal and the convex top mirror (H) is hyperboloidal; (2) the long axis of the ellipsoidal mirror is aligned with the axis of the hyperboloidal mirror and the optical axis of the camera (C); and (3) a locus (B) of the hyperboloidal mirror coincides with one locus of the ellipsoidal mirror, and the other locus coincides with the nodal point (C) of the real camera. Thus, the single viewpoint of the "panoramic virtual camera" is right at the second locus (O) of the ellipsoidal mirror.



(a) PAL model (b) PAL image (768*576)

Fig. 1. Geometric model of the PAL sensor



(a). Cylindrical image, without eliminating radial distortion

(b). Cylindrical image, after eliminating radial distortion

Fig. 2. Image unwarping and rectification

The refraction does not add too much complexity; instead, it changes the vertical viewing angle range. The first refraction through the ellipsoidal surface (R_1) changes the vertical viewing range from [0°, 90°) to [α_1, +α_2], where $\alpha_1<0^\circ<\alpha_2<90^\circ$, which is often desired for panoramic imagery in robotics applications. For the PAL-3802 camera system we have approximately α_1=-15° and α_2=20°. The second refraction through the planar surface (R_2) only moves the point of convergence of rays that are reflected from the top mirror up some distance. In conclusion, a ray from a 3D point P is first refracted by an ellipsoidal surface R_1, and passes through one of the loci (i.e. the viewpoint O of the virtual camera) of the concave mirror E. Next, the ray is reflected by the ellipsoidal mirror E to its second locus B (which is also a locus of the hyperboloidal mirror H). Then it is reflected by the convex mirror H to the second locus of H, i.e. the nodal point C of the real camera. Thus the annular image in the target plane of the real sensor can be viewed as being captured by a "virtual" camera located at viewpoint O. Fig. 1b show an image captured by the PAL sensor.

2.2. Image Unwarping and Rectification

Having the PAL geometry, we have developed an empirical method to unwarp and rectify the PAL image, which consists of two steps:

(1) Center determination and image unwarping - 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. 1b). If more than two such lines are detected in an original PAL image $I(x,y)$, the center point (x_0,y_0) can be determined by their intersection. Then a cylindrical panoramic image $I(\rho,\theta)$ (Fig. 2a) can be generated by a polar transformation

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

(2) Vertical distortion rectification - Distortion exists in the vertical direction of the unwarped cylindrical image (or the radial direction in the original PAL image) due to the non-linear reflection and refraction of the 2nd-order mirror surfaces. Note the unequal widths of the black-white bars on the white board in Fig. 2a, caused by the vertical distortion (the widths are equal in the real world). We use an N-order polynomial to approximate the distortion:

$$v = \sum_{i=0}^{N} a_i \rho^i \qquad (2)$$

where ρ is the vertical coordinate in the original cylindrical image, and v is in the rectified cylindrical image. Fig 2b shows the rectification result using a 2nd-order polynomial approximation. Note the equal intervals of the black bars on the white board.

2.3. Calibration of the virtual cylindrical sensor

After image unwarping and distortion rectification, we have cylindrical images generated from the virtual viewpoint O (Fig. 3). The next thing we will do is to find the viewpoint and the effective focal length of the panoramic "virtual camera". We assume that the sensors and all objects rest in a common planar surface (the floor). The viewpoint of the virtual panoramic camera

corresponds to a horizontal circle in the cylindrical image, the intersection of the cylindrical image and the horizontal plane passing through the viewpoint. The projection in the horizontal direction is a circular projection, and the effective focal length in the horizontal direction (i.e. the radius of the cylindrical image) can be expressed as $F_h = \frac{W_\theta}{2\pi}$, where W_θ is the perimeter (in pixels) of the cylindrical image. The projection in the vertical direction is modeled as a linear perspective projection, so the "focal length" in the vertical direction can be estimated as

$$F_v = D\frac{v_1 - v_2}{H_1 - H_2} \tag{3}$$

where H_1 and H_2 are the heights of two points in a vertical pole (measured from the floor), v_1 and v_2 are the vertical coordinates (with an arbitrary origin) of their image projections, and D is the distance of the pole from the camera (Fig. 3). By moving the same pole from distance D to distance D', we have a new projection v'_1 for the point of height H_1. Then the vertical coordinate (v_0) of the horizon circle and the height (H_0) of the virtual camera can be calculated as

$$v_0 = \frac{Dv_1 - D'v'_1}{D - D'}, \quad H_0 = H_1 - \frac{D(v_1 - v_0)}{F_v} \tag{4}$$

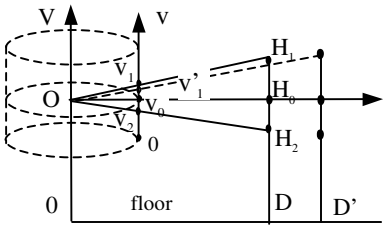


Fig. 3. Intrinsic parameter estimation

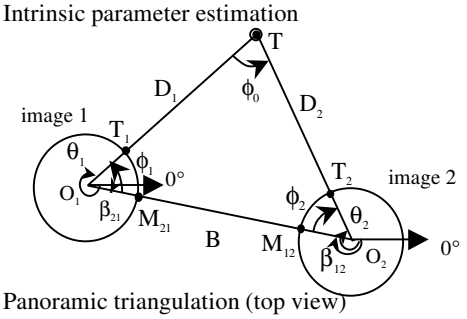


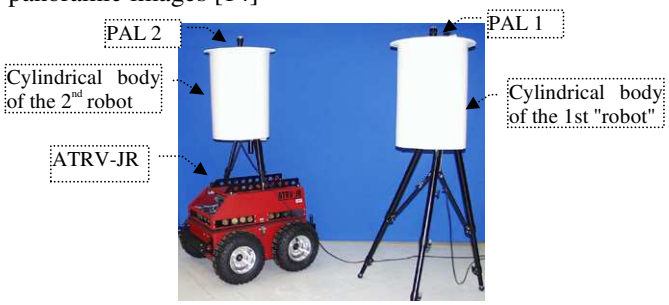
Fig. 4. Panoramic triangulation (top view)

3. PANORAMIC VIRTUAL STEREO GEOMETRY

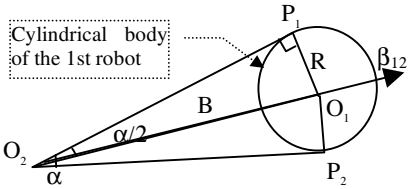
Assume we have two panoramic cameras with the same parameters, both of which are subject to planar motion on the floor and are at the same heights above the floor. Suppose that in Fig. 4, O_1 and O_2 are the viewpoints of the two cameras and they can be localized by each other in the panoramic images as M_{21} and M_{12}. B is the baseline (i.e. distance O_1O_2) between them. The projection of a target point T is represented by T_1 and T_2 in the two panoramic images. Then a triangle O_1O_2T can be formed. By defining an arbitrary starting orientation for each cylindrical image, angles ϕ_1, ϕ_2 (and ϕ_0) of the triangle can be calculated from the following four bearing angles: θ_1 and θ_2, the bearings of the target in image 1 and image 2 respectively, β_{12} and β_{21}, the bearing angles of camera 1 in image 2 , and camera 2 in image 1 respectively. Therefore the distances from the two cameras to the target can be calculated as

$$D_1 = B\frac{\sin\phi_2}{\sin(\phi_1 + \phi_2)}, \quad D_2 = B\frac{\sin\phi_1}{\sin(\phi_1 + \phi_2)} \tag{5}$$

A numerical analysis of the triangulation error will be given in Section 5. In the extreme case, if the target is aligned with the baseline O_1O_2, the triangulation relation is invalid. However, we can still estimate the 3D location of the target by using the size-ratio of the target in the two panoramic images [14]



(a) Setup of the panoramic virtual stereo



(b) Find orientation and distance by a cylinder (top view)

Fig. 5. Dynamic calibration by cylinders

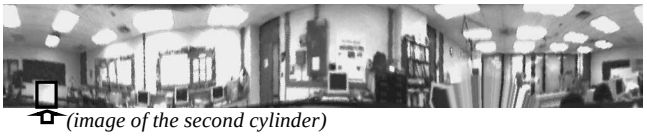
3.1. Dynamic calibration

In order to estimate the distance of a target, we need first to estimate the baseline (B) and the orientation angles (β_{12} and β_{21}) of the two panoramic cameras by a dynamic calibration procedure. We have proposed several practical approaches for this purpose [13]. The basic idea is to make the detection and computation robust and simple. 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, Fig. 5a), which can be easily seen and extracted in the image of the other robot's camera. We assume that the rotation axis (i.e. OV in Fig. 3) of each panoramic image 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, e.g., from the image of camera 2 we have

$$B = R/\sin(\frac{\alpha}{2}) \tag{6}$$

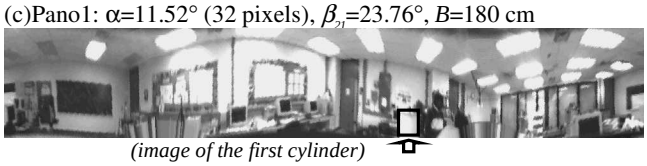
where α is the angle between two occluding projection rays measured in the image of camera 2, and R is the radius of the 1st cylindrical body (Fig. 5b). The orientation angle (β_{12}) of O_2O_1 is simply the average of the bearings of two occluding boundary points P_1 and P_2. We can do the

same in the image of camera 1. Fig.6 show a calibration result. The cylindrical body of each robot is detected and measured in the panoramic image of the other camera. In the experiment, the perimeter of the cylindrical image is 1000 pixels, so the angular resolution χ is 360/1000=0.36° per pixel (or χ=6.28 mrad/pixel). The radius of the cylindrical body of each robot is R=18 cm.



(image of the second cylinder)

(c)Pano1: α=11.52° (32 pixels), β_{21}=23.76°, B=180 cm

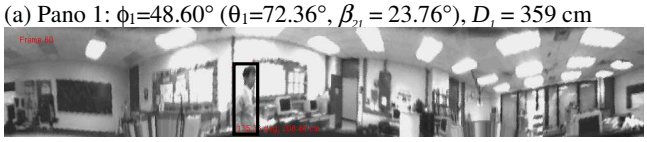


(image of the first cylinder)

(d) Pano2: α=11.52° (32 pixels), β_{12} = 227.88°, B = 180 cm

Fig. 6. A dynamic calibration example

(a) Pano 1: ϕ_1=48.60° (θ_1=72.36°, β_{21} = 23.76°), D_1 = 359 cm

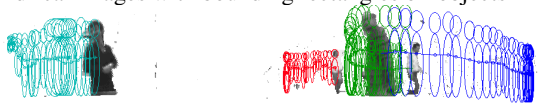


(b) Pano 2: ϕ_2=92.52° (θ_2=135.36°, β_{12} = 227.88°), D_2 = 208 cm

Fig. 7. 3D estimation by triangulation

Fig. 7 shows the experimental results of 3D estimation of a target using the calibration result in Fig. 6. The relative error is about 5% of the distance, which is caused by the errors in dynamic calibration and the bearing angle measurements from two different views of the target (Section 5). Note that a 1-pixel error in the width of the robot cylinder in a panoramic image will introduce about 5-cm error in calculating the baseline in the configuration of Fig. 6, which is a function of the resolution of the panoramic image and the size of the cylindrical robot (Eq. (14)).

(a) Cylindrical images with bounding rectangles of objects



(b) Object tracks, each track is for the last 32 frames

Fig. 8. Moving object detection and tracking

4. SELECTED 3D MATCHING APPROACH

Since our primary goal is to detect and to track moving targets (humans) in 3D, the panoramic virtual stereo method is applied to each object (region) as a whole that has already been extracted from the two panoramic images. A fast moving object extraction and tracking algorithm has been developed, which integrates frame difference, background subtraction, background updating and quasi-connected region grouping techniques [13]. Fig. 8 depicts a multiple moving human object detection and tracking procedure. Multiple moving objects (4 people) were detected in real-time while moving around in the scene in an unconstrained manner; the panoramic sensor is stationary. Each of the four people was completely extracted from the complex background as depicted by an object image and its contour. The dynamic track, represented as an icon (elliptic head and body) for the last 30 frames of each person is shown. The final object image is depicted at the end of the corresponding track. The frame rate for multiple object detection and tracking is about 5 Hz in a Pentium 300M Hz PC for 1080*162 panoramic images.

Since the appearance of an object will vary significantly depending on viewpoint, the information only from 2D images will produce ambiguity in object match. Hence, we explore the 3D-related measurements as well as 2D features. We will explain our 3D matching algorithm assuming that we have the centroid and the bounding rectangle of each object region. Extension to the contours of objects is straightforward. From each panoramic image, a set of objects (regions) is extracted, which is represented by a set

$$\mathbf{T}^{(k)} = \{T_i^{(k)} = (I_i^{(k)}, w_i^{(k)}, h_i^{(k)}, \theta_i^{(k)}, v_i^{(k)}), i = 1,..., N_k\} \quad (7)$$

where k (1 or 2) is the number of cameras, $I_i^{(k)}, w_i^{(k)}, h_i^{(k)}, \theta_i^{(k)}, v_i^{(k)}$ are the photometric feature, the width and height of the image region, bearing angle (in degree) of the region's centroid of the target i in camera k, and the vertical coordinate of the top of the image region. For each object i in $T^{(1)}$, we try to match it with every object j in set $T^{(2)}$, and we derive the following measurements for each possible match:

(1). *Ray convergence* - A meaningful match must satisfy the condition that two rays from the viewpoints to the target images converge. The degree of ray convergence is calculated as a ratio between 0 and 1:

$$r_{rc}(i,j) = \begin{cases} 0: & \text{if rays } O_1T_i^{(1)} \text{ and } O_2T_j^{(2)} \text{ diverge} \\ \phi_0 & \phi_0 \le B / D_{\max} \\ 1 & \text{otherwise} \end{cases} \quad (8)$$

where ϕ_0 is the angle between rays $O_1T_i^{(1)}$ and $O_2T_j^{(2)}$ (see Fig. 9c) , and D_{max} is the maximum distance where an object can be detected by the panoramic stereo system.

(2). *Width consistency* - By calculating the distances $D_i^{(1)}$ and $D_j^{(2)}$ of the "hypothesized" target to viewpoints O_1 and O_2 by matching $T_i^{(1)}$ and $T_j^{(2)}$, the size of the hypothesized target estimated from two images should be very close for the correct match. A width ratio is calculated to account for this width consistency in two images

$$r_{wc}(i,j) = \frac{\min(w_i^{(1)} D_i^{(1)}, w_j^{(2)} D_j^{(2)})}{\max(w_i^{(1)} D_i^{(1)}, w_j^{(2)} D_j^{(2)})} \in [0,1] \quad (9)$$

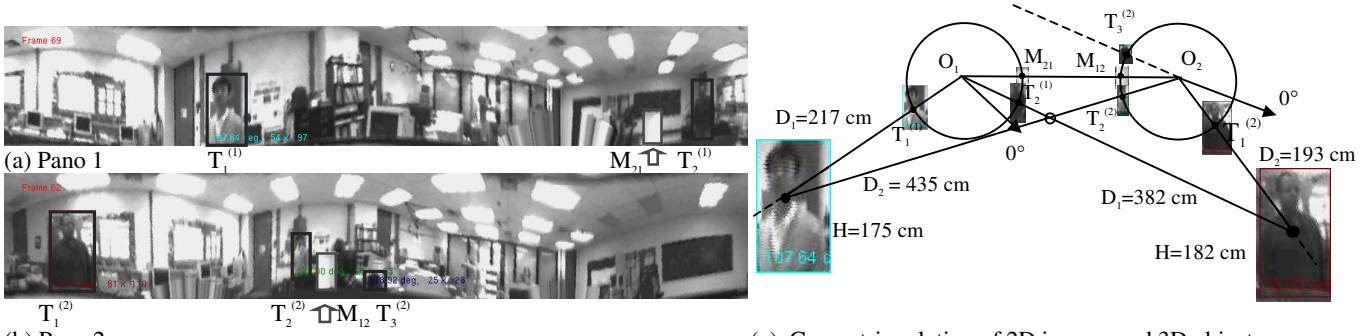


Fig. 9. 3D estimation of multiple objects (M_{ij} is the image of robot *i* in panoramic image *j*, which is used for calibration)

(3). *Height consistency* - First we calculate the height of the top of the hypothesized object from both images, as $H_i^{(1)}$ and $H_j^{(2)}$ (from Eq.(4)), which should be the same. Then a height ratio accounting for the degree of height consistency can be calculated as

$$r_{hc}(i,j) = \frac{\min(H_i^{(1)}, H_j^{(2)})}{\max(H_i^{(1)}, H_j^{(2)})} \in [0,1] \qquad (10)$$

(1). *Photometric similarity* - Assuming that the photometric feature is a scalar value, then the following photometric ratio can be derived

$$r_{ps}(i,j) = \frac{\min(I_i^{(1)}, I_j^{(2)})}{\max(I_i^{(1)}, I_j^{(2)})} \in [0,1] \qquad (11)$$

In the current algorithm, we use a scale value - the median intensity of the image region of an object as the photometric feature. More complex features can be used to calculate the photometric ratio between two image regions.

Therefore, a *match measurement matrix* $M = [r(i,j)]_{N1xN2}$ can be constructed where

$$r(i,j) = r_{rc}(i,j) r_{wc}(i,j) r_{hc}(i,j) r_{ps}(i,j) \in [0,1] \qquad (12)$$

Note that in the match measurement features, all match constraints (except the last one) contain 3D information, namely orientation constraints, width constraints and height constraints. Now the task is to find the best match for each object of the first image, based on the match matrix. The current implementation is based on an exclusive "greedy" match algorithm as follows:

Step 0: Initialization: m = 0; $M^{(0)} = M$;

Step 1: Find the maximum value in matrix $M^{(m)}$, and its corresponding indices (i, j). Then object i in the first image matches object j in the second image;

Step 2. Delete row i and column j in the matrix $M^{(m)}$ to form a sub-matrix $M^{(m+1)}$; assign m = m+1;

Step 3. Go to step 1 until no rows or columns left in $M^{(m)}$.

Fig. 9 gives an example where two moving persons were detected in panoramic image 1 (Pano 1) and three persons were detected in Pano 2. Fig. 9c visually shows the geometrical relation between images and 3D objects, and Fig. 10 gives the match measurement matrix $[r(i,j)]$ as well as all the four measures $r_{ps}(i,j), r_{rc}(i,j), r_{wc}(i,j)$ and $r_{hc}(i,j)$. By applying the "greedy" match algorithm, the correct final matches are $(T_1^{(1)}, T_2^{(2)})$ and $(T_2^{(1)}, T_1^{(2)})$. The estimates of distances and heights of the two people are labeled in Fig. 9c. In this experiment, the panoramic image parameters are F_h =159 pixels, F_v =258 pixels, v_0= 58 pixels, H_0=138 cm, β_{21} = 313.56°, β_{21} = 156.60°, and B = 239 cm. A modification of this algorithm is needed in order to find matches for some of the objects that are merged together, missed in detection and/or occluded.

	$T_1^{(2)}$	$T_2^{(2)}$	$T_3^{(2)}$
$T_1^{(1)}$	0.0 (0.46, 0.0, *, *)	**0.72** **(0.86,1.0,0.93,0.91)**	0.0 (0.40,0.0, *, *)
$T_2^{(1)}$	**0.988** **(1.0,1.0,0.99,0.99)**	0.32 (0.54,1.0,0.62,0.95)	0.0 (0.86,0.0, *, *)

Fig. 10. Match measurement matrix ("*" means "don't care" when two rays do not converge, .i.e. $r_{rc} = 0.0$)

5. ERROR ANALYSIS AND VIEW PLANNING

Here we derive an analysis of the error in estimating distance D_1 from the first camera to the target. The upper bound of the estimated triangulation error can be computed by using partial differentials of Eq. (5) as

$$\partial D_1 = \frac{D_1}{B}\partial B + D_1\left|\cot(\phi_1+\phi_2)\right|\partial\phi_1 + \frac{D_2}{\sin(\phi_1+\phi_2)}\partial\phi_2 \qquad (13)$$

where ∂B is the error in computing the baseline B, and $\partial\phi_1$ and $\partial\phi_2$ are the errors in estimating the angles ϕ_1 and ϕ_2 from the two panoramic images. Analyzing Eq. (5), we have found that the distance error comes from three separate error sources: calibration error, matching error and triangulation error, which will be discussed below.

5.1. Calibration error

Dynamic calibration estimates the baseline B, and the bearing angles β_{12} and β_{21} of the two cameras. The error in estimating the baseline (by Eq. (6)) can be derived as

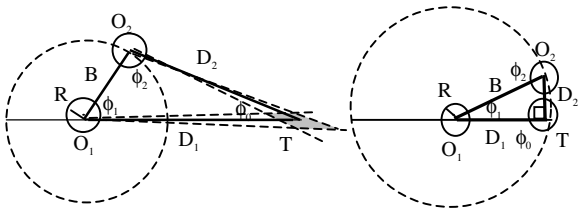
$$\partial B = \frac{B\sqrt{B^2-R^2}}{2R}\partial\alpha \le \frac{B^2}{2R}\partial\alpha \qquad (14)$$

where R<<B, and $\partial\alpha$ is the error in estimating the angle α in an image. From Eq. (14) we can find that the baseline error ∂B is inversely proportional to the dimension of the cylindrical body for dynamic calibration given the same

angle error $\partial\alpha$. Given the radius R and the angle error, the baseline error is roughly proportional to the square of the baseline itself. The errors in the bearing angles β_{21} and β_{12}, which introduce errors to ϕ_1 and ϕ_2 (Fig. 4), can be roughly modeled as the same as $\partial\alpha$, i.e. $\partial\beta_{21} = \partial\beta_{12} = \partial\alpha$.

5.2. Matching error

Assuming that we want to find the distance of a *given* point T_1 in view 1 by finding its corresponding point T_2 in view 2. In this sense, there will be no error in providing the bearing angle θ_1 in view 1, which implies that the error $\partial\phi_1$ is solely determined by β_{21} via calibration, i.e. $\partial\phi_1 = \partial\alpha$. However, the view difference in O_1 and O_2 will introduce a "matching error" (denoted as $\partial\theta$) in θ_2, the localization of T_1's matching point T_2, which could be a function of the location of the view point O_2 (related to O_1). Thus, $\partial\phi_2 = \partial\alpha + \partial\theta$ is a (complicated) function of the viewpoint location and is generally larger than $\partial\phi_1$.



(a). $B < D_1$ (b) $B > D_1$

Fig. 11. Configuration for best viewing angle (given the baseline)

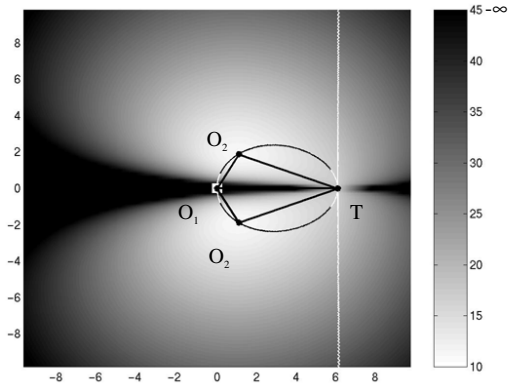


Fig. 12. Error map for distance D_1 when camera O_2 is in different locations of the map by fixing camera O_1 and the target T (D_1 = 34R = 6m, R = 18 cm). Minimum error is $\partial D_1 / D_1 \partial\phi$ =11.2 when B=220cm, $\phi_1 = 62.1°$ (We have two such symmetric locations for O_2). The upper bound of the relative error is $\partial D_1 / D_1$=7.0% when $\partial\phi$ is equivalent to 1 pixel. The labels in the two axes are distances (in meters); the black-white curve shows where the minimum errors can be achieved for viewpoint O_2 on circles around O_1 *(see explanation in the text)*; the error value ($\partial D_1 / D_1 \partial\phi$) is encoded in intensity: see the corresponding bar.

5.3. Triangulation error and view planning

Now we want to find a numerical result of the following problem: for a certain distance D_1 from camera 1 to the target, what is the error distribution for different locations of camera 2, which determines configurations of baselines and angles of the panoramic stereo? Since it is not easy to give a numerical function of the error $\partial\phi_2$ vs. the location O_2, we will use the same measure errors bound in angles, i.e. $\partial\alpha = \partial\phi_1 = \partial\phi_2 \equiv \partial\phi$. Later on we will re-consider this matching error qualitatively. We decompose the analysis into two steps. First, by fixing the baseline, we find the optimal angle ϕ_1. It is equivalent to finding the optimal position of O_2 on a circle of origin O_1 and radius B (Fig. 11). Second, under the optimal angle configuration, we find the optimal baseline B. An additional consideration is that a human has a size comparable to the robots, so the distances between a robot and the target should be at least greater than the dimension of the robot, $2R$. Due to the sign of cotangent in Eq. (13), we mainly have the following two cases (A complete coverage of cases and their mathematical derivation is given in [14]):

Case (1). Given that $B \le D_1$-$2R$ (Fig. 11a, Fig. 12), the best estimation (minimum error) can be achieved when

$$\cos\phi_1 = \frac{3BD_1}{2D_1^2 + B^2} \tag{15}$$

Note that in this case th error in Eq. (13) can be represented as a function of only ϕ_1, and it is increases from the minimum value to ∞ when the angle ϕ_1 changes from the optimal value to 0° and 180° respectively (see Fig. 12 for each "virtual" circle entered at O_1). The minimum error is achieved when $\phi_1 < 90°$, $\phi_2 > 90°$ and $\phi_0 < 90°$. Inserting Eq. (14) and Eq. (15) into Eq. (15) we have

$$\partial D_1 = \left(\frac{D_1\sqrt{B^2 - R^2}}{2R} + \frac{\sqrt{(D_1^2 - B^2)(4D_1^2 - B^2)}}{B}\right)\partial\phi < D_1\left(\frac{B}{2R} + \frac{2D_1}{B}\right)\partial\phi \tag{16}$$

From Eq. (16), it is intuitive that the larger is the baseline, the better the triangulation will be (term 2), however the estimated error in the baseline is also larger (term 1). The global minimum value can be achieved when

$$B = 2\sqrt{D_1 R} \tag{17}$$

which means that (1) more accurate baseline estimation can be obtained given a larger cooperative robotic target (i.e. R), hence the optimal baseline for estimating distance D_1 *can* be larger, and (2) the farther the target is, the larger the baseline *should* be. The error in this global optimal configuration (O_2 in Fig. 12) is

$$\partial D_1^+ = D_1\left(\frac{\sqrt{4D_1R - R^2}}{2R} + \frac{\sqrt{(D_1 - 4R)(D_1 - R)}}{\sqrt{D_1 R}}\right)\partial\phi < 2D_1\sqrt{\frac{D_1}{R}}\partial\phi \tag{18}$$

Case (2) . When $B \ge \sqrt{D_1^2 + 4R^2}$ (Fig. 11b, Fig. 12), the best estimation can be achieved when

$$B = \sqrt{D_1^2 + 4R^2}, \quad \cos\phi_1 = \frac{D_1}{B} \tag{19}$$

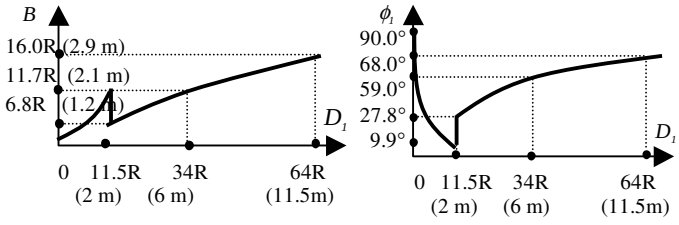
and the error in the optimal configuration is

$$\partial D_1^- = D_1\left(\frac{\sqrt{D_1^2 + 3R^2}}{2R} + \frac{2R}{D_1}\right)\partial\phi \tag{20}$$

Note that in this case, the minimum error is achieved when $\phi_1 < 90°$ (which is the minimum angle by physical constraint of the minimum object distance), $\phi_2 < 90°$ and $\phi_0 = 90°$ (see the vertical white lines in Fig. 12). The optimal baseline in Eq. (19) means the second camera should be as close as possible to the target.

By some tedious mathematics comparing Eq. (18) and Eq. (20) under different D_1, we arrive at the following observation: *If the distance from camera 1 (the main camera) to the target is greater than 11.5 times the radius of the robot, i.e. D_1 > 11.5 R, we have $\partial D_1^+ < \partial D_1^-$, which means that the best configuration is* $B = 2\sqrt{D_1 R}$, $\cos\phi_1 = \frac{3BD_1}{2D_1^2 + B^2}$ *(Eqs. (17) and (15)). Otherwise, we have* $\partial D_1^+ >= \partial D_1^-$, *i.e. the best configuration is* $B = \sqrt{D_1^2 + 4R^2}$, $\cos\phi_1 = \frac{D_1}{B}$ *(Eq. (19)).*

This observation will be used in the optimal view planning. The distance error map under different viewpoint of camera O_2 is given in Fig. 12 in the case of D_1= 34R=6m to verify the above conclusion. The selection of optimal viewing angle and baseline for different distance is shown in Fig. 13. Note that parameters in Fig. 13 are slightly different from those in Fig. 12 because the curves in Fig. 13 are drawn using Eq. (18) and (20) with some approximation and practical consideration. A comparison of the error between the flexible baseline and the fixed baseline, triangulation method and size-ratio method is given in [14], which has shown that the flexible baseline triangulation method is almost always more accurate.



(a) B~D_1 curve (b) ϕ_1~D_1 curve

Fig. 13. Best baselines and angles vs. distance curves (The numbers in the parentheses are given when R = 0.18m)

The best triangulation configuration is derived when all the angular errors ($\partial\alpha$, $\partial\phi_1$, $\partial\phi_2$) are treated as the same and independent to the view configuration of the panoramic stereo. However, as we discussed in Section 5.2, the error $\partial\phi_2$ may be a function of the position of O_2 (given the locations of O_1 and T). A quantitative result can be derived in the same manner as above if the function is known or can be approximated; but here we only give a qualitative analysis The error map in Fig. 12 also shows that there is a relatively large region with errors that are less than twice that of the minimum error. (In Fig. 12 it is the region around the black part of the minimum curve.) The large errors only occur when angle ϕ_0 is very close to 0° and 180°. It implies that a tradeoff can be made between the matching error due to large view difference and the triangulation error due to the small view difference (A further work is needed here). As we did in Section 4, the match is between the centroids of the regions in two panoramic images, thus the 3D estimation gives the distance of a point near the center of the human target. In addition, it is interesting to note that larger view difference can give a better measurement of the dimension of the 3D object (person), which is similar to the volume intersection method.

6. EXPERIMENTAL SYSTEM

In the panoramic stereo vision approach, we face the same problems as in traditional motion stereo: dynamic calibration, feature detection, and matching. However, our goal is not to get a detailed depth map of the surfaces, but rather the location of the moving objects. Thus, the following cooperative strategies can be explored between two robots (and their panoramic sensors) to achieve this goal: monitor-explore working mode, mutual awareness, information sharing and view planning.

In the two-robot scenario for human searching, one of the robots is assigned the role of "monitor" and the other the role of "explorer". The role of the "monitor" is to monitor the movements in the environment, including the motion of the "explorer". One of the reasons that we have a "monitor" is that it is advantageous for a stationary camera (mounted on the "monitor") to detect and extract moving objects. On the other hand, the role of the "explorer" is to follow a moving object of interest and/or find a "better" viewpoint for constructing the virtual stereo geometry with the camera on the "monitor". Mutual awareness of the two robots is important for their dynamic calibration of relative orientations and the distance between the two panoramic cameras. In the current implementation, we have designed a cylindrical body with known radius and color so it is easy for the cooperating systems to detect each other. On the other hand, it is possible to share information between the two panoramic imaging sensors, since they have almost identical geometric and photometric properties. For example, when a certain number of moving objects are detected and extracted by the stationary "monitor", it can pass the information of the object number, geometric and photometric features of each object to the explorer who may be in motion. Thus it makes the explorer easier to track the same objects.

View planning is applied whenever there are difficulties in object detection due to occlusion and 3D estimation. Since both the robots and the targets (humans) are in motion, it is hard to combine multiple 3D estimation along the path of the moving robot to estimate the location of the moving person. Instead, we define the view planning as the process of adjusting the view point of the exploring camera so that best view angles and baseline can be achieved dynamically for the monitoring camera to estimate the distance to the target of interest (see Section 5.3). Note that the explorer is always trying to find a best position in the presence of a target's motion.

In our experimental system, we mounted one panoramic annual lens (PAL) camera on an RWI ATRV-Jr. robot (the "explorer"), and the other PAL camera on a tripod (the "monitor")(Fig. 5a). Two Matrox-Meteor frame grabbers were installed on the ATRV-JR and a desktop PC respectively, both with 333M Hz PII processors. The communication between two platforms is through a client-server model over an Ethernet link (wireless Ethernet communication will be used in the future system). 3D moving object detection and estimation programs run separately on the two machines at about 5 Hz. Only camera and object parameter data (i.e., baseline, bearing angles, sizes, and illuminate features) were transmitted between two platforms so the delay in communication can be ignored at the current processing rate (5Hz). In the current implementation, we assume that the most recent results from both platforms correspond to the events at same time instant. Synchronized image capture is being considered by using the 30-frame buffering capability of the frame grabbers in order to avoid the motion delay of moving objects in two images.

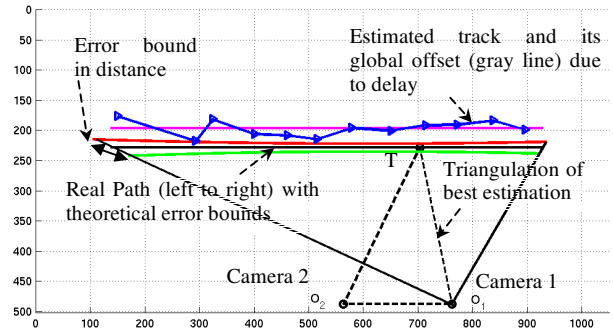


Fig. 14. Panoramic stereo tracking result (axis in cm)

Fig. 14 shows the result from an experiment to evaluate the panoramic stereo's performance of tracking a single person walking along a known path when the two cameras were stationary. The theoretical error bounds were computed assuming that all the angular errors in Eq. (13) and Eq. (14) were equivalent to 1 pixel. The target (T) position where the theoretical best triangulation on this track can be expected is shown in the figure, which is validated by the real experimental result. Even if the localization errors in images may be larger than 1 pixel, the average error of the estimated track is comparable to the theoretical error bounds, taking a global offset into account (The offset of the estimated track from the real path is due to the delay of the processing of the explorer (O_2) working in "telnet" mode). Further experiments on view planning and evaluation is being undertaken.

7. CONCLUSION AND DISCUSSION

This paper has presented a panoramic virtual stereo approach for two cooperative mobile platforms. There are four key features in our approach: (1) omni-directional stereo vision with single viewpoint geometry and a simple camera calibration method, (2) cooperative mobile platforms for mutual dynamic calibration and best view planning, (3) selected 3D matching after object extraction and (4) near real-time performance. The integration of omni-directional vision with mutual awareness and dynamic calibration strategies allows intelligent cooperation between visual agents. Experiments have shown that this approach is encouraging. On-going and future work include the following topics: (1) ***Improvement of the calibration accuracy*** - By integrating a panoramic camera with a pan/tilt/zoom camera, the system can increase the capability (in both viewing angle and image resolution) to detect the cooperative robots as well as the targets. Robust and accurate dynamic calibration is the key issue in the cooperative stereo vision. (2) ***Improvement of 3D matching*** - By using the contours of object images and more sophisticated features, more accurate results can be expected. This is another main factor that affects the robustness and accuracy of the 3D estimation. (3) ***Tracking of 3D moving objects*** - We need to develop more sophisticated algorithms to track the moving objects in the presence of occlusion, and by moving cameras as well as stationary cameras, and to integrate the 3D estimation during tracking.

REFERENCES

[1]. Pentland, A., A. Azarbayjani, N. Oliver and M. Brand, Real-time 3-D Tracking and Classification of Human Behavior, Proc. DARPA IUW, May 1997: 193-200.

[2]. Haritaoglu, I., D. Harwood and L. Davis, W4S: A Real-time System for Detection and Tracking People in 2.5D, ECCV98.

[3]. Lipton, A. J., H. Fujiyoshi, R. S. Patil, Moving Target Classification and Tracking from real-time Video, Proc. DARPA IUW, v 1, Nov.1998, 129- 136.

[4]. Baker, S. and S. K. Nayar, A theory of catadioptric image formation, ICCV98.

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

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

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

[8]. Zhu, Z, S. Yang, G. Xu, X. Lin, D. Shi, Fast road classification and orientation estimation using omni-view images and neural networks, *IEEE Trans Image Processing*, vol 7, no 8, August 1998: 182-1197.

[9]. Ishiguro, H., M. Yamamoto and S. Tsuji, Omni-directional Stereo, IEEE Trans. PAMI, Vol. 14, No.2, 1992: 257-262

[10]. Konolige, K. G., R. C. Bolles, Extra set of eyes, Proc. DARPA IUW, v 1, Nov.1998, 25- 32.

[11]. Boult, T., E., R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, A. Erkan, Frame-Rate omnidirectional surveillance and tracking of camouflaged and occluded targets, Second IEEE Workshop on Visual Surveillance, June 1999: 48-58.

[12]. Ng, K. C., H. Ishiguro, M. Trivedi and T. Sogo, Monitoring dynamically changing environments by ubiquitous vision system, IEEE Workshop Visual Surveillance, 1999: 67-73

[13]. 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.

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