Analysis and Design of Panoramic Stereo Vision Using Equi-Angular Pixel Cameras

Mark Ollis, Herman Herman and Sanjiv Singh

CMU-RI-TR-99-04

The Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

January 1999

This work was sponsored by the Defence Research Establishment, Suffield, Canada.

Abstract

In this report we discuss methods to perform stereo vision using a novel configuration of devices that allow imaging of a very wide field of view (full 360 degrees in the azimuth and up to 120 degrees in elevation). Since the wide field of view produces signficant distortion that varies with viewpoint, we have developed a method to do correlation matching and triangulation for stereo vision that incorporates mirror shape. In this report we evaluate various configurations of cameras and mirrors that could be used to produce stereo imagery, including one that can use a single camera to produce stereo images. Range from panoramic stereo imagery can be used in many applications that require the three dimensional shape of the world. The chief advantages of this method is that it can made cheaply with no moving parts and can provide dense range data. Specifically, this kind of a sensor could be used for telepresence, autonomous navigation by robots, automatic mapping of environments and attitude estimation.

Table of Contents

1. Introduction

Recent research in tele-operated and autonomous systems has shown the utility of imaging that is able to span a very wide field of view. If instead of providing a small conic section of view, a camera could capture an entire half-sphere at once, several advantages can be accrued. Specifically, If the entire environment is visible at the same time, it is not necessary to move the camera to fixate on an object of interest or to perform exploratory camera movements. This also means that it is not necessary to actively counteract the torques resulting from actuator motion. Processing global images of the environment is less likely to be affected by regions of the image that contain poor information. Generally, the wider the field of view, the more robust the image processing will be. For autonomous robots, panoramic imagery allows the use of regions of the image that correspond to the features transverse to the motion of the vehicle for mapping, while forward and rear pointing regions can be used for purposes such as tele-presence and obstacle avoidance.

Today, the alternative method to capture the entire scene in a single image is to use a ultra-wide-angle lens at the cost of weight, expense and optical distortion. A typical 180 degree lens can weigh over a kilogram and cost thousands of dollars. Reflective optics offer a solution to the problem of global imaging. A camera placed below a convex reflective surface can produce a large field of view provided an appropriate mirror shape. Such a configuration (Figure 1) is eminently suited to miniaturization and can be produced inexpensively.

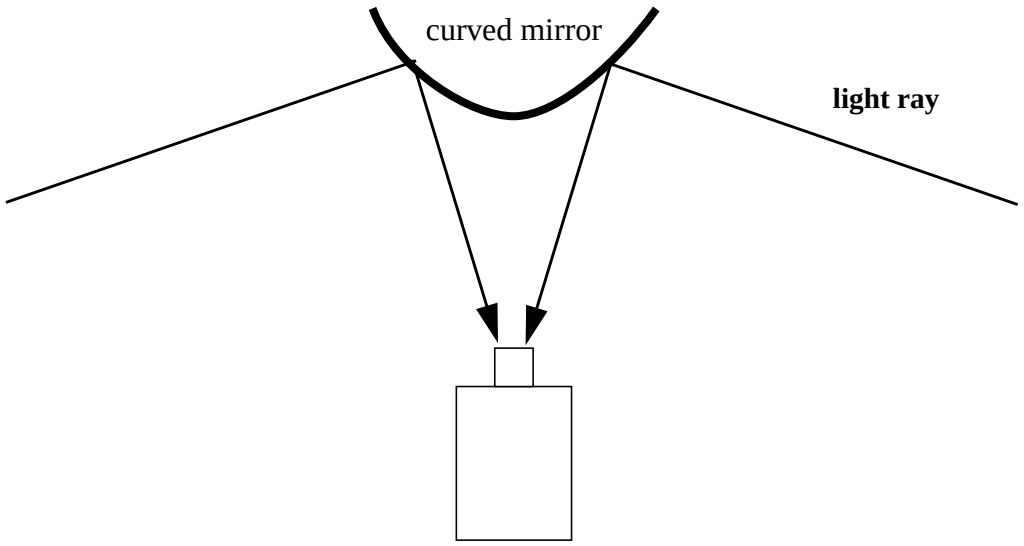


Figure 1 Configuration of a panospheric camera.

With appropriately shaped mirrors the camera can image a full 360 degrees in azimuth, and in some cases up to 120 degs in elevation. An sample image taken from a color camera matched with a curved mirror is shown in Figure 2. For purposes of tele-presence, one common application of this imagery is to unwarp the raw spherical image into a cylindrical one as shown in Figure 3.

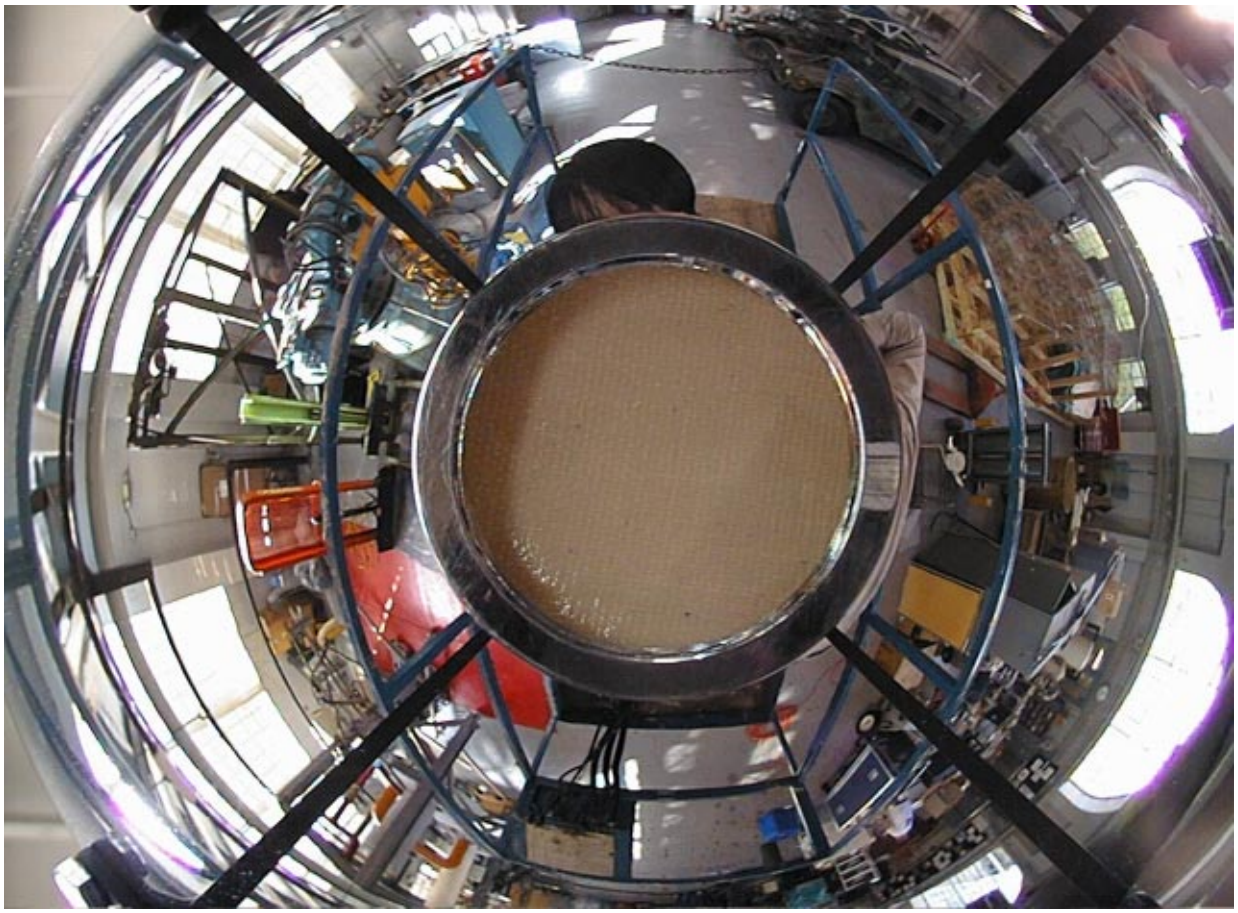


Figure 2 Raw image from panospheric camera

Figure 3 Raw image from Figure 2 displayed as a rectangular image using a projection onto a sphere.

The use of such imagery has distinct advantages: First it is a passive sensor, so power requirements are minimal. Second, it has the potential to be extremely robust, since the sensor is purely solid state and has no moving parts. Third, curved mirrors can be made free of optical distortion that is typically seen in lenses. Third, the large field of view available offers substantial advantages for target tracking, obstacle detection, localization, and tele-navigation.

1.1. Panoramic Stereo Vision

Typically stereo vision has been conducted with small field of view cameras because of lens distortion. The largest field of view that has been used to do stereo is roughly around 110 degrees. Panoramic stereo vision is possible if mirrors are used rather than lenses because of low optical distortion. Not only is panoramic stereo more favorable than narrower field of stereo systems employed for telepresence or autonomy, it offers several advantages for applications that might otherwise use inertial sensor or active ranging. Since such systems can be built inexpensively, it is possible to substitute one panospheric system for many monocular ones. As an example, motion estimation from optical flow becomes well conditioned with panoramic stereo imaging.

In this report we report on the analysis and design of a panospheric stereo vision system. Such a system could provide complete appearance, that is, both the shape and the reflectance properties of the environment. While reflectance information is simple to produce as shown above, depth perception is typically a difficult problem.

We have evaluated a number of configurations of cameras and mirrors to produce dense panoramic range maps. We have done the following:

- Developed a mirror shape that is truly equi-angular. The implication of this is that each pixel in the image spans exactly the same angle. Only approximately equi-angular mirror shapes have been reported in the literature.
- Developed a novel method of correlating features in panoramic images. Correlation is the process of matching features between images.
- Developed a novel method of triangulation that accounts for mirror distortion to calculate range. Triangulation is the process of determining range from an image disparity
- Developed a prototype stereo vision engine that produces range maps from panoramic images.
- Examined 5 configurations or mirrors and cameras and evaluated them on the basis of field of view, range precision and experimental validation with sample simulated (ray-traced) imagery.

1.2. Outline of the report

In Section 2 we discuss different mirror shapes that have been used by other groups and present a mirror shape that has the property that every pixel images an equal angle. In Section 3 we evaluate 5 configurations of panoramic stereo systems. In Section 4 we discuss stereo algorithms and in Section 5 we discuss other uses of panoramic imaging.

2. Panospheric Mirrors

At least three families of mirrors have been used to create panoramic images:

- Spherical: These mirrors have constant curvatures and are easy to manufacture but have no compelling optical advantages. These mirrors were used to provide telepresence for the Atacama mission[2] and the Antarctic meteorite search mission [4].
- Parabolic: Columbia University has developed a panoramic imaging system that uses a parabolic mirror and an orthographic lens to produce a perspective panoramic image. The disadvantage of this system is that it requires orthographic lenses which can be expensive [3].
- Equiangular: Chahl and Srinivasan have proposed a family of equi-angular mirrors that are parameterized by a few parameters [1]. The idea with these mirrors is that each pixel spans an equal angle irrespective of its distance from the center of the image.

Given the geometry shown in Figure 4, the general form of these mirrors is given in (1)

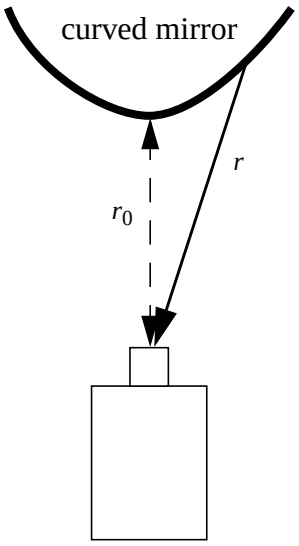


Figure 4 Mirror geometry from Srinivasan and Chahl paper.

$$\cos\left(\theta\frac{1+\alpha}{2}\right) = (r/r_0)^{-(1+\alpha)/2} \quad (1)$$

For different values of α, different mirror shapes produced. Examples are shown in Figure 5. The advantage of these mirrors is that resolution is unchanged when the camera is pitched or yawed. However, since the combination of the mirror/camera is no longer a projective device, stereo algorithms must be modified.

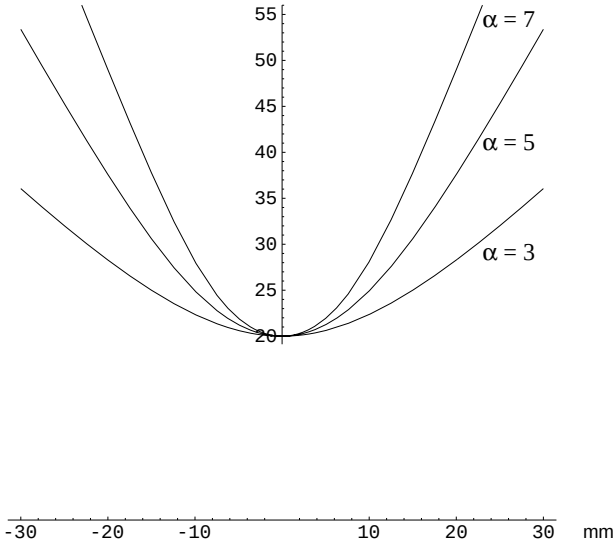


Figure 5 Mirror shape for α = 3, 5 and 7. Arbitrary angular magnification allowed by controlling the α.

Chahl and Srinivasan make a small angle approximation when they claim that each pixel spans an equal angle. They use the following approximation to derive mirror shape:

$$\frac{dr}{d\theta} = r\cot\left(k\theta + \frac{\pi}{2}\right) \qquad k = (-1-\alpha)/2 \tag{2}$$

Since the camera is still a projective device this only works for very small fields of view. Surfaces of mirrors in which each pixel truly corresponds to an equal angle are shapes which satisfy the polar coordinate equation

$$\frac{dr}{d\theta} = r\cot\left(k\tan\theta + \frac{\pi}{2}\right) \tag{3}$$

instead of the approximation used in their paper. The advantage of using (2) is that the surfaces produced have a close-form solution where as (3) must be solved numerically. The difference in the mirror shapes is shown in Figure 6.

For this report we have used the approximated mirror shapes since they allow analytic solutions of several problems which would otherwise need to be estimated numerically. In a real system, however, it is likely that all the calculations will be implemented in lookup tables for speed; thus, for a real system, there is no advantage to using an approximation; it is merely a useful tool for analysis.

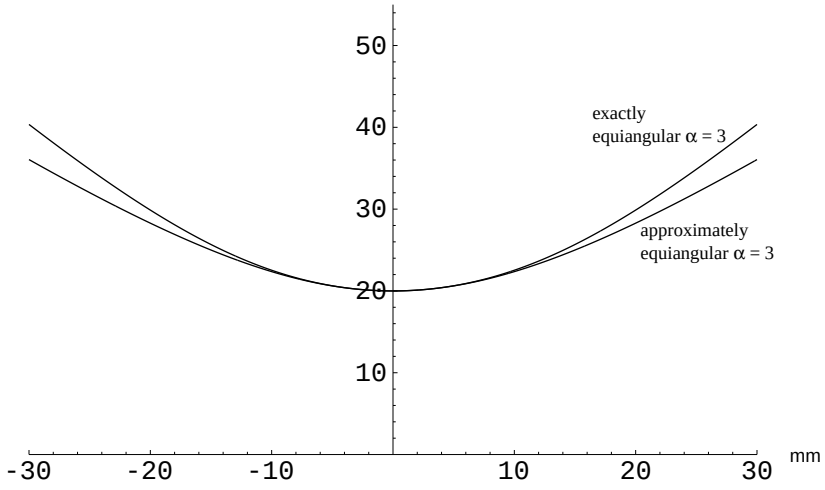


Figure 6 Mirrors ($\alpha = 3$) that provide approximately equal angles for each pixel and exactly equal angles for each pixel.

3. Stereo Configuration Analysis

In this section, we consider five different physical configurations for producing stereo images. These configurations are shown in Figure 7.

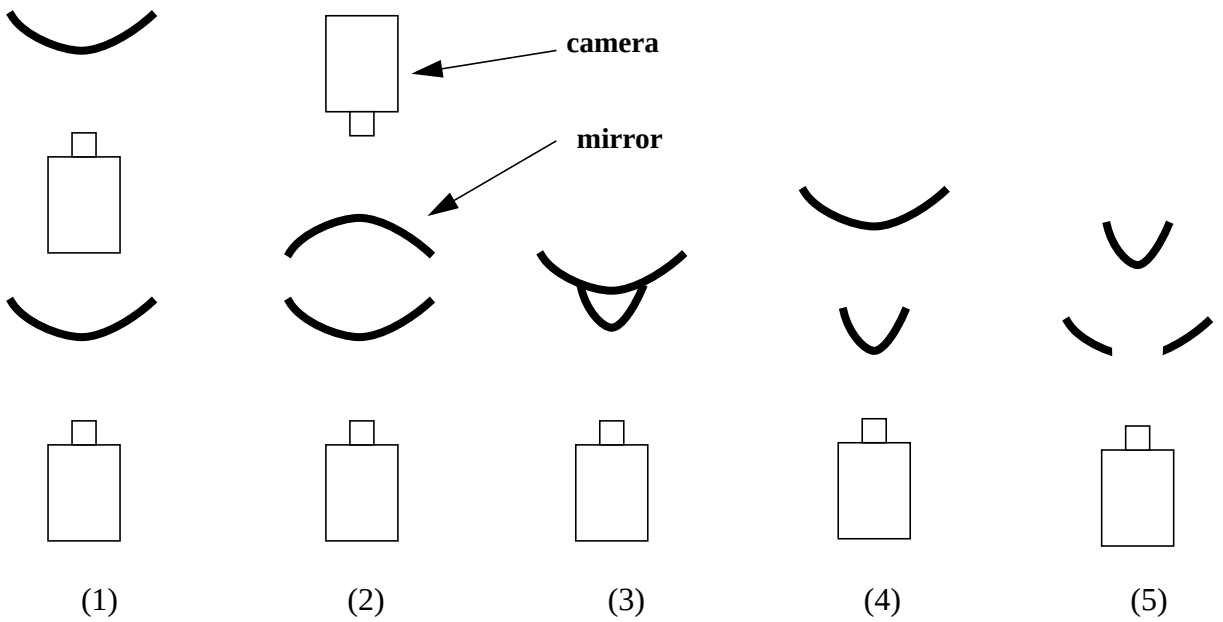


Figure 7 Side view showing the five configurations examined. (1) and (2) use two cameras. (3), (4) and (5) use a single camera.

The primary issues we consider in this report are the following:

- **Field of View**: As with all stereo systems, depth can only be computed for points which are imaged with both cameras. In order to make best use of the camera imaging areas, the two camera views should overlap as much as possible.
- **Resolution**: While it is possible to do stereo matching on two images whose effective spatial resolution is significantly different, it has several disadvantages. The chief drawback is that the range precision will be determined by the lower-resolution camera, and so the detail provided by the higher-resolution camera will be wasted.
  Another drawback is that in general, smoothing of the higher resolution image will be required in order to do the match. The smoothing required will depend on how badly the resolutions mismatch, which may vary strongly from one part of the image to another; this smoothing will thus have to be done computationally, rather than by simply defocussing the lens.
- **Range Accuracy:** For a standard stereo system, the range accuracy is determined by the separation of the cameras, their resolution, and the quality of the disparity calculation. Once the physical cameras and the matching algorithm are selected, the separation of the cameras (the baseline) becomes the dominant issue.
  The situation is somewhat more complex for the camera systems described below. For example,

an analytic expression for range precision of the two camera/two hyperbolic mirror case can be computed from (10); it will depend on the distance of the cameras from the mirrors, the distance between the mirrors, and depends as well on the location of the point being imaged. By making some assumptions about the field of view of the original camera (60 degrees), the resolution of the CCD (640x480), and the quality of the feature matching (within 1 pixel), we analyzed some configurations to see how the range accuracy deteriorates with distance.

- **Simulated range image quality**: For some of the configurations, we have actually calculated range maps from simulated images of a simple, flat, textured planar environment. The mathematics involved in this calculation are explicitly shown for one configuration (using two cameras and two hyperbolic mirrors); the calculations involved in the other cases are analogous, but not shown here. Further discussion of the stereo algorithm is deferred until section 2.

While obviously not as compelling as real data, the simulated images do serve to illustrate many of the advantages and disadvantages of each system under ideal conditions.

3.1. Configuration 1

Consider the mirror-camera system shown in Figure 8, where the two cameras are separated by a baseline, b. An image of a planar, checkerboard world with this configuration produces image pairs such shown in Figure 9.

3.1.1. Computing Range

If the cameras are vertically aligned such that their optical axes are co-incident and the images are not rotated with respect to each other, then any point in the world will be imaged somewhere along the same radial line (starting from the center of the image and going outward) in each image. If we can find the image points r_1' and r_2' (see Figure 7) that correspond to a point in the world (r_g, y_g), then it is possible to estimate r_g and y_g. Determination of the match, i.e. the process of finding r_1' and r_2' is done by a process called correlation (see 4.1. for more details). Below we show how range estimation is done.

First, it is necessary to develop an expression relating an imaged point (r_g, y_g) to its focal plane coordinate r' for a single camera-hyperboloid mirror configuration. We begin with the standard projective equation, which relates r' to its location on the mirror, (r_m, y_m) (f is the camera focal length):

$$r' = \frac{f r_m}{y_m} \tag{4}$$

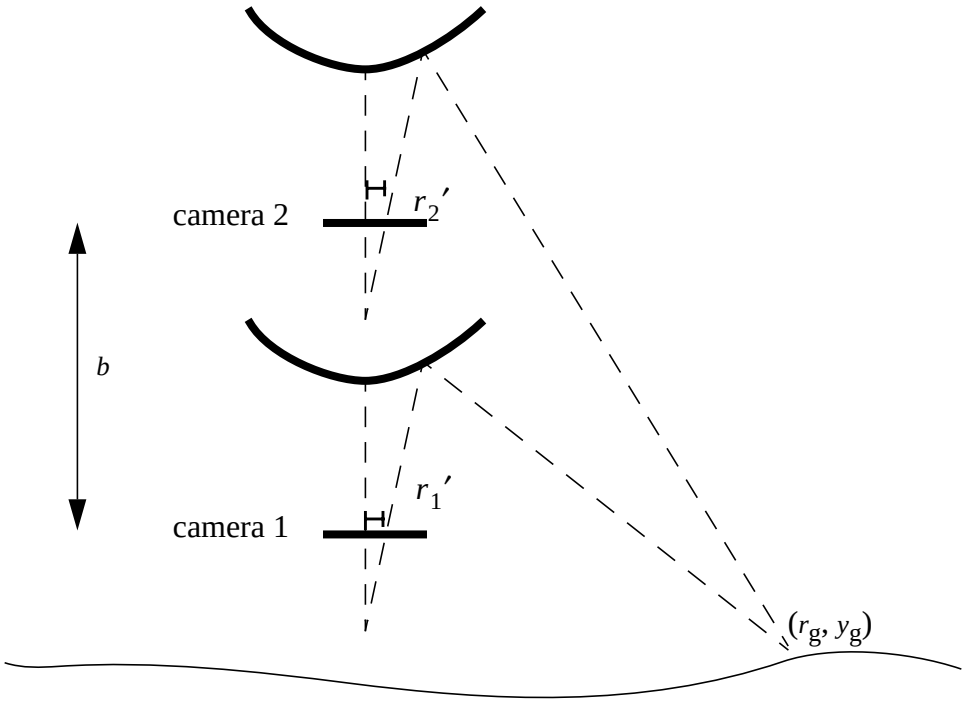
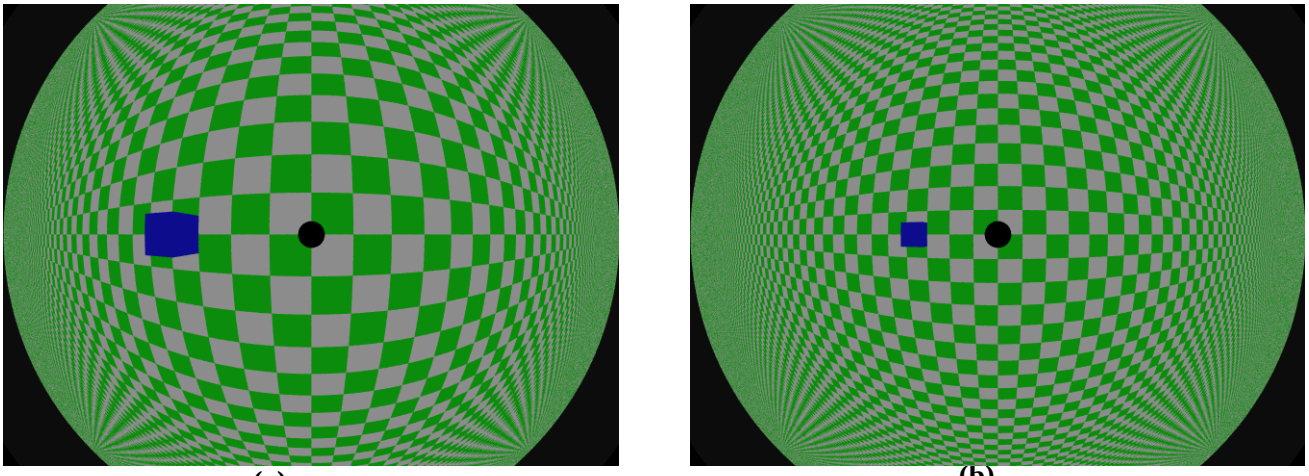


Figure 8 **Configuration 1. Two hyperbolic cameras and two mirrors with a vertical baseline.**



(a) **(b)**

Figure 9 **Synthesized images for configuration 1. (a) lower camera (b) upper camera. The mirrors are hyperbolic (alpha = 3); the baseline is 30 cm. In this example, the lower camera is 20 cm from the floor; the upper camera is 50 cm from the floor; and the two hyperbolic mirrors are mirrors are displaced 40 cm and 70 cm above the floor.**

Since we know the equation for the hyperboloid mirror is $y_m = \sqrt{r_0^2 + r_m^2}$, we can rewrite (1) as

$$r' = \frac{f r_m}{\sqrt{r_0^2 + r_m^2}} \tag{5}$$

We can now solve for (r_m, y_m) in terms of f and r':

$$r_m = \frac{r_0 r'}{\sqrt{f^2 + r'^2}} \qquad y_m = \sqrt{r_0^2\left(1 + \frac{r'^2}{f^2 + r'^2}\right)} \tag{6}$$

Since we know that a ray coming from the camera at incidence angle θ bounces off the mirror at angle 3θ, we can write

$$r_g = r_m + (y_m - y_g)\tan(3\theta) \tag{7}$$

where $\theta = \text{atan}\left(\frac{y_m}{r_m}\right)$. Substituting in the expressions for r_m and y_m from (3), and using the trigonometric identity,

$$\tan(3\theta) = \frac{3\tan\theta - \tan^3\theta}{1 - 3\tan^2\theta} \tag{8}$$

We finally derive the equation relating (x_g, y_g) to a radial distance r' measured from the center of the camera CCD:

$$r_g = \frac{r_0 r'}{\sqrt{f^2 - r'^2}} + \frac{(3f^2 r' - r'^3)\left(\sqrt{\frac{f^2 r_0^2}{f^2 - r'^2}} - y_g\right)}{f^3 - 3fr'^2} \tag{9}$$

For two identical camera-mirror systems separated by baseline b, we can now write a system of two equations relating r_1' and r_2' to r_g and y_g:

$$r_g = \frac{r_0 r_1'}{\sqrt{f^2 - r_1'^2}} + \frac{(3f^2 r_1' - r_1'^3)\left(\sqrt{\frac{f^2 r_0^2}{f^2 - r_1'^2}} - y_g\right)}{f^3 - 3f r_1'^2} \tag{10}$$

$$r_g = \frac{r_0 r_2'}{\sqrt{f^2 - r_2'^2}} + \frac{(3f^2 r_2' - r_2'^3)\left(\sqrt{\frac{f^2 r_0^2}{f^2 - r_2'^2}} - y_g - b\right)}{f^3 - 3f r_2'^2}$$

Given any two of these parameters, we can solve for the other two using these equations. For stereo, we are given r_1' and r_2), and thus can solve for r_g and y_g. There is, in fact, an algebraic closed form solution for r_g and y_g in terms of r_1' and r_2', but it is somewhat lengthy and thus omitted here.

3.1.2. Depth map results

Since the checkerboard patterns are artificial, we have used synthesized panoramic views of a flat plane with a marbleized texture as shown in Figure 10. It is expected that the performance of the stereo matching will vary as some of the radial lines will be orthogonal to the texture which will help the matching process while others will be parallel to the texture which will make matching difficult.

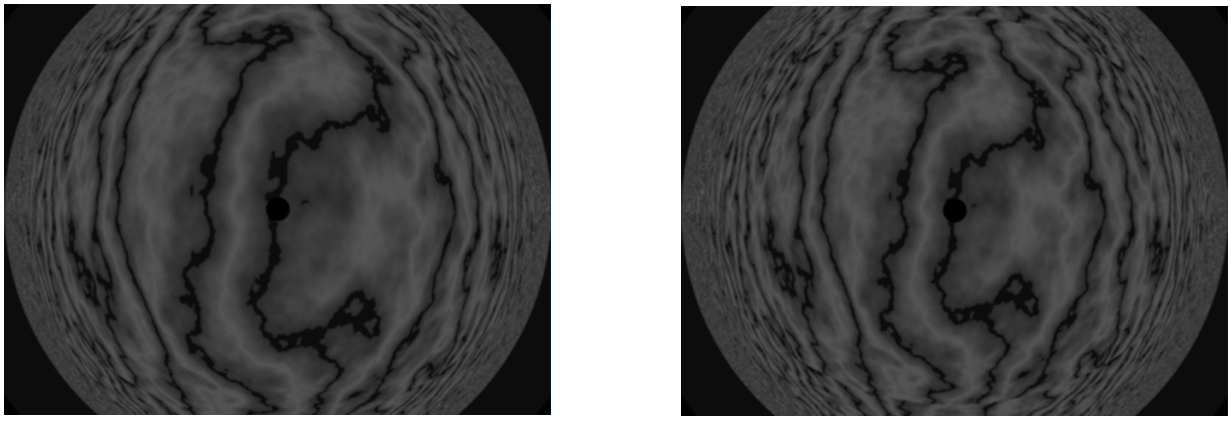
Stereo imaging produces radial patterns as shown in Figure 11.

Figure 12shown more quantitative results for this configuration. Figure 12 shows a 3D view of the reconstructed terrain.

3.1.3. Discussion

This configuration has many desirable qualities. The resolution difference between the two cameras, while significant, varies much less than in the other configurations, because both mirrors are equally curved; and the resolution of both falls off at an approximately equal rate. The angular field of view matches as well.

This configuration requires two cameras, which drives up both the expense and the effort involved for calibration. However, it does allow the range precision to be easily controlled simply by increasing the spread between the two camera-mirror systems. This kind of adjustment is much more difficult in for the



(a) **(b)**

Figure 10 Synthesized images for computing the depth map for configuration 1. (a) lower camera (b) upper camera. The mirrors are hyperbolic (α= 3); the baseline is 30 cm. In this example, the lower camera is 20 cm from the floor; the upper camera is 50 cm from the floor; and the two hyperbolic mirrors are mirrors are displaced 40 cm and 70 cm above the floor.

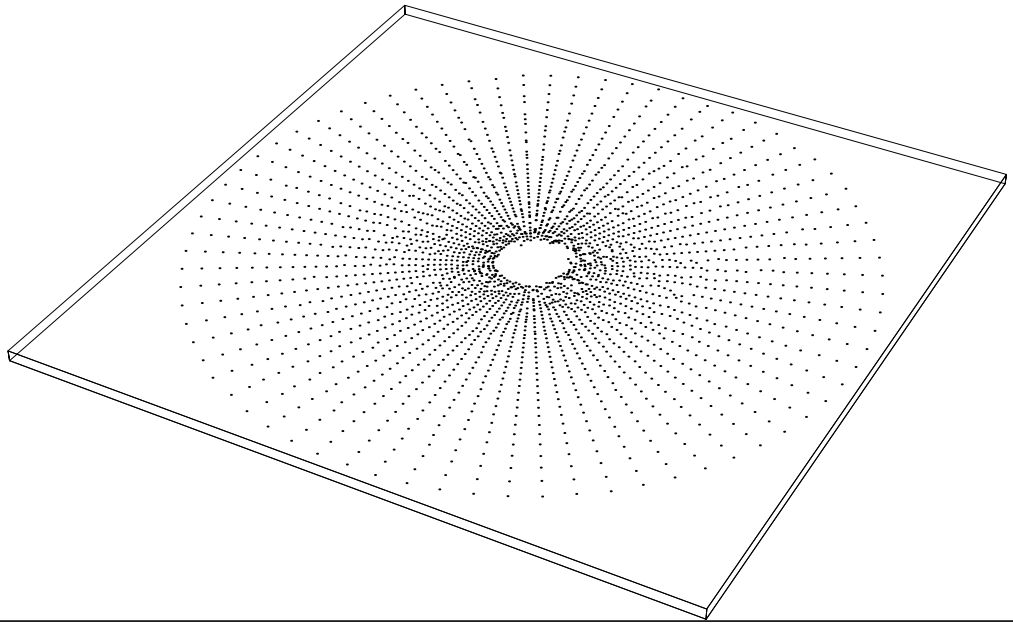


Figure 11 Radial patterns created by panospheric stereo imaging.

single-cameras, multi-mirror configurations; the only possible adjustment in those cases is to alter the distance between the cameras and the mirrors, which necessitates changing the mirror shape in order to retain the roughly equal-angle qualities described in Section 2.

A graph of range precision versus distance is shown Figure 14 (assuming a 60 degree FOV camera with 640x480 resolution, and features matched to 1 pixel). Precision is a measure of the best possible resolution possible. Of the five configurations we have analyzed, this configuration has the best range precision.

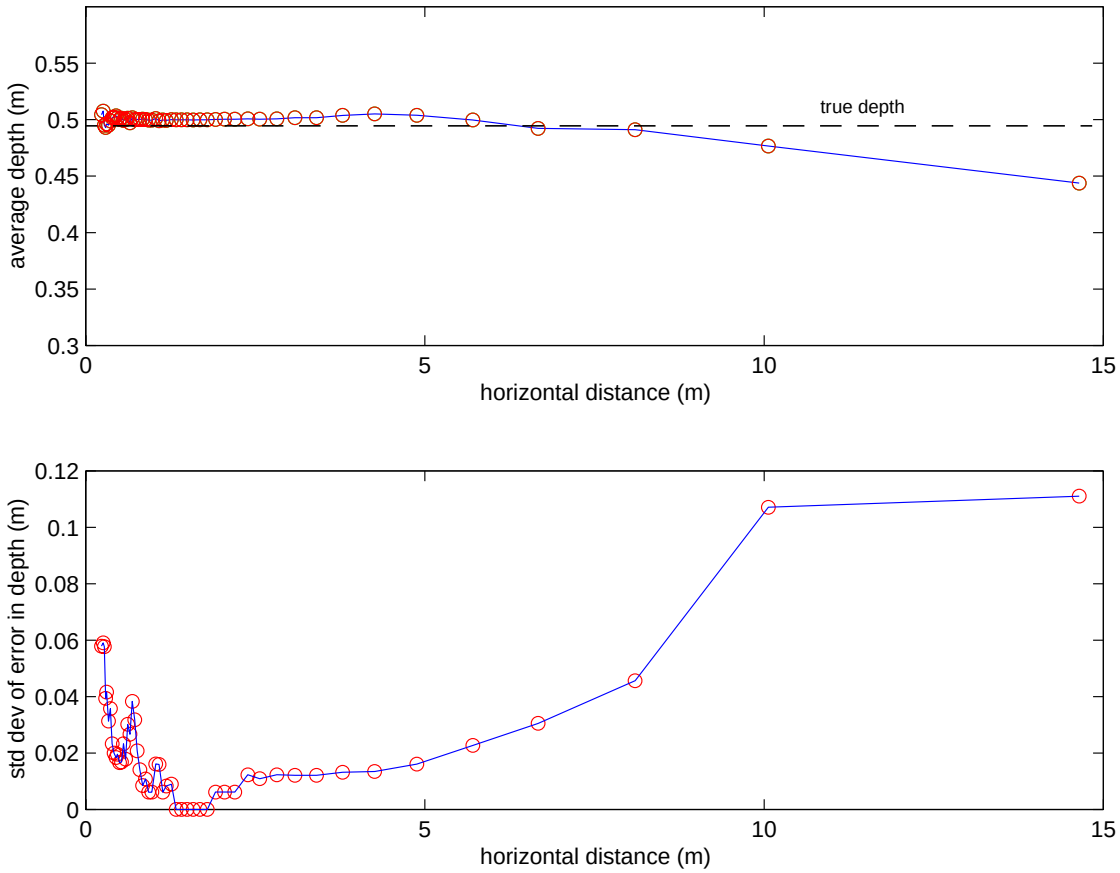


Figure 12 Reconstruction of the terrain using configuration 1. Depth error is due to differences in texture on the plane (a) A 2D profile of the terrain along the radial distance. (b) Standard Deviation in the depth error

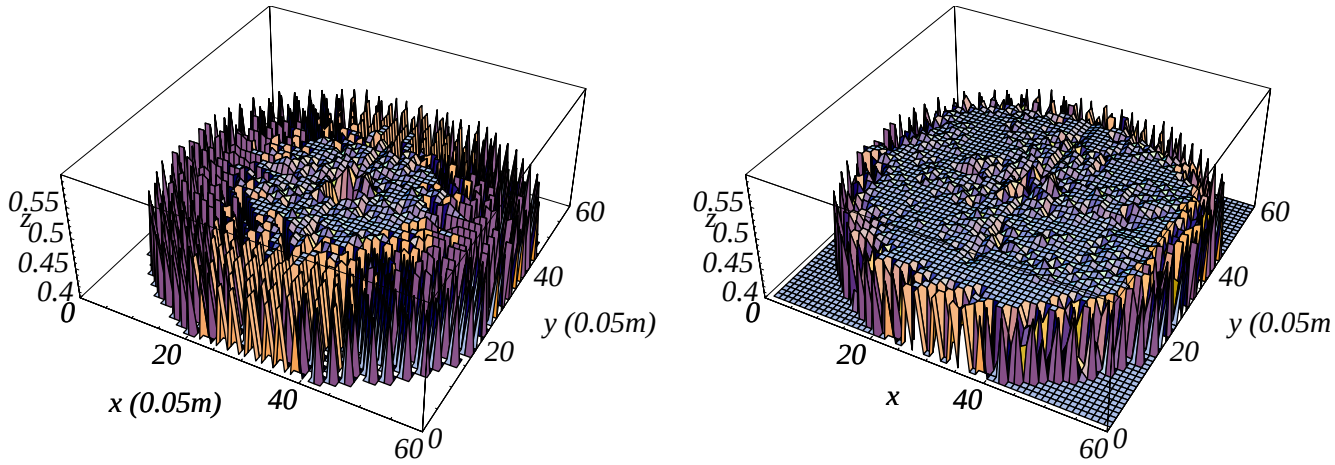


Figure 13 3D view of reconstructed terrain using configuration 1. 3D points are integrated into a regularly spaced grid of height values.(a) Raw terrain map (b) Interpolated terrain map. Unknown points are estimated using known neighbors.

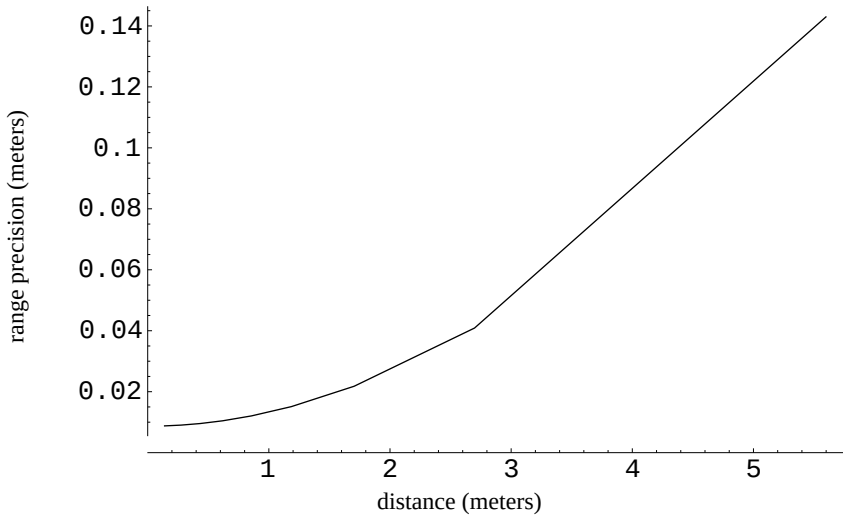


Figure 14 Range precision with configuration 1 as a function of distance to the object. Here, the baseline between the two mirrors is 30 cm.

3.2. Configuration 2

A second configuration that was examined is shown in Figure 15. Imaging the same checkerboard world as in the first example with this configuration produces image pairs shown in Figure 16.

3.2.1. Discussion

Simply by viewing the two images above, it is clear that this configuration is lacking on several accounts. The field of view overlap is minimal, and is limited to the extreme edges of both cameras. Further, since the two mirrors are so close together, the triangulation necessary for range value calculation will have a relatively small baseline compared to the total space taken up by the system. Because of these drawbacks, no attempt was made to calculate range maps from this configuration.

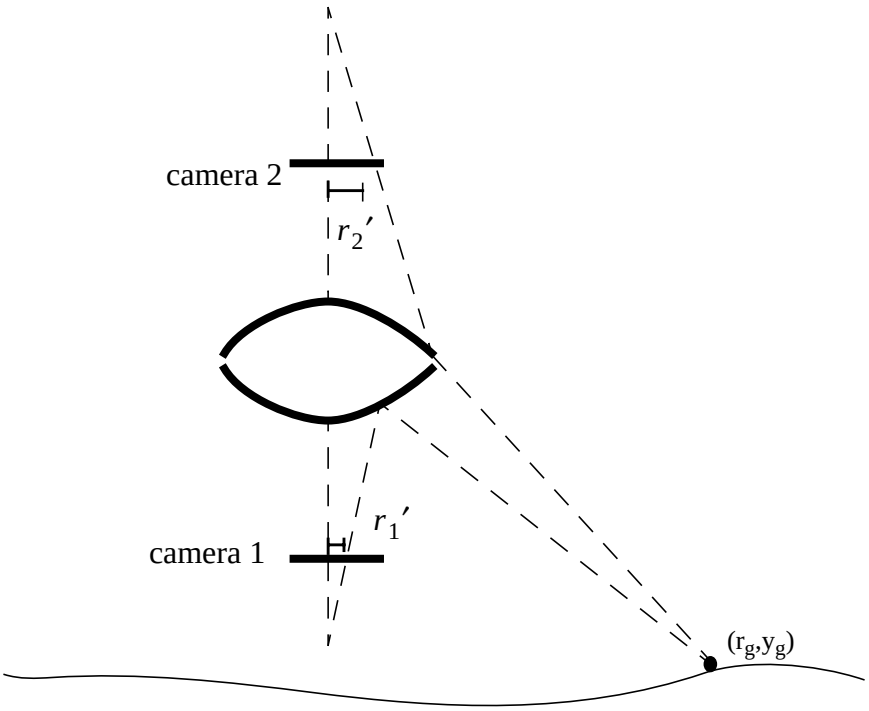
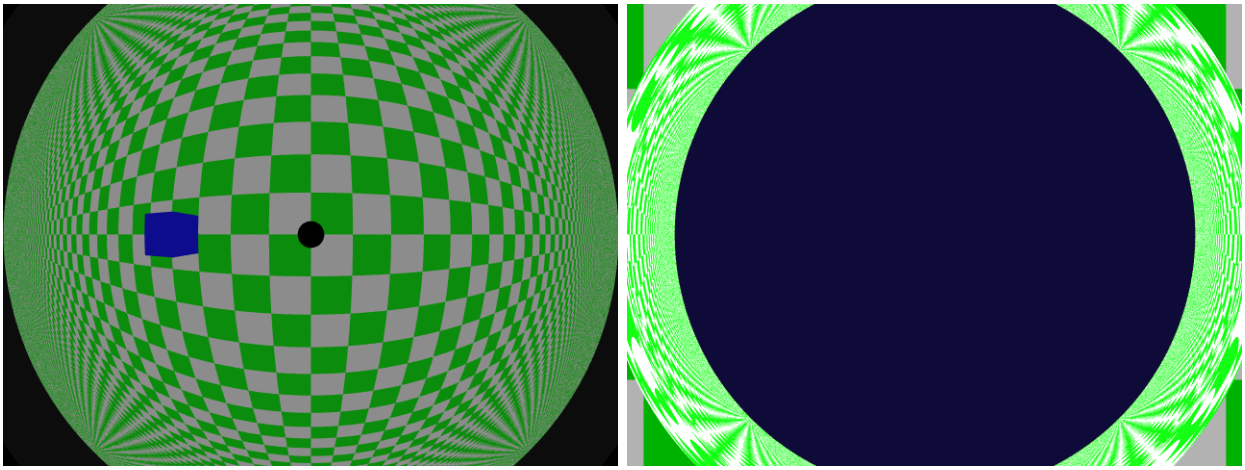


Figure 15 Configuration 2. Two cameras and two mirrors with a vertical baseline.



(a) **(b)**

Figure 16 Synthesized images for configuration 1. (a) lower camera (b) upper camera. Both mirrors in this example are hyperbolic ($\alpha = 3$); the vertices of the two mirrors are 10 cm apart. In this example, the lower camera is 20 cm from the floor; the upper camera is 70 cm from the floor; the mirrors are 40 cm and 50 cm above the above the floor, respectively.

3.3. Configuration 3

A third configuration that was examined is one in which mirrors of two curvatures are used with a single camera as shown in Figure 17. Imaging the same checkerboard world as in the first example with this

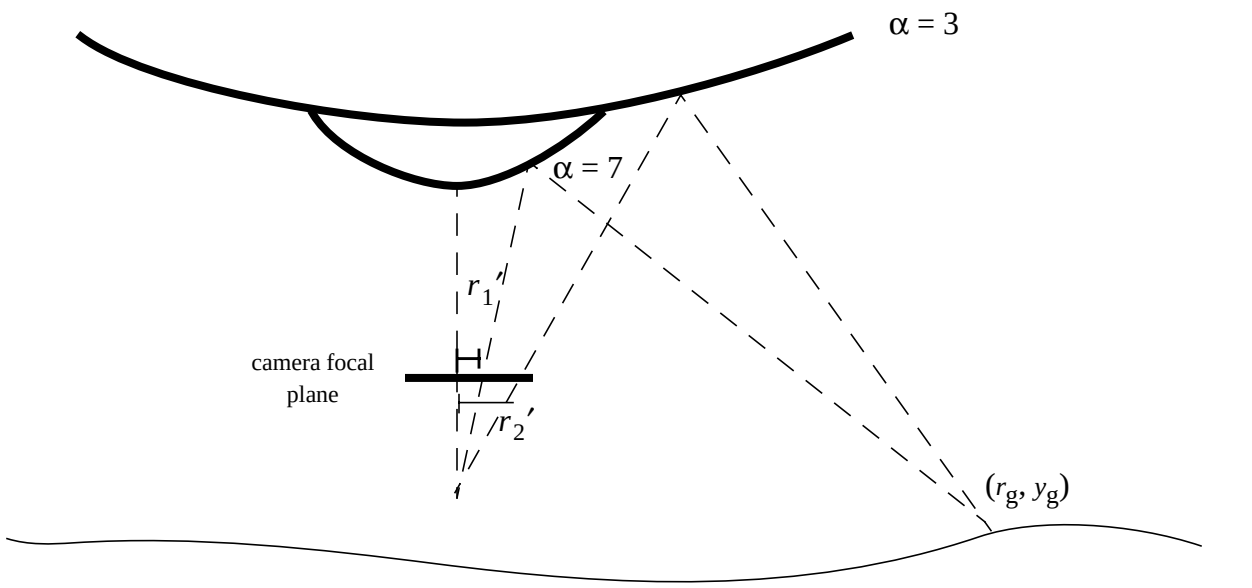


Figure 17 Configuration 3. A single camera is used with a dual lobed mirror. A lower curvature mirror is above and a higher curvature mirror is below.

configuration produces an image shown in Figure 18.

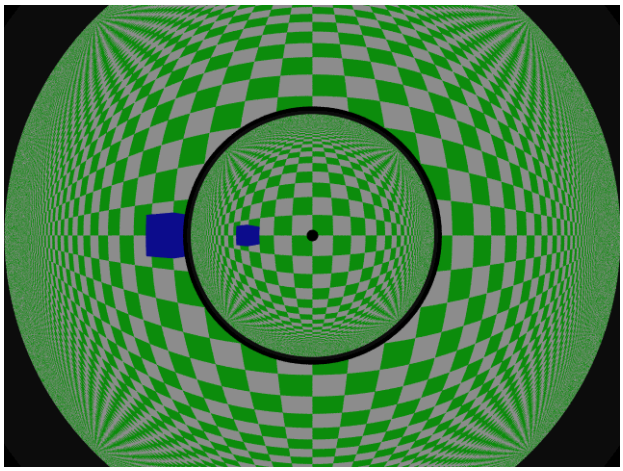


Figure 18 Single image from the dual lobed mirror shown in Figure 17. in this example, the camera is 50 cm from the floor; the α = 7 mirror is 68.5 cm from the floor; and the α=3 mirror is 70 cm from the floor)

3.3.1. Depth map results

Once again we used a textured flat world to synthesize images for configuration 3 as shown in Figure 19.

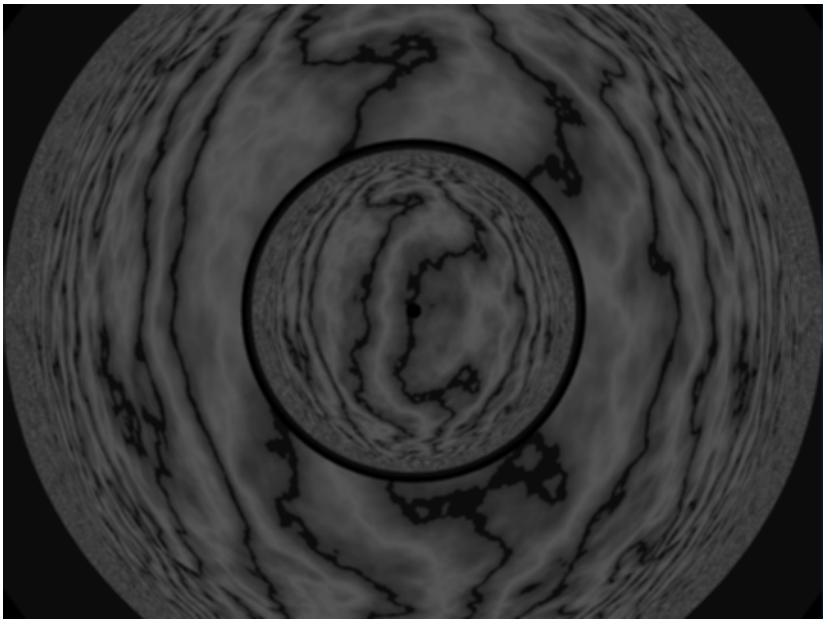


Figure 19 Single image from the dual lobed mirror shown in Figure 17. This image is used to generate the depth map using stereo. In this example, the camera is 50 cm from the floor; the α = 7 mirror is 68.5 cm from the floor; and the α=3 mirror is 70 cm from the floor)

Quantitative results for this configuration are shown in Figure 20. Terrain maps generated from the single image in Figure 19 are shown in Figure 21.

3.3.2. Discussion

The field of view of the two mirrors can be chosen to overlap almost completely by appropriately sizing the two mirrors, although, obviously, the lower mirror occludes a portion of the upper mirror's view. Also, unlike the first two configurations, this one requires only one camera, reducing both the cost, mechanical infrastructure and the calibration effort required.

The real disadvantage of this method is the relatively small baseline it provides. Since the two mirrors are so close together, the effective baseline for the stereo calculation is quite small. In the figure below, we plot the range precision (assuming a 60 degree FOV camera with 640x480 resolution, and features matched to 1 pixel). For objects more than just a few meters away, the uncertainty soon exceeds the actual range!.

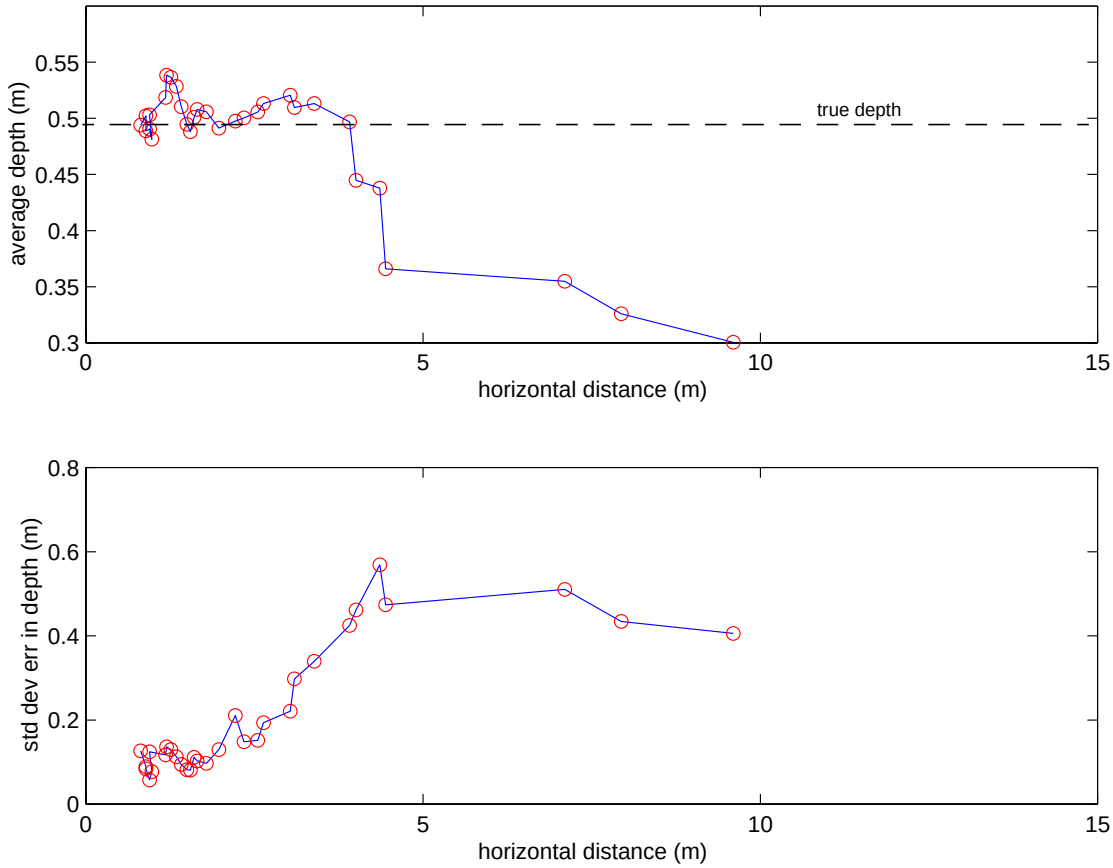


Figure 20 Reconstruction of the terrain using configuration 3. (a) A 2D profile of the terrain along the radial distance. (b) Standard Deviation in the error due to the differences in texture.

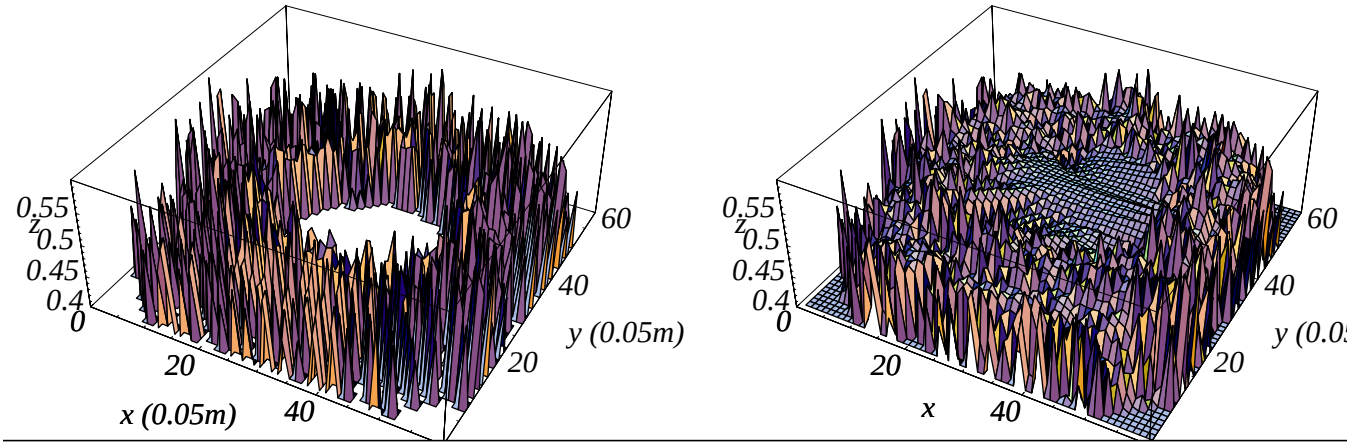


Figure 21 3D view of reconstructed terrain using configuration 3 (a) Raw terrain map (b) Interpolated terrain map.

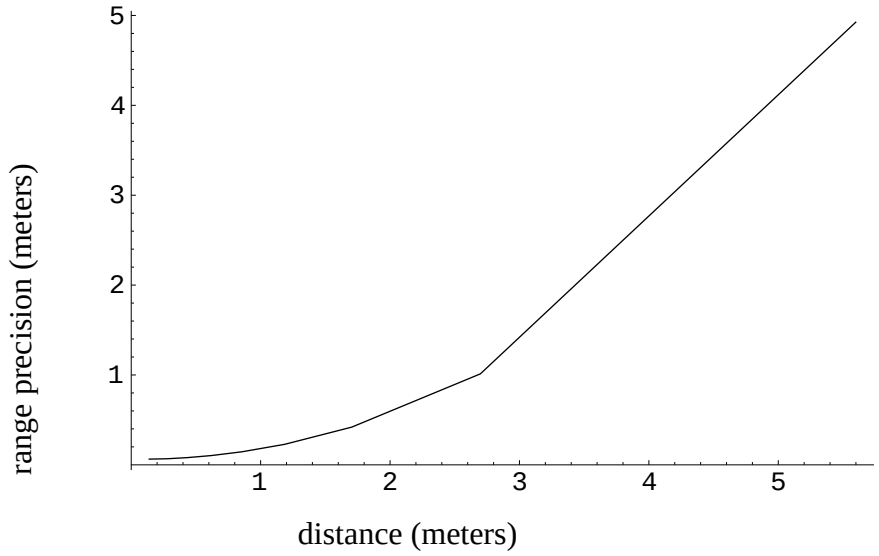


Figure 22 Range precision with configuration 3 as a function of distance to the object. The vertices of the lower and upper mirrors here are separated by only 1.5 cm.

3.4. Configuration 4

A fourth configuration examined is shown in Figure 23. This configuration is similar to configuration #3 except that the two lobes of the mirror are separated to provide a larger baseline. Imaging the same checkerboard world as in above examples with this configuration produces an image shown in Figure 24. Notice that in this image, the cube appears only in the part of the image that corresponds to the higher curvature (lower) mirror. This is because the cube is at such an orientation that the lower mirror occludes the upper mirror. In general, for a fixed mirror size, the greater the baseline, the smaller this type of occlusion.

3.4.1. Discussion

As can be seen, this configuration combines the equal-resolution aspects of configuration #3 with the increased baseline of configuration #5. For imaging objects below the horizon, this is our configuration of choice. A graph of range precision as a function of distance is presented in Figure 25, showing an improvement over all the configurations except #1.

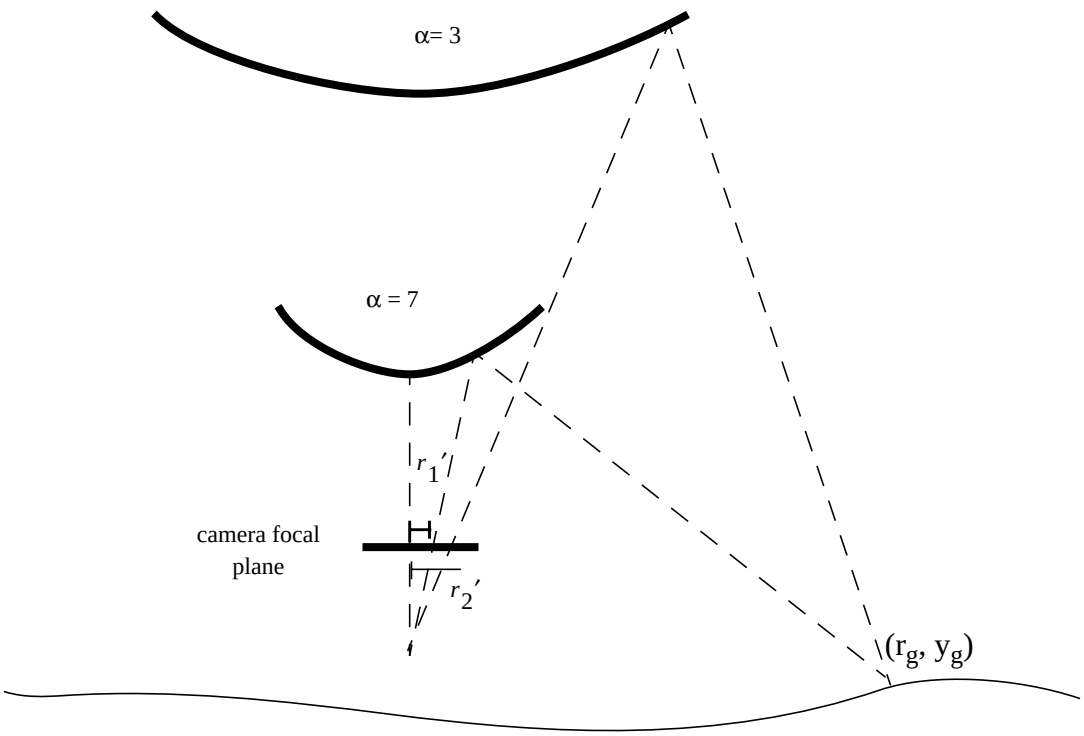


Figure 23 Configuration 4 uses two mirrors and 1 camera. The lower curvature mirror is placed above the higher curvature mirror.

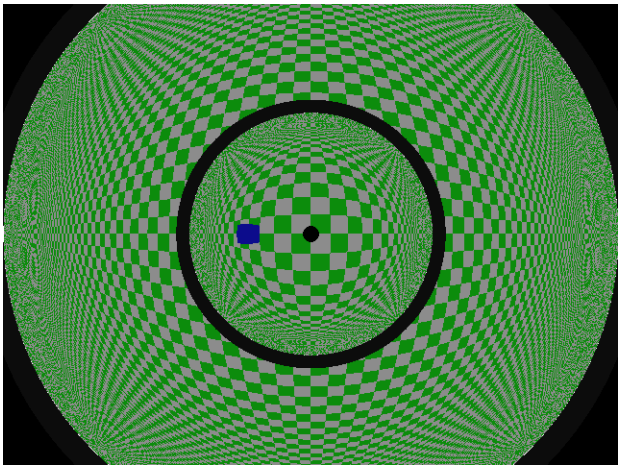


Figure 24 Single image from the dual lobed mirror shown in Figure 23. Here, the vertices of the two mirrors are separated by 30 cm. The image of the blue box is out of view of the higher mirror. (in this example, the camera is 50 cm from the floor; the α = 7 mirror is 70 cm from the floor; and the α =3 mirror is 100 cm from the floor

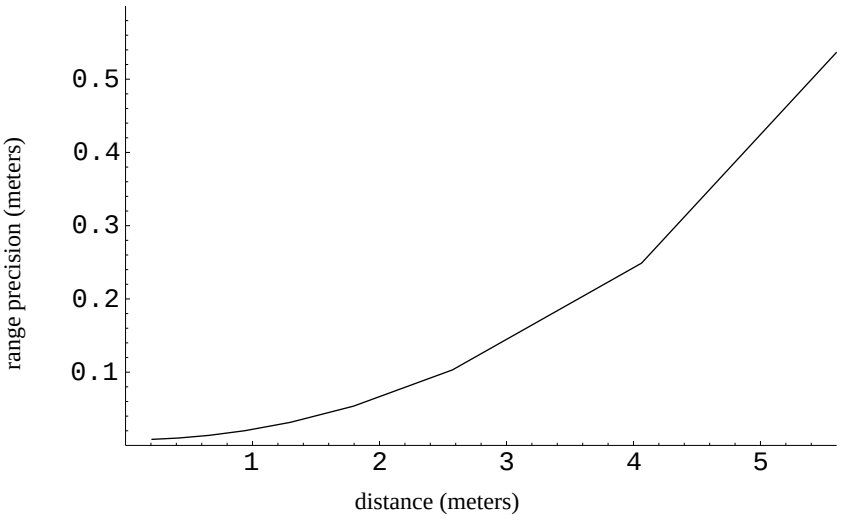


Figure 25 Range precision with configuration 4 as a function of distance to the object. The vertices of the lower and upper mirror here are separated by 30 cm.

3.5. Configuration 5

The final configuration examined is shown in Figure 26. Here a high curvature mirror is placed above the low curvature mirror which has a hole in the middle. Imaging the same checkerboard world as in the first

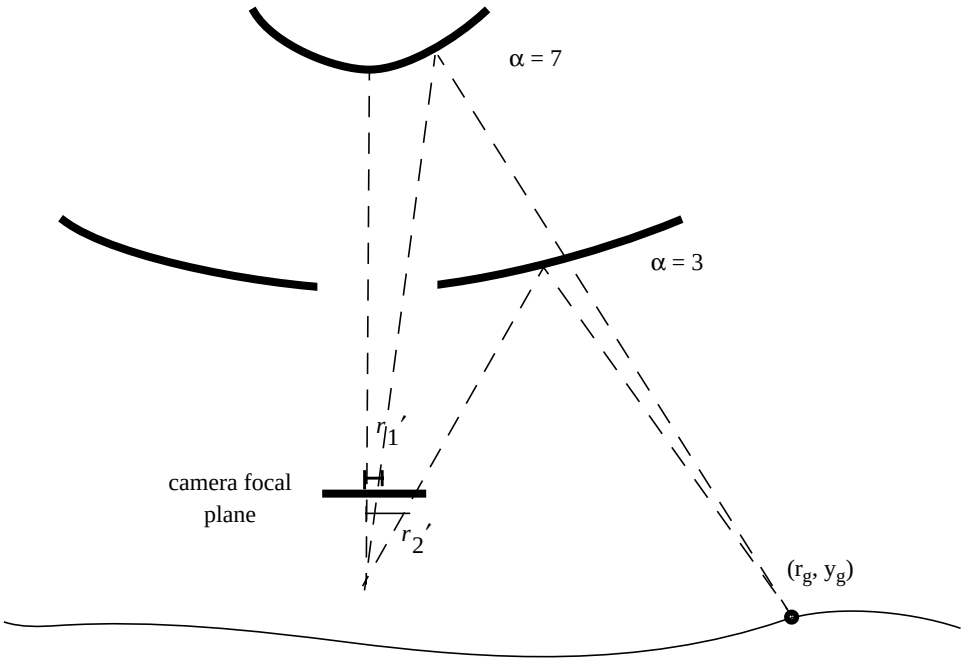


Figure 26 Configuration 5. A higher curvature mirror is placed over a lower curvature mirror. Two mirrors and a single camera are used.

example with this configuration produces an image such as this (in this example, the camera is 50 cm from the floor; the $\alpha = 7$ mirror is 100 cm from the floor; and the $\alpha = 3$ mirror is 70 cm from the floor. Note that once again the lower mirror occludes the view of blue cube for the upper mirror.

3.5.1. Depth map results

For configuration 5, we used the image shown in Figure 28 to generate a terrain elevation profile.

Figure 29 shows quantitative results for configuration 5. Figure 29 shows a reconstructed terrain map for this configuration.

3.5.2. Discussion

This is another variant of the configuration 3, with many of the same advantages and disadvantages. There are a few points worth noting, however. The principal disadvantage of configuration 3 has been overcome; the separation between the two mirrors provides much improved precision, as shown in the graph below. The quality of the depth maps has improved correspondingly.

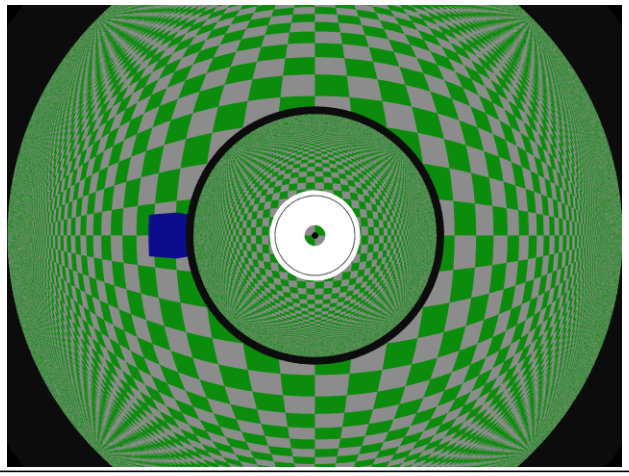


Figure 27 Single image from the dual lobed mirror shown in Figure 26. Here, the vertices of the two mirrors are separated by 30 cm. The image of the blue box in the inner, higher mirror is blocked by the outer, lower mirror.

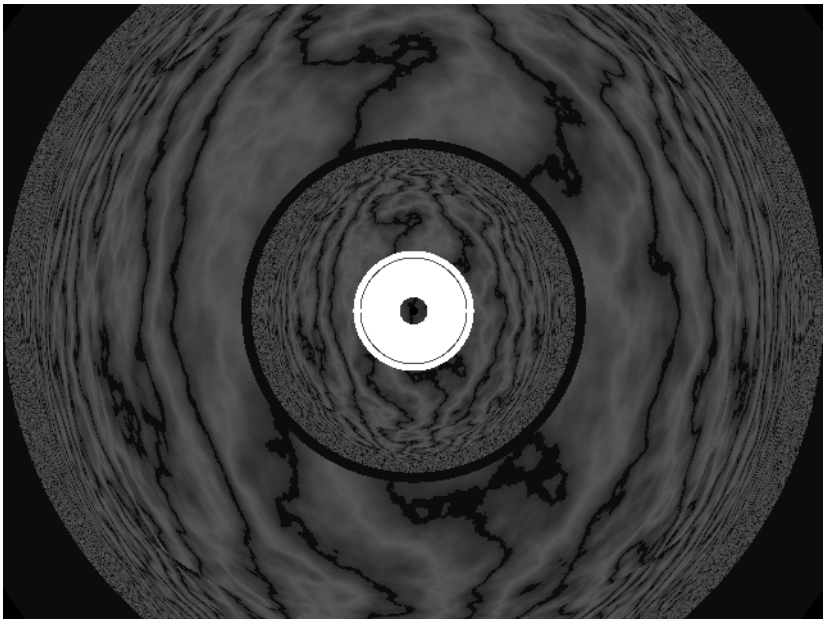


Figure 28 Single image from the dual lobed mirror shown in Figure 26. This image is used to generate the depth map using stereo. Here, the vertices of the two mirrors are separated by 30 cm. The image of the blue box in the inner, higher mirror is blocked by the outer, lower mirror.

One disadvantage, however, is that the resolution in the two mirrors is less well matched than in configuration #3. Objects below the horizon appear smaller in the central mirror because it is both more tightly curved and higher from the ground. For objects below the horizon, therefore, this configuration is in general inferior to configuration 4.

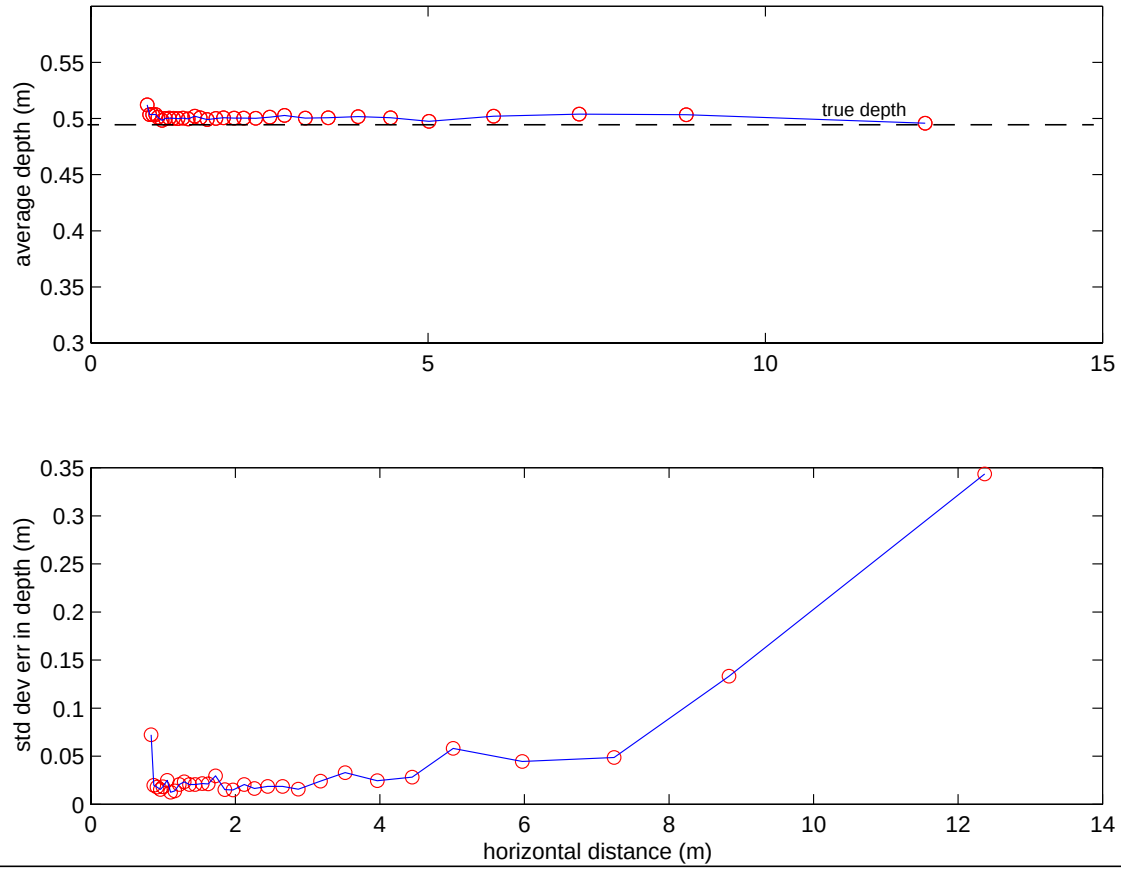


Figure 29 Reconstruction of the terrain using configuration 5. (a) A 2D profile of the terrain along the radial distance. (b) Standard Deviation in the error due to the differences in texture.

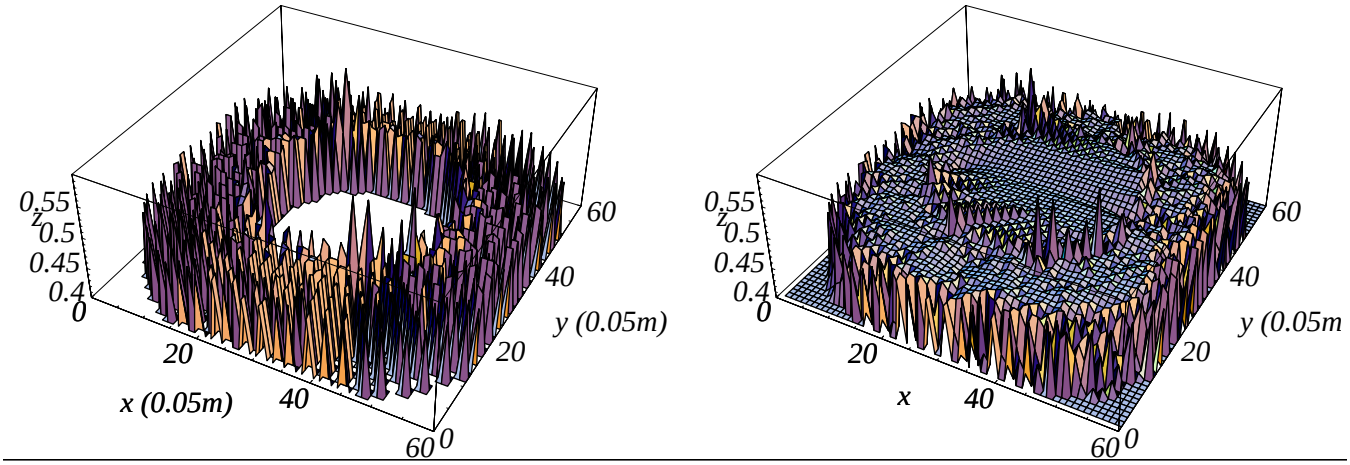


Figure 30 3D view of reconstructed terrain using configuration 5 (a) Raw terrain map (b) Interpolated terrain map.

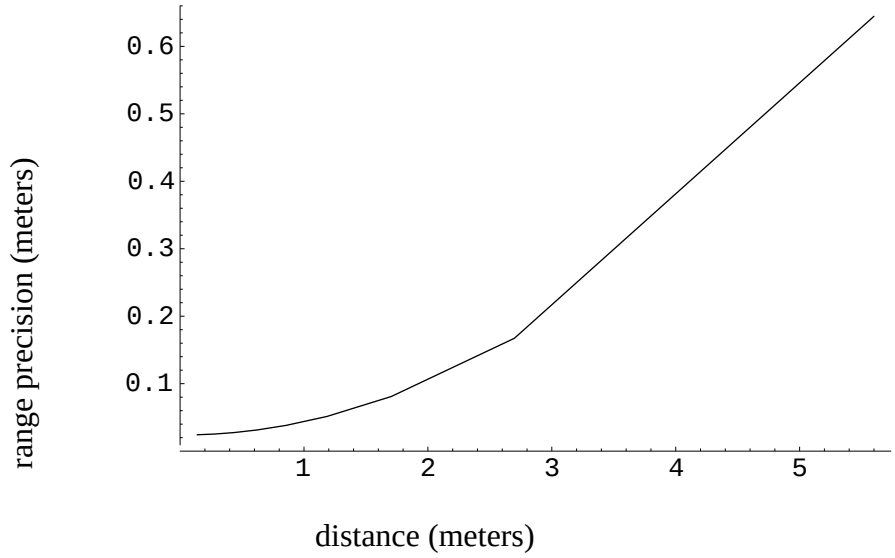


Figure 31 Range precision with configuration 5 as a function of distance to the object.

4. Stereo Algorithms

As discussed in the introduction, the algorithms typically used by stereo vision systems are not strictly valid for the configurations described above, largely because these systems are no longer projective devices. In this section, we discuss the modifications necessary to make stereo work on images from these systems. We first discuss the correlation process, by which a window in one mirror is matched to a window in another; next, we discuss how this matching information is used to estimate the 3-D location of the feature.

4.1. Correlation

Stereo vision requires matching of features in one image to another. The process typically consists of correlating a window in one image to windows along an "epi-polar" line in the other image as shown in Figure 32.

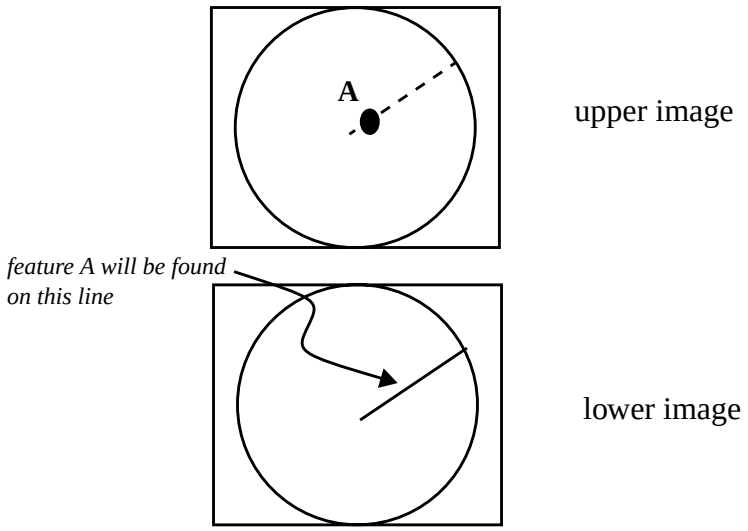


Figure 32 The epi-polar constraint for correctly aligned (co-incident optical axes, with zero rotation) panospheric cameras. For points in the intersecting field of view between the two cameras every feature in one image can be found in the other image on the radial line from the center of the image. A straightforward extension applies to the case of two mirrors and one camera.

The process of correlation is well studied and there are few well known algorithms that are used by researchers. We have used a correlation process known as "Sum of Squares Difference" (SSD). This correlation minimizes the normalized sum of squared difference of the two windows. We used a 21 x 21 window for correlation.

While the correlation methods are standard, simple correlation of square windows along the epi-polar lines makes strong assumptions-- projective cameras that image surfaces normal to the camera. While this assumption doesn't hold for most stereo applications, because in general the imaged surface will not be normal to the optical axis, for our panospheric imaging systems, this assumption is even more dangerous because square windows in one image do not correspond to square windows in the other image. Figure 33 shows how an example of how a rectangular window in one image warps when mapped to the correct match in the other image.

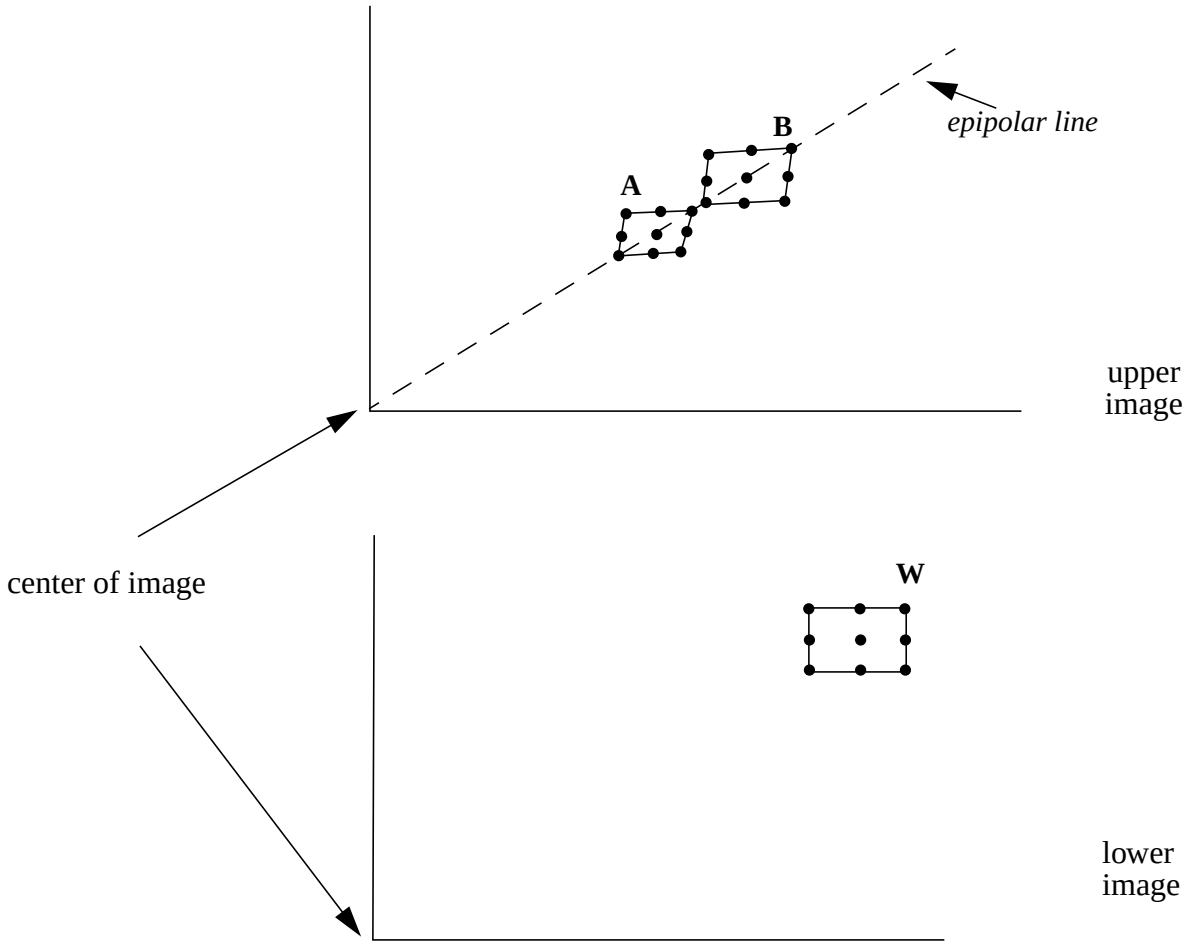


Figure 33 An example of correctly mapping a window from one image to another with mirrors of equal curvature. The top right hand quadrant of two images from cameras in configuration 1 (both mirrors α = 3, baseline 30 cm) are shown. The window W in the lower image will match with window A if the depth to the patch imaged is 5 cm and with window B if the depth is 50 cm. In both cases, a flat plane is assumed. Note that only discrete points from each window are used. This illustration shows the matching of a 3x3 window.

This effect is noticed even when mirrors of the same curvature are used but is much more pronounced when the mirror curvatures are different as shown in Figure 34. Thus, during the correlation process, the actual shape of the window in one image needs to be warped appropriately as the window is translated

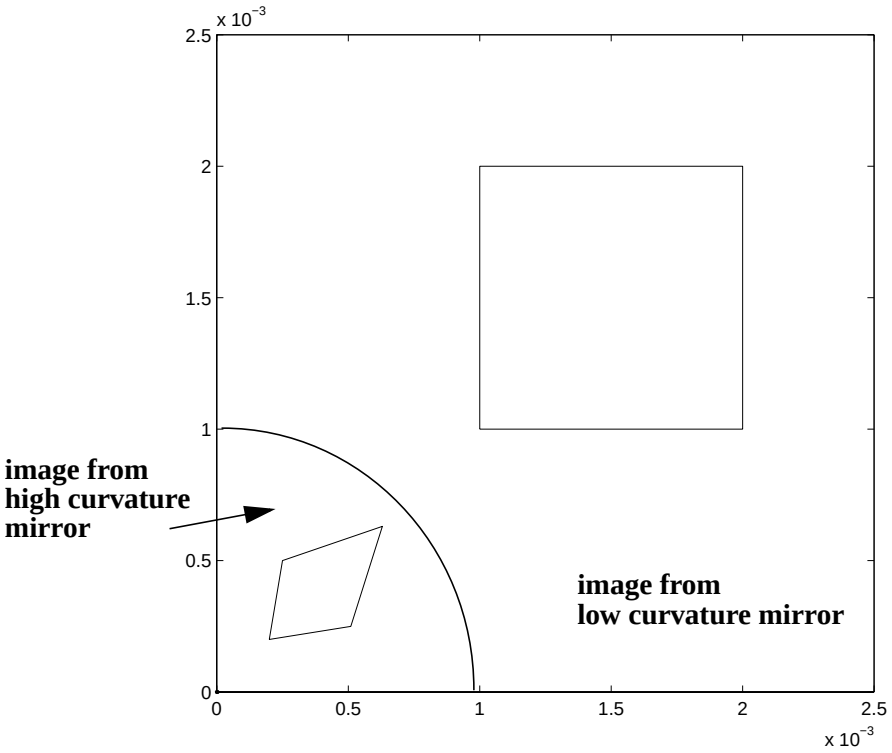
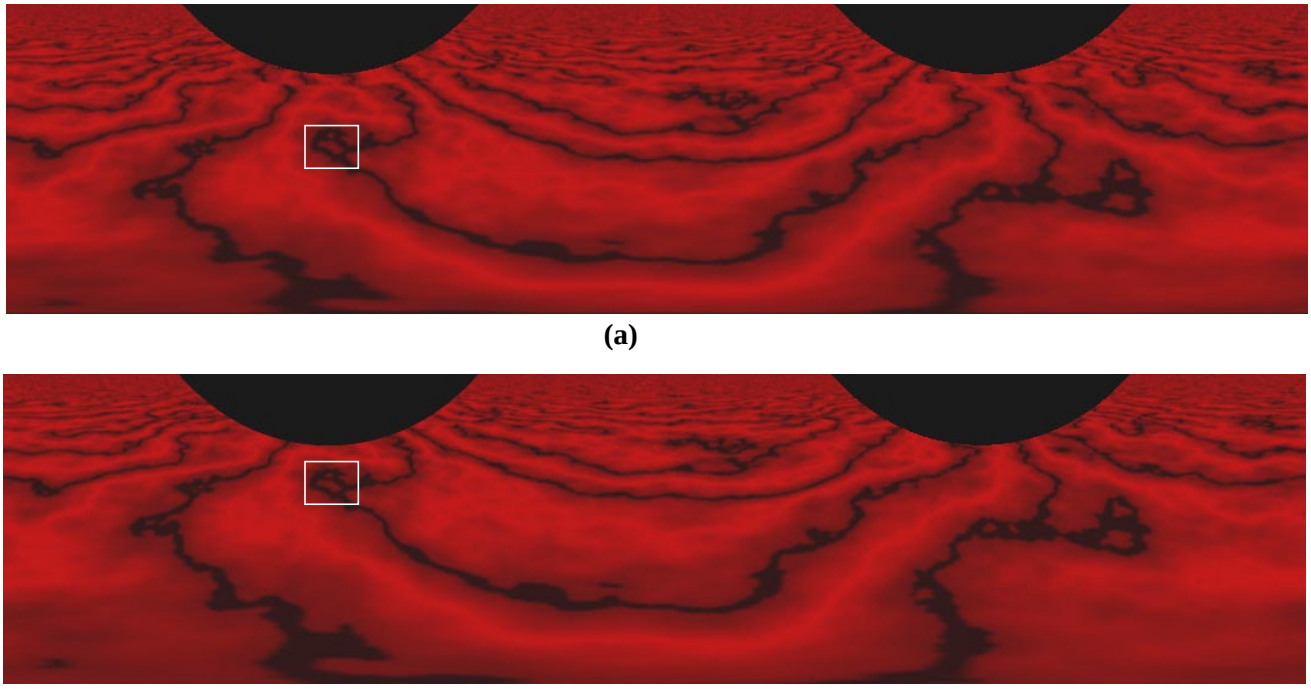


Figure 34 Example of window warping when mirrors of varying curvature are used. Points in a window on the image corresponding to the low curvature mirror must be must be remapped severely to produce the correct match. In this case configuration #4 is used. The warping effect is some what exaggerated because the lower mirror is very close (5cm) to the surface being imaged.

around the epi-polar line. Since the calculations required for this warping are somewhat complex, it is necessary to generate a lookup table to map pixels in one mirror onto pixels in another; this is the method used with all the images in Section 4.

To show the necessity of incorporating mirror geometry into the correlation we conducted the following test. We used the configuration with the best precision (configuration #1) to generate unwarped images (Figure 35). That is, the raw images were transformed into rectangular images assuming a spherical projection. In this test, we conducted correlation on the unwarped images but triangulation with the warped images using (10). Doing so ignores the distortion from the mirrors and simply uses standard window-based matching methods to generate the correspondences. Such a method, however, can introduce substantial inaccuracies into the range map.

The quantitative results from using this method are shown in Figure 36. Figure 37 shows the terrain maps generated from these images.



(a)

(b)

Figure 35 Unwarped images used for correlation. Note that features (e.g. in box) are shifted vertically between images. (a) image from upper camera (b) image from lower camera.

4.2. Triangulation

Given a matching feature in the two mirrors, the correct way to compute the 3D location of the feature is to use (10) in Section 4 (or the equivalent, in the case of the other stereo configurations) to compute the intersection of the two lines from each mirror. An approximation can be made by simply estimating each mirror as a point, and then using the angular magnification factor to approximate the direction of the line. This approximation for an $\alpha=k$ mirror is shown in Figure 38.

Other than mathematical convenience, there is no compelling reason to make this approximation since for the major computational cost is in the correlation routines, not in the computation of 3D position once the match is found. The result of ignoring the geometry of the mirror in triangulation in addition to in correlation is shown in Figure 39. Terrain maps generated using this method are shown in Figure 40.

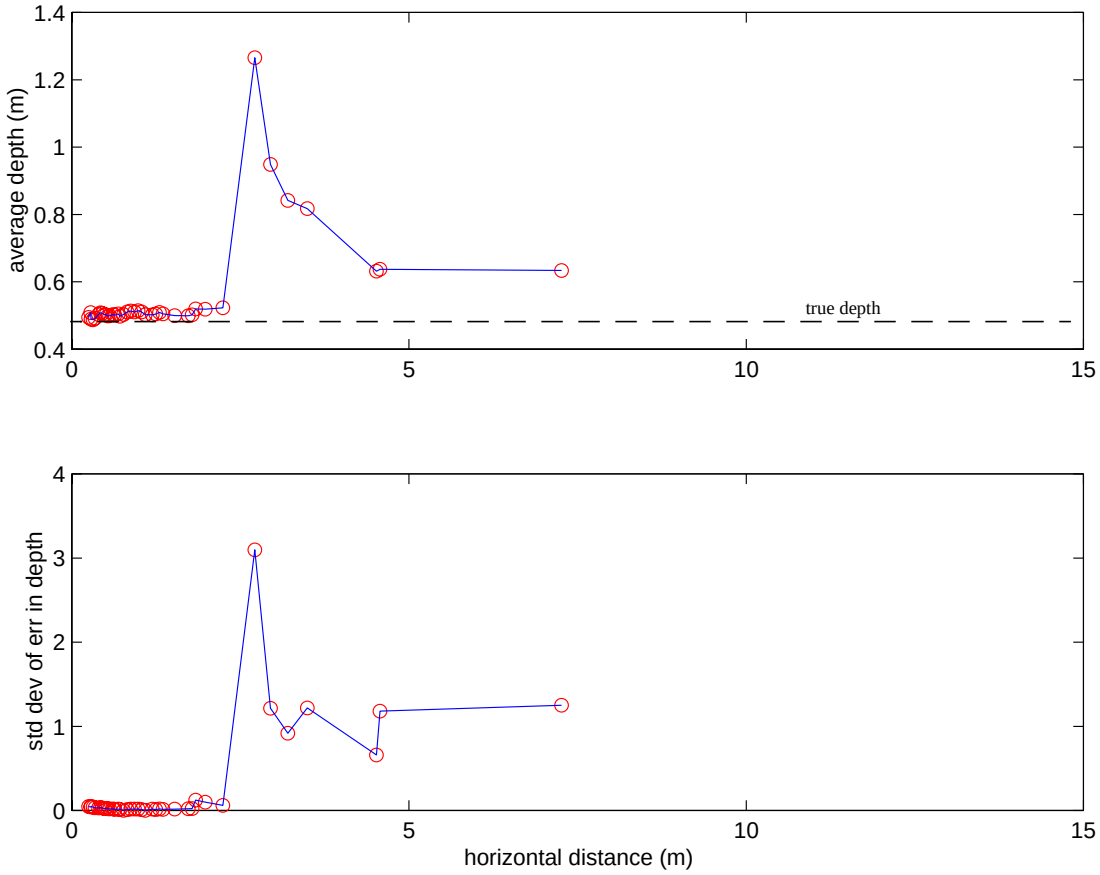


Figure 36 Reconstruction of the terrain using configuration 1. Correlation is done with the unwarped images but triangulation is done with proper consideration of mirror geometry (a) A 2D profile of the terrain along the radial distance. (b) Standard Deviation in the depth error which is typically due to the differences in texture.

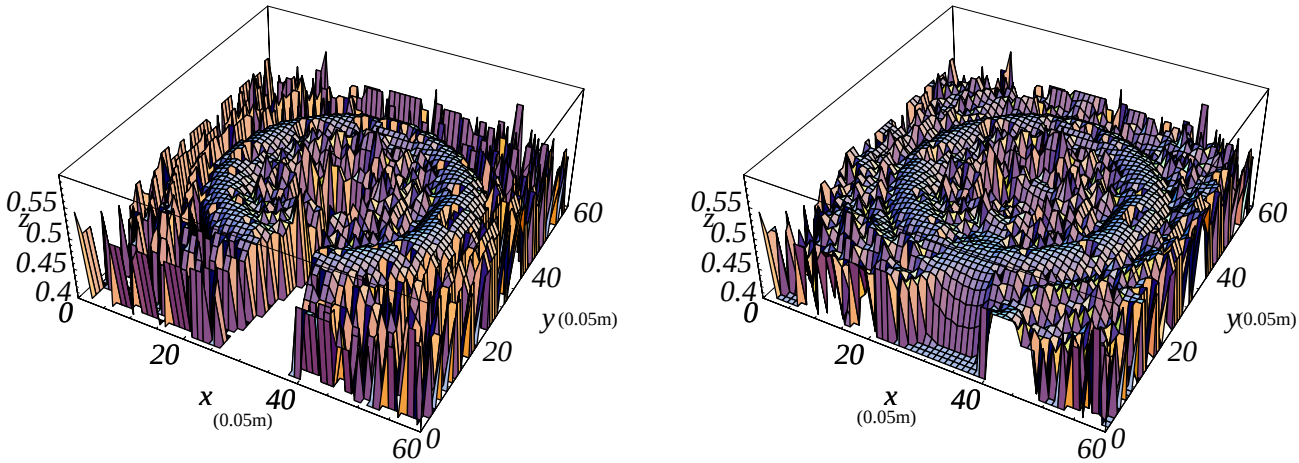


Figure 37 Terrain maps generated from the images in Figure 35. (a) Raw terrain map. The missing section is an artifact due to the correlation window— each end of the unwarped image is clipped by half the width of the correlation window. (b) Interpolated terrain map.

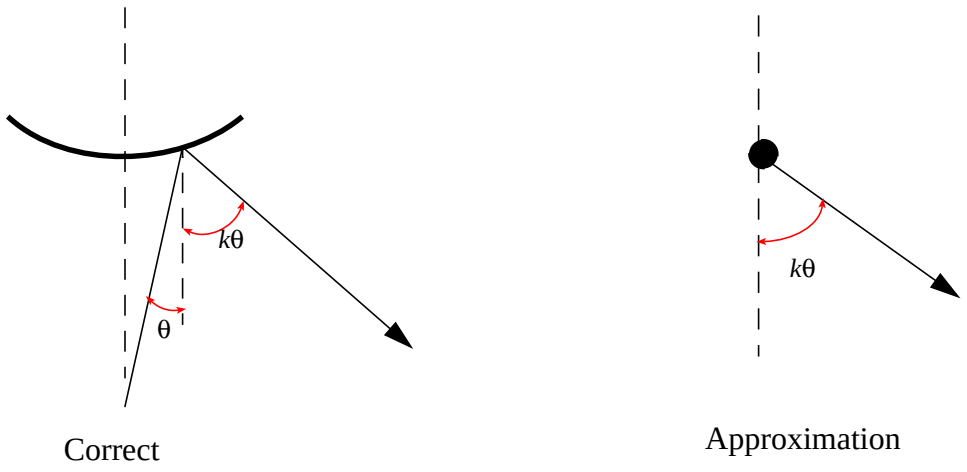


Figure 38 Approximation vs. exact solution of point imaged in curved mirror.

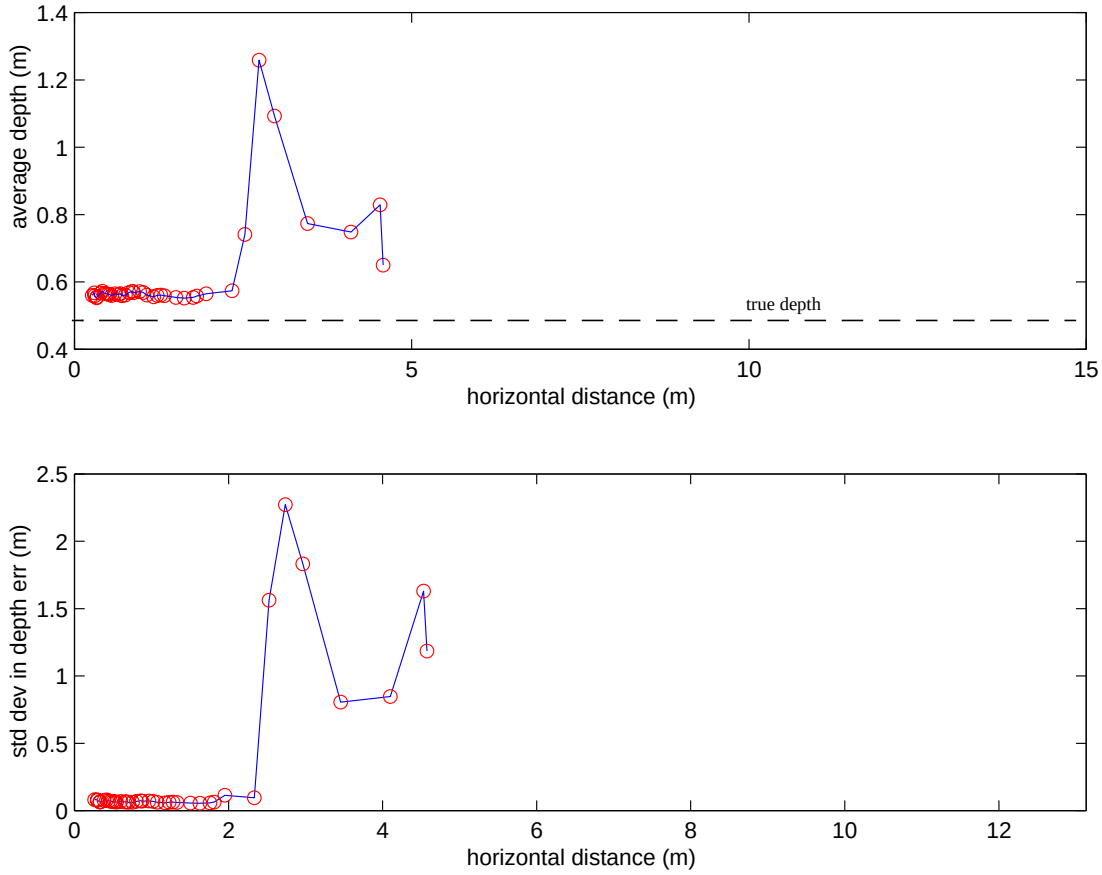


Figure 39 Reconstruction of the terrain using configuration 1. Correlation and triangulation both use the unwarped images from Figure 35 (a) A 2D profile of the terrain along the radial distance. (b) Standard Deviation in the error due to the differences in texture.

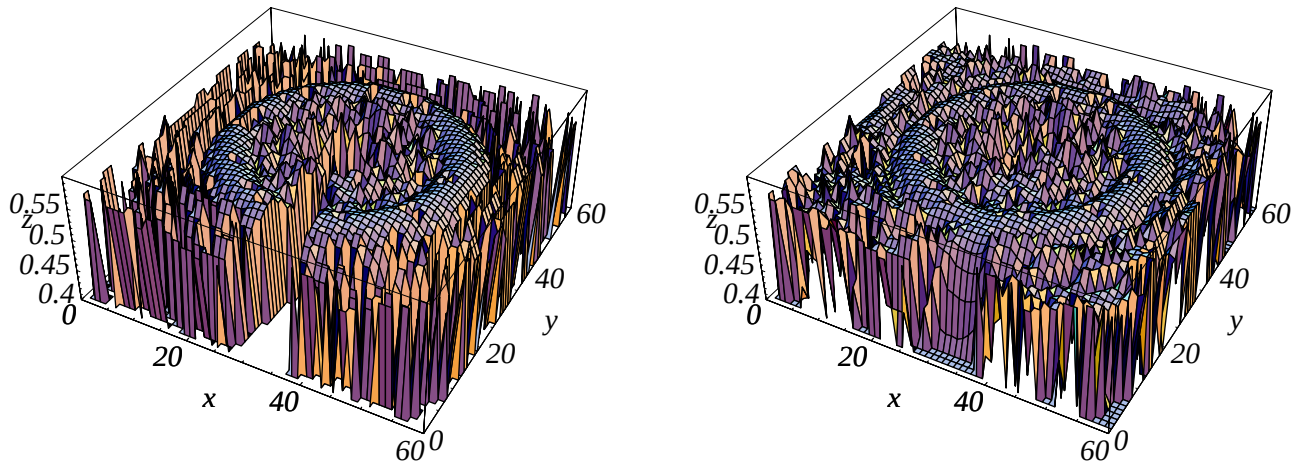


Figure 40 Terrain maps generated from the images in Figure 35. In this case both correlation and triangulation is done using the unwarped images. (a) Raw terrain map. The missing section is an artifact due to the correlation window— each end of the unwarped image is clipped by half the width of the correlation window. (b) Interpolated terrain map.

4.3. Stereo Algorithms: Caveats

Stereo vision is an inherently circular process. We would like to know the shape and of the world and distance to points in the world. To compute these precisely, it is necessary to warp the correlation windows based on, in part, the shape of the world, one of the properties we are trying to compute in the first place! As mentioned above, most stereo algorithms make the assumption (often implicitly) that the surface being computed is normal to the camera. In the results above, for the purposes of warping the correlation windows, we assume that the world is flat and level. In effect we are using perfect knowledge of the shape (the surface normals), although not the distance to the world. If the true shape of the world is approximately level and flat (as in most outdoor environments), the recovered shape is expected to approximate the true shape. If the world is fully three dimensional, we would expect severely degraded results in the regions where the true surface normals vary significantly from the assumed surface normals. One way to recover from such problems is to iterate. That is, one set of range data can be used to approximate the surface normals used to compute the warping of the correlation window.

One reason that there is such a noticeable difference between the performance of the methods that work directly on the warped image vs. the unwarped images is that in the former case, correlation is done in the depth space, effectively resulting in sub-pixel accuracy. That is, proper consideration of the mirror geometry allows a search in the euclidean space. The correlation window is moved along the epi-polar line based on a step taken in the euclidean space. In our implementation, when the unwarped images were used, the correlation was done only the image space, without an attempt to get sub-pixel accuracy.

5. Other uses of panoramic imaging

5.1. Attitude Determination

We also show an example of the effect of rotating the camera in Figure 41 and Figure 42. Figure 41 shows the two orientations of the camera. The first image in Figure 42 shows a panoramic image with the camera pointed vertically. Then we took another image with the camera tilted 45 degrees to the left. This image is shown as the second image in Figure 42. It can be seen from the two images that the features in the direction of the tilt are moved radially. This fact can be used to determine the direction of the tilt. In addition, the amount that the features move in the radial direction can be used to determine the amount of the tilt. Therefore, by tracking features in the image and determine the direction and the magnitude of their movement, change in orientation can be computed.

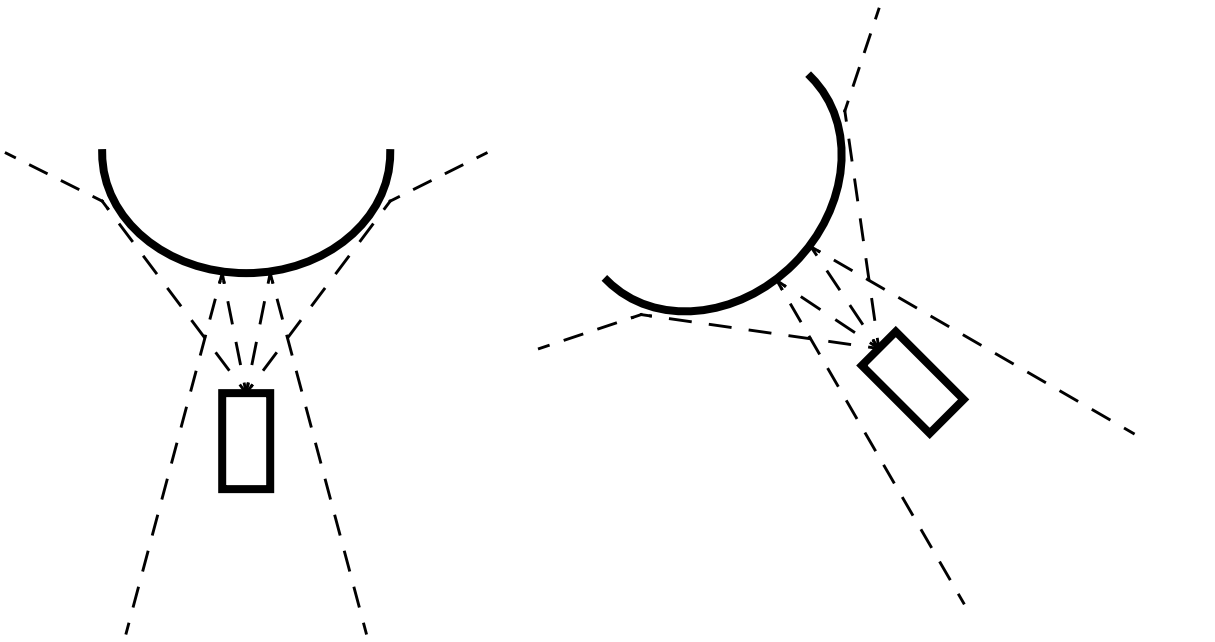


Figure 41 The top image in Figure 42 is taken with the left configuration. The bottom image in Figure 42 is taken with the right configuration.

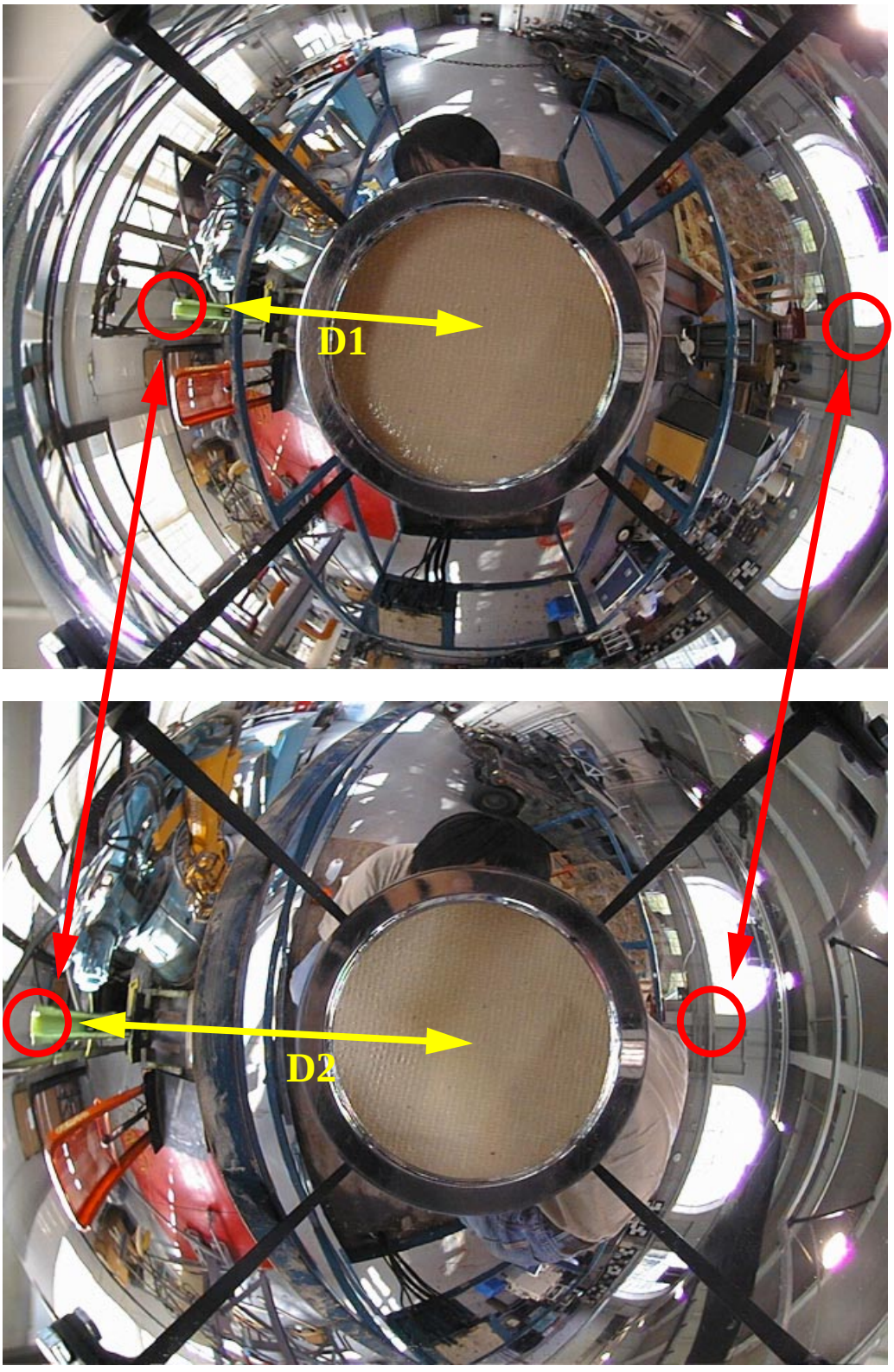


Figure 42 The top image is taken with an upright orientation. The bottom image is taken when the camera is tilted sideway by 45 degrees. Attitude can be estimated by tracking distinct features across sequences of images.

5.2. User Assisted Mapping

Figure 43 illustrates how the proposed system would operate. Figure 43(a) shows a plan view of a robot vehicle moving through the environment. As it moves, position tagged panoramic images (such as shown in Figure 43(b-d)) are stored and later presented to a human operator. The operator chooses features in one of the images and these are tracked through a sequence of images, automatically.

This type of operation can be divided into three phases:

- **Input**: This phase involves display of images and selection of distinct features in an image by an operator.
- **Localization**: In this phase the selected features are tracked through a sequence of images and the 3-D pose corresponding to the features is computed.
- **Modeling**: This phase uses the 3D poses corresponding to tracked features to generate a fractal terrain and maps the texture from the images onto the surface of the reconstructed terrain.

The flow chart in Figure 44 illustrates the sequence of steps in reconstructing 3D terrain from a sequence of position tagged panoramic images. The three phases illustrated in Figure 43 are modular and independent enough that any one of them can be replaced with a different set of tasks to create a new functionality. This is illustrated in Figure 45. That is, the system can be modified to reconstruct solid objects instead of terrain by changing just the modeling phase.

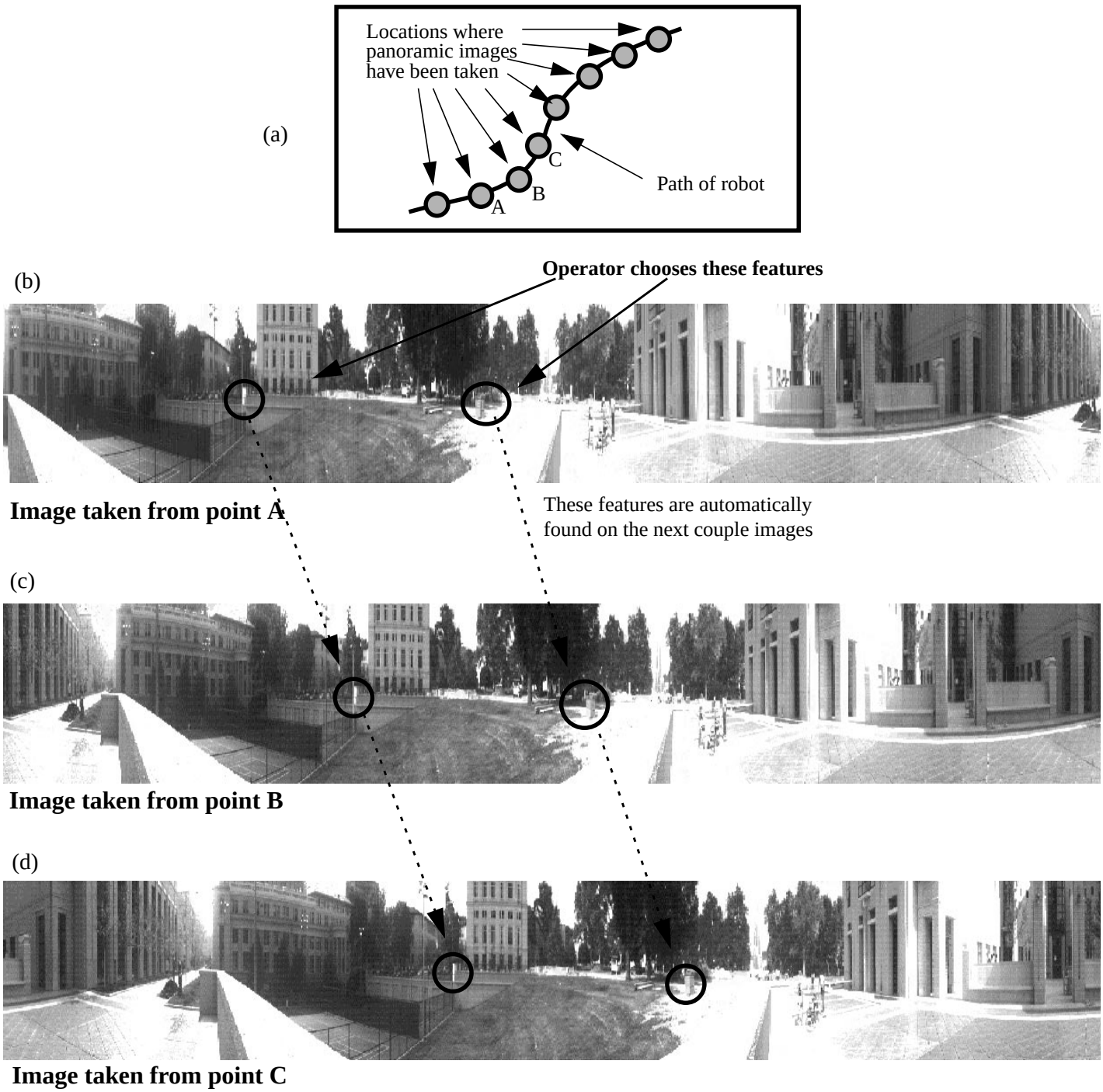


Figure 43 (top) a plan view of a robot moving along a path. (bottom) Tracking of selected features in a sequence of images taken at locations A, B, &C.

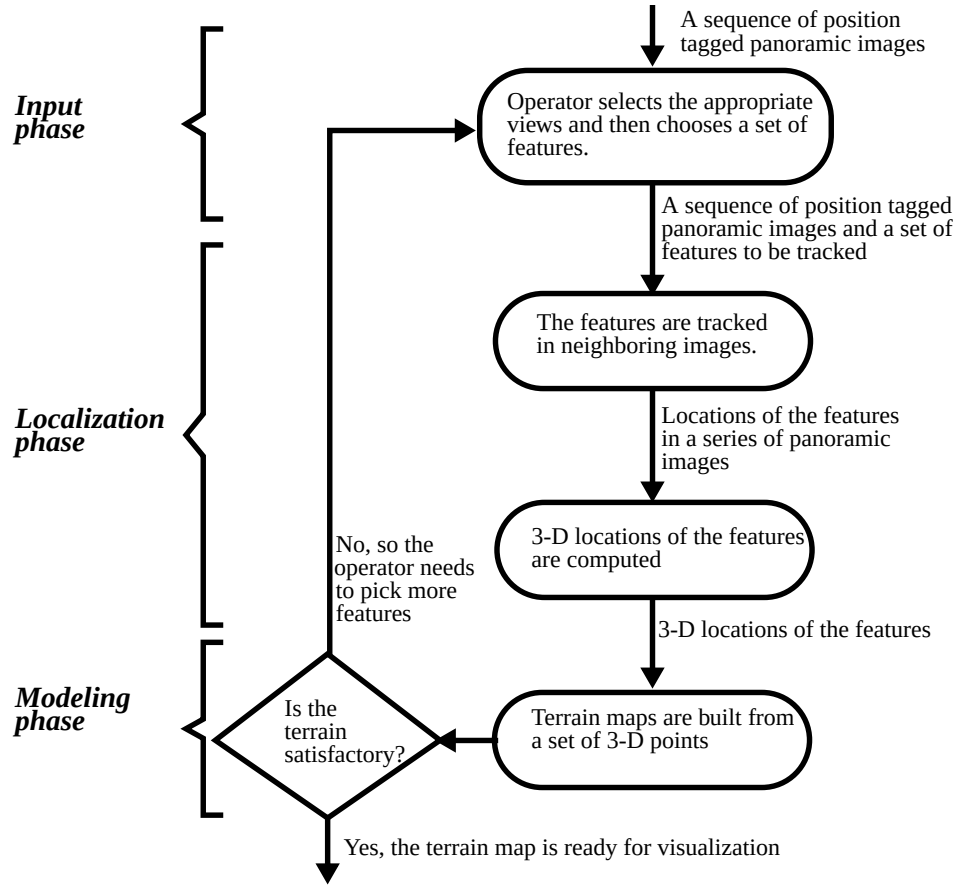


Figure 44 Reconstruction of 3-D terrain topology from a series of panoramic images.

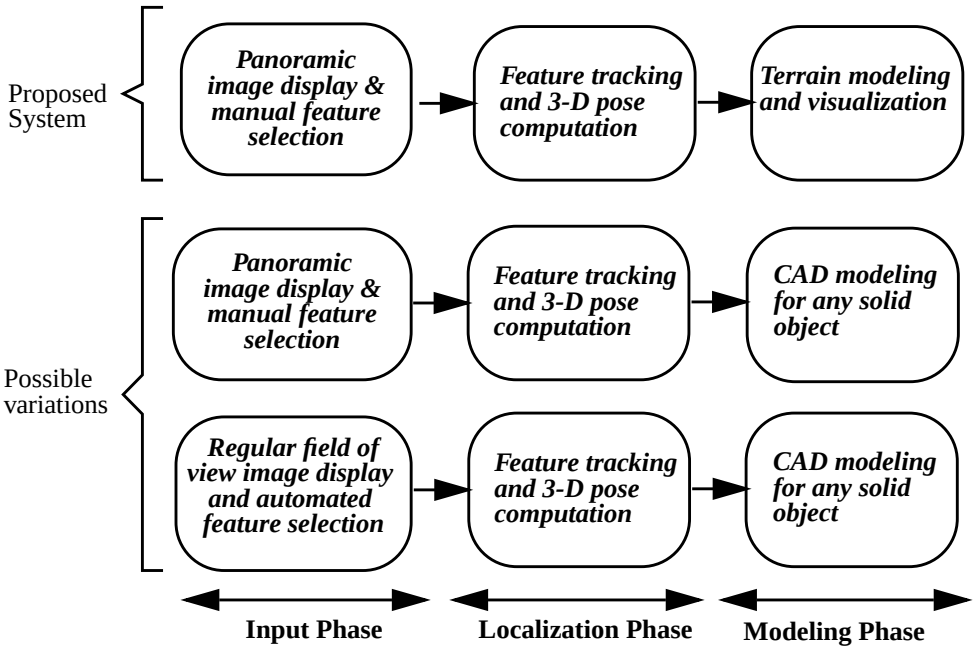


Figure 45 The three processing phases in the proposed system and variations of the basic theme.

6. Conclusions

Our analysis leads us to conclude the following:

- Consideration of the mirror geometry is necessary in both correlation and triangulation. This is most clearly demonstrated by examining the reconstructed surfaces when a flat plane is imaged by stereo cameras. For example, compare Figure 37 and Figure 40 to Figure 13.
- Between correlation and triangulation, mirror geometry is more important in the correlation.
- If range precision is the most important criteria, then a system based on configuration #1 is preferable.
- Configuration 2 is not practical because it is hard to produce a sufficiently large intersecting field of view.
- Configuration 3 is not recommended because of the small baseline involved. This causes significant problems in matching features with high spatial frequency and the geometric precision that is possible is very low (5m of depth precision at 5m of horizontal range)
- The advantage of using one camera is that it is much easier to deploy, is cheaper and is much easier to calibrate. If the precision afforded by these cameras is acceptable, then either configuration #4 or configuration #5 should be chosen.
- #4 is preferable if the objects being imaged are below the horizon. On the other hand if the objects being imaged are above the horizon, then configuration #5 should be used.

References

[1] J. S. Chahl and M. V. Srinivasan, “Reflective Surfaces for Panoramic Imaging,” Applied Optics, Vol 36, No. 31, November 1997.

[2] J. Murphy, “Panospheric Video for Robotic Tele-Exploration,” Ph.D. Thesis, Robotics Institute, Carnegie Mellon University, May 1998.

[3] S. Nayar, “Catadioptric Omnidirectional Camera,” In Proceedings Conference on Computer Vision and Pattern Recognition, June, 1997.

[4] Antarctic Mission Team, “Antarctic Search for Meteorites,” Technical Report CMU-RI-TR-98-109, July, 1998.