Panoramic range estimation using a moving camera and a specially shaped reflective surface

J. S. Chahl and M. V. Srinivasan
Centre for Visual Science
Research School of Biological Sciences
Australian National University
P.O. Box 475, Canberra, A.C.T. 2601, Australia
email: Javaan.Chahl&anu.edu.au, M.Srinivasan&anu.edu.au *

Abstract

A technique is presented for obtaining an omnidirectional range map using a panoramic, moving visual sensor. The sensor consists of a single, standard video camera viewing a specially shaped reflective surface [Chahl and Srinivasan, 1997b]. The surface has the property that a given change in the angular elevation of view in the external environment maps to a constant radial displacement in the camera's image.

Estimation of range is based on the fact that the local, motion-induced deformation of the panoramic image is range-dependent, irrespective of the nature of the deformation. For a translatory motion, the nature of the deformation of the image on a sphere depends upon the the azimuth and elevation of the viewing direction: it is an expansion in the direction of motion, a contraction in the opposite direction, a translation for viewing directions perpendicular to the direction of motion, and a complex deformation along other viewing directions. The range in each direction of view is estimated by comparing the magnitude of the image deformation measured in that direction, with the magnitude of the deformation that would be produced by a reference environment consisting of a "virtual" spherical shell of predetermined radius surrounding the sensor. The local image deformation is measured using a version of the image interpolation algorithm, developed previously in our laboratory [Srinivasan, 1994b].

The technique is capable of delivering omnidirectional range maps in quasi-real time. It has potential applications in surveillance systems, as well as in visual guidance and control of autonomous flying vehicles.

*This work was supported by the Australian Defence Science and Technology Organisation, and the United States Airforce.

1 Introduction

The development of strategies for navigation of autonomous robots can benefit substantially from the use of wide angle visual sensors. Visual estimation of range has been achieved in a number of ways [Jarvis, 1983], many involving the comparison of images resulting from a known displacement of the sensor in some known direction. Numerous algorithms exist for computing range using two paraxial cameras or optical sensor arrays e.g. [Nagle *et al.*, 1993][Goshtasby and Gruver, 1993][Stange *et al.*, 1991]. This is the so-called 'range from stereo' approach. An equivalent approach would be to use a single camera to capture views from two different locations, and infer range from the motions of corresponding features in the two images. This is the so-called "range from motion" approach, [Ma and Olsen, 1990][Zhuang *et al.*, 1994][Ishiguro *et al.*, 1992][Yakimovsky and Cunningham, 1978][Skifstad and Jain, 1989].

In this paper a technique is developed for the estimation of range in panoramic visual scenes that encompass almost the entire surrounding environment.

Panoramic range estimation requires, first of all, a technique for capturing panoramic images. Here we use an omnidirectional visual sensor which enables a single camera to capture a panoramic image at video rates. The sensor consists of a specially shaped, curved mirror and a conventional camera as described in [Chahl and Srinivasan, 1997b], and illustrated in figure 1B. The shape of the curved reflector is such that it provides a linear relationship between the angle of incidence of light onto the surface, and the angle of reflection onto the CCD array with respect to the centre of the array, as does a planar mirror. This property ensures that the camera provides uniform resolution of the environment in the vertical plane, independent of elevational angle [Chahl and Srinivasan, 1997b]. The gradient of the linear relationship can be chosen as desired to produce a larger or smaller field of view. The shape of the profile of the reflecting surface depends upon the desired gradient, which we term the "elevational gain". An ele-

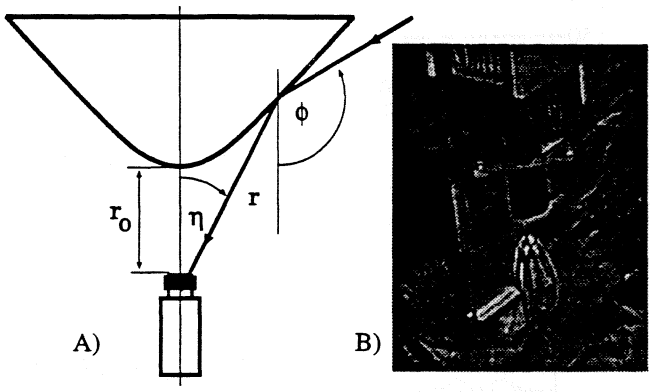

Figure 1: A) Basic metrics of the panoramic imaging system. B) Implementation.

vational gain of 11, for example, provides a field of view of 330° in the vertical plane for a camera with a visual field of 30°. The profile of a reflecting surface with a gain of 11 (ie. $\phi = 11 * \eta$ see figure 1A) is specified by

$$r^6 = \frac{r_0^6}{\cos 4\eta} \quad (1)$$

where r is distance from the camera nodal point, r_0 is minimum distance from the nodal point, and η is angle of view outward from the centre of the image [Chahl and Srinivasan, 1997b].

With this imaging arrangement, a pixel in the camera's image plane, located along an azimuthal direction θ at an angular distance of η from the centre of the image represents the intensity of a point in the environment whose azimuth is θ and whose elevation is $\phi = \eta * k$ where k is the elevational gain of the imaging surface.

The technique presented here delivers omnidirectional range maps by using images captured by a moving panoramic sensor of the kind described above. Range is computed by measuring local image deformations and comparing the magnitude of the deformations that would be produced if the environment were at a known, standard distance. The advantage of measuring complex image deformations rather than simple image translation (local optic flow) is that useful range information can be obtained even in regions of the image where image motion does not consist of a simple translation. If the sensor moves in a straight line, the nature of the local deformation of the panoramic image depends upon the azimuth and elevation of the direction of view: it is an expansion in the direction of motion, a contraction in the opposite direction, a translation for viewing directions perpendicular to the direction of motion, and a complex deformation along other viewing directions. Image deformation is measured by using a version of the image interpolation algorithm [Srinivasan, 1994b], outlined below. The procedure is applicable to all regions of the panoramic image, and it deals with all of the various types of complex image deformation, (expansion in the direction of motion, contraction in the opposite direction, and translation in direction orthogonal to the motion) in a unified way.

2 Theory

Consider that the sensor captures a panoramic image at a particular location. We denote this image by $f_o(\theta, \phi)$. The camera then translates through a distance h before capturing a second panoramic image. We denote this image by $f(\theta, \phi)$. The aim is to compute a panoramic range map, given these two images. To do this, we begin by first calculating what the second image would have been if the environment consisted of a spherical shell of radius R_0 centred on the camera's new location. We assume that the radius R_0 of this "virtual" sphere is smaller than the distance to the nearest object in the real environment (R_0 can always be chosen appropriately to satisfy this). The second image is calculated by considering the motion of the image of each point of the environment, as the sensor moves from the first location to the second within the virtual sphere. It can be shown that the initial azimuth (θ_0) and initial elevation (ϕ_0) of any point on the shell of the virtual sphere are related to the new azimuth (θ_1) and elevation (ϕ_1) by

$$\theta_0 = \cos^{-1}\left(\frac{R_0 \cos\theta_1 \cos\phi_1}{\sqrt{h^2 + R_0^2\cos^2\phi_1 - 2R_0 h\cos\phi_1 \sin\theta_1}}\right) \quad (2)$$

and

$$\phi_0 = \tan^{-1}\left(\frac{R_0 \sin\phi_1}{\sqrt{h^2 + R_0^2\cos^2\phi_1 - 2R_0 h\cos\phi_1 \sin\theta_1}}\right) \quad (3)$$

Using these relationships we can predict what the second image should be, given the first image and the assumption that the environment is a spherical shell of radius R_0. We denote the predicted second image by $f_1(\theta, \phi)$. In reality, the actual second image, f, will be different from the predicted image f_1 because the actual distance of the environment is not R_0 in every direction, as in the case of the virtual sphere.

The image deformations that occur between f_0 and f_1 represent an upper bound on the deformations that can be expected, because we have assumed that the actual distance of the environment from the sensor is at least R_0 units in every direction. The actual range in any viewing direction is given by the ratio of the maximum possible image deformation (i.e the deformation between f_0 and f_1) to the image deformation that actually occurs (i.e the deformation between f_0 and f). If the actual image deformation is small compared to the maximum possible deformation, the range is much larger than R_0; if it is

equal to the maximum possible deformation, the range is R_0; and if it is half the maximum deformation, for example,the range is $2R_0$.

To compute range, therefore, it is necessary to measure the amount of actual image deformation as a fraction of the maximum deformation that can occur along each viewing direction (θ, ϕ). This measurement is done by using the *Image Interpolation algorithm* described below.

2.1 The image interpolation algorithm

The image interpolation algorithm [Srinivasan, 1994b] [Nagle and Srinivasan, 1996] [Chahl and Srinivasan, 1997a] is used to compute the local image deformation due to the motion of the sensor. This algorithm computes the fractional deformation between $f_0(\theta,\phi)$ and $f(\theta,\phi)$, relative to the deformation between $f_0(\theta,\phi)$ and $f_1(\theta,\phi)$. It assumes that, for small displacements of the sensor, $f(\theta,\phi)$ can be approximated by $\widehat{f(\theta,\phi)}$, a weighted linear combination of $f_0(\theta,\phi)$ and $f_1(\theta,\phi)$

$$\widehat{f(\theta,\phi)} = f_0(\theta,\phi) + \alpha\,[f_1(\theta,\phi) - f_0(\theta,\phi)] \qquad (4)$$

where the weighting variable α specifies the fractional deformation of $f(\theta,\phi)$ in going from $f_0(\theta,\phi)$ to $f_1(\theta,\phi)$. α will have a value ranging between 0 and 1.

The value of α for which $\widehat{f(\theta,\phi)}$ best approximates $f(\theta,\phi)$ is determined by minimizing the integral squared error between $\widehat{f(\theta,\phi)}$ and $f(\theta,\phi)$ as follows

$$\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} \Psi(\theta,\phi)\left[\widehat{f(\theta,\phi)} - f_0(\theta,\phi)\right]^2 d\theta d\phi = \int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} \Psi(\theta,\phi)[f_0(\theta,\phi) + \alpha(f_1(\theta,\phi) - f_0(\theta,\phi)) - f(\theta,\phi)]^2 d\theta d\phi \qquad (5)$$

where Ψ is a window function that restricts the area of interest of the error computation to a region centred around the direction in which range is to be determined (typically, the function chosen for Ψ is Gaussian). Minimising this expression with respect to α yields the following expression for α:

$$\alpha(\theta_0,\phi_0) = \frac{\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} \Psi_{\theta_0,\phi_0}(\theta-\theta_0,\phi-\phi_0)\,[f(\theta,\phi) - f_0(\theta,\phi)]}{\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} \Psi_{\theta_0,\phi_0}(\theta-\theta_0,\phi-\phi_0)\,[f_1(\theta,\phi) - f_0(\theta,\phi)]\,d\theta d\phi} \frac{[f_1(\theta,\phi) - f_0(\theta,\phi)]\,d\theta d\phi}{[f_0(\theta,\phi) - f_1(\theta,\phi)]^2\,d\theta d\phi} \qquad (6)$$

The above expression specifies the fractional image deformation in the viewing direction (θ_0,ϕ_0). Note that the shape of the window function Ψ depends on where its centre (θ_0,ϕ_0) is located (see figure 2.2C).

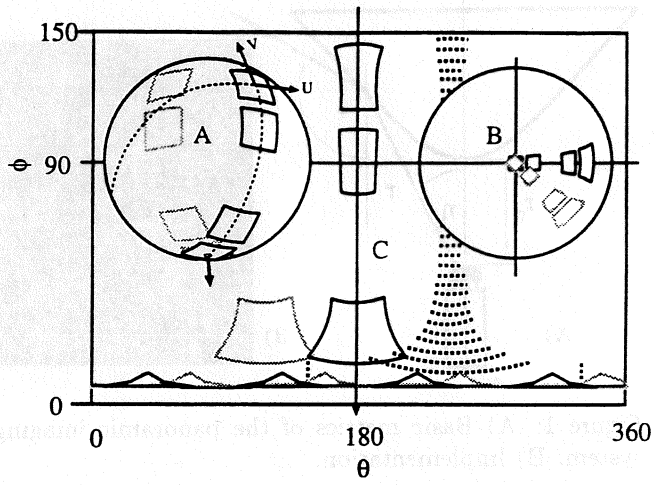

Figure 2: Geometric relationships for filtering panoramic images.

Generally, the images are low-pass filtered before attempting to compute deformation. The aim of this procedure is to reduce aliasing by undesirable high frequency components in the image. The same effect can be achieved by using an optical low-pass filter or even defocusing of the camera lens.

The entire process, including the prefiltering, windowing and image deformation computation can be expressed more compactly as a series of convolutions, which is convenient for implementation on image processing hardware:

$$\alpha(\theta_0,\phi_0) = \frac{\Psi \otimes [\Phi \otimes (f - f_0)] \cdot \Psi \otimes [\Phi \otimes (f_1 - f_0)]}{\Psi \otimes [\Phi \otimes (f_1 - f_0)]^2} \qquad (7)$$

where Φ is the low-pass filter.

This expression yields a "relative image deformation map" which is obtained purely by separable convolutions and arithmetic operations on corresponding pixels. The range in each direction $R(\theta_0,\phi_0)$ is then given by

$$R(\theta_0,\phi_0) = \frac{1}{\alpha(\theta_0,\phi_0)} \qquad (8)$$

2.2 Homogeneous processing of the panoramic image

We have seen above that, in order to compute the range map, it is necessary to spatially filter the external environment in a homogeneous fashion. In other words, if we consider the environment to be mapped on to a "viewsphere", we need to filter the image on the viewsphere homogeneously. The window function, Ψ, should also be applied homogeneously over the viewsphere.

Implementation of the above requirements is not completely straightforward. This is because a given object

will produce an image that differs in size and slope depending upon the viewing direction. Homogeneous filtering is achieved by ensuring that each patch symmetrically straddles the longitudinal and elevational great circles passing through the centre of the patch, as shown in 2.2A. We refer to the directions of these great circles as the principal axes of the filter at that location (V is directed along the local longitude, and U along the orthogonal great circle, see figure 2.2A. If the filter is square, then the boundaries of the filter will also lie on great circles.

Consider a square filter patch of angular width w along each of its principal axes, with its centre at elevation ϕ_c and azimuth θ_c. It can be calculated that the projections of the principal axes U and V of the patch on the image plane are given by: Projection of U axis:

$\theta = \theta_c + \tan^{-1}\left[\frac{\tan(\theta_c+u)}{\cos\phi_c}\right]$

$\phi = \cos^{-1}\left[\cos(\theta_c+u)\sin\phi_c\right]$

Projection of V axis:

$\theta = \theta_c \qquad \phi = \phi_c + v$

where θ and ϕ refer to the pixel locations in the image as specified by polar co-ordinates (θ representing the azimuth and ϕ being proportional to distance from the image centre, or elevation angle in the external environment). u and v are directional angles measured along the principal axes U and V, respectively. These angles each vary from $-\frac{w}{2}$ to $+\frac{w}{2}$. Similar expressions can be derived for the boundaries of the filter patch, because they also lie on great circles.

The projections of these filters on the image plane are shown in figure 2.2B. Filtering the image using the projected shapes will correspond to filtering the external environment homogeneously using square filters of angular width w. This technique achieves a representation that is a locally orthogonal coordinate system.

Another advantage of using filters whose weighting functions are separable along the principal axes, U and V, is that the filtering operation (2-D convolution) can be carried out by two successive 1-D convolutions, first along U and then along V or vice versa. This speeds up the computation of 2-D convolution considerably, as is well known in the case of standard 2-D filtering of perspective projection images [Goodman, 1968].

The situation here, however, is somewhat more complex than with standard 2-D filtering of a perspective projection image, because the convolution cannot directly be performed on the view sphere: it has to be performed in the image plane, where the information resides. Figure 2.2B shows how the square filter patches on the surface of the viewsphere map onto the image plane. These filter functions are clearly not separable along the θ and ϕ axes.

To simplify the analysis, let us consider the image that is obtained by unwarping the camera image and representing it in Cartesian co-ordinates with θ as the abscissa, ϕ as the ordinate. The result is shown in figure 2.2C. It is clear that the width of the filter along the ϕ axis is constant, equal to w. However, the width along the θ axis is variable and depends upon the elevational angle ϕ. The principal axis V of each filter maps onto a line parallel to the ϕ axis, as shown in figure 2.2A,C. But the principal axis U is, in general, curved and of variable length. The curvature and length depend upon the elevation ϕ, as shown by the dotted lines in figure 2.2C. The set of curved line segments represent the mappings of the principal axes of filters whose centres all look in the same azimuthal direction but at different elevations, ranging from the minimum ("straight down", $\phi = 0°$) to the maximum value ($\phi = 150°$) dotted lines, figure 2.2C. Note that the U axis corresponding to the filter that looks "straight down" consists of two vertical line segments, 180° apart in azimuth. This corresponds to a segment of a great circle passing through the "south pole" of the view sphere.

The following procedure ensures that the camera image is filtered homogeneously with respect to the external environment,and in a separable fashion:

(i) Choose a particular azimuth, corresponding to a vertical column of pixels in the unwarped image 2.2C. For each pixel in this vertical column, sum up the values of all of the pixels that are crossed by the projection of the U principal axis that passes through that pixel.

(ii) In a new, Cartesian θ versus ϕ image (which we term image A), set the corresponding pixel value equal to this sum. Do this for each pixel in the column.

(iii) Repeat (i) and (ii) for the column corresponding to each azimuth. This completes image A. This operation represents convolution along the U axis, and the image that is obtained is the result of filtering the external environment (or, equivalently, the image on the viewsphere) uniformly, along the local U axis.

(iv) Create a new image B, in which each vertical (constant azimuth) column is convolved in the vertical ($V-$ or $\phi-$ direction) with a one-dimensional rectangular window function of angular width w. This operation represents the result of filtering image A everywhere along the local V axis.

The image B represents the final result of filtering the external environment (or, equivalently, the image on the viewsphere) homogeneously by a square filter of angular with w along each principal axis. This image can be mapped back onto the image plane of the camera, if desired. The $\theta - \phi$ plane is used to eliminate costly trigonometric computations that would be required if the original camera image was used directly.

The technique described above is not restricted to square filters. It is applicable to any spatial filter whose

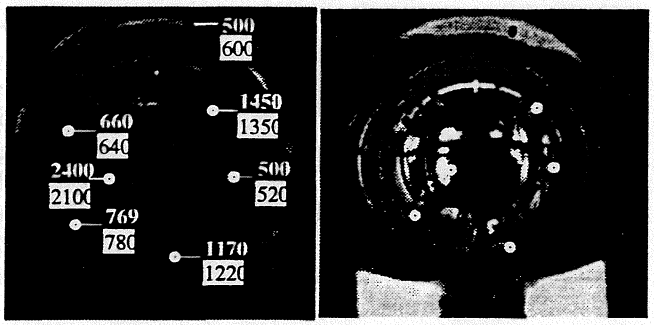

Figure 3: Range Image with ground truth measurements in white text, and computed ranges in black. On the right is the camera image showing the points where range was computed.

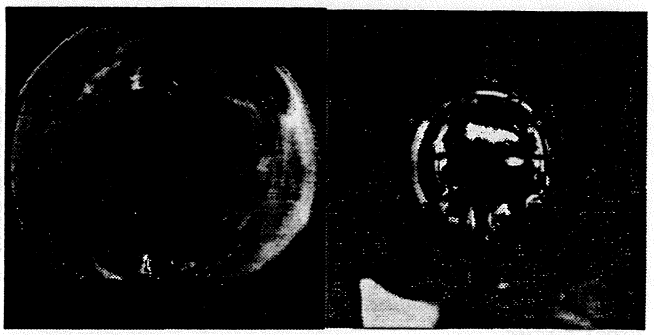

Figure 4: Range image in a scale model of a ravine, and the corresponding camera image on the right.

weighting function is separable along the principal axes U and V. For example, one can implement filtering by a 2-D Gaussian function with its major and minor axes oriented along U and V, by using one-dimensional Gaussians for the convolutions along the U and V axes.

2.3 Physical implementation details

The reflective surface, of elevational gain 11, was machined from aluminium rod stock (2011), and polished with domestic metal polish. The surface was designed to be placed 30mm from the nodal point of the camera.

Alignment of the mirror with the camera is critical if the linear relationship between elevation angle and camera viewing angle is to be maintained, as shown in [Chahl and Srinivasan, 1997b]. To this end a 1mm hole was drilled through the axis of the surface before removing it from the chuck of the CNC lathe. This allowed precise alignment of the axis of the surface, with the axis of the camera.

The surface was suspended above the camera using a single aluminium strut so as to avoid the effects of internal reflection when using the more stylish option of glass or plastic cylinders. The images created by the reflective surface were captured by a conventional $\frac{1}{3}''$ CCIR, CCD camera, and CS mount lens. The field of view of the camera was 30^o, leading to an effective field of view of $30 * 11$ or 330^o for the camera-mirror assembly. The images were digitised by an 8 bit frame grabber inside an Intel Pentium-Pro based workstation.

Precise unidirectional motion control was achieved by mounting the sensor assembly on a robotic gantry. The positioning accuracy of the actuator was better than 0.1mm, and repeatability was better than 0.05mm.

3 Experimental results

The imaging system was moved within the workspace of the gantry. A number of objects were placed at various ranges around the camera. Ground truth ranges were measured using a tape measure, with an estimated accuracy of $\pm 10mm$. The lowpass filter was square in shape, and of width 5^o. The computational window function Ψ was also square and 15^o wide. The sensor was translated "North" by 10mm between the two image acquisitions. The virtual circle radius was 200mm.

Figure 3 shows the results. The largest errors were readings of 600mm at 500mm, and 2100mm at 2400mm. Close objects at intermediate elevation give the most accurate results, probably due to high image velocity.

Figure 4 is a range image and camera image captured from a location within a scale model of a ravine. The walls of the ravine are clearly visible as is the path through the ravine (North).

The speed with which the range maps are computed is limited primarily by the time taken for the two-dimensional convolution operations. Using an Intel Pentium-Pro workstation the processing time for 2-D convolution of a 288x288 pixel image with a filter of width 15^o (16 radial pixels on the image) is 150 msec. Range maps were computed at a rate of 1.0Hz.

4 Discussion

The range measurements in directions close to the axis of sensor motion are less accurate than those in directions perpendicular to the motion, as anticipated. The range resolution provided by the algorithm is limited by the smallest change in image deformation (or displacement) that can be measured above the noise level of the camera.

Sensor motion (or the stereo baseline in the case of a two camera system) can cause a problem common to many forms of visual range-finding; the occlusion of one object in the field of view by another *between* the acquisition of the two views. When the motion/deformation computation is performed on such an image pair the results will be unpredictable, as there is no overlap between the corresponding regions in the two images. Fortunately, the image interpolation algorithm has been shown to degrade gracefully when faced with poor or ill-

conditioned data [Srinivasan, 1994b][Srinivasan, 1994a]. The simplest means of limiting the effect of this phenomenon is to ensure that the motion (or stereo baseline) of the system is small. On the other hand, the minimum size of the baseline is constrained by the level of noise in the system, so that the optimum size of the baseline is governed by a tradeoff.

The information provided by the part of the image facing the direction of motion is of low quality. This is to be expected, as no motion occurs in the regions corresponding to the two poles of the flow field, which are located along the axis of sensor motion. In these situations the algorithm gleans depth information from the edges of the Ψ function. This effect explains some of the irregularities of the range image in directions near the poles of motion.

Major assumptions made in implementing the algorithm were that the nearest object falls outside the virtual sphere, and that the virtual sphere is large compared to the displacement of the sensor. If these conditions are not met, the image interpolation algorithm would be attempting to measure deformations larger than the reference deformation (extrapolation rather than interpolation). In this situation the estimates of deformation, and hence the measurements of range, will become unreliable [Srinivasan, 1994b].

5 Conclusion

The range estimation technique developed here delivers a panoramic range profile that is accurate to within ten per cent for ranges over one hundred times the separation of successive sensor positions. The accuracy, utility and speed of the algorithm suggest applications for visually guided mobile robots, particularly flying vehicles.

References

[Chahl and Srinivasan, 1997a] J. S. Chahl and M. V. Srinivasan. Range estimation using a panoramic visual sensor. *Journal of the Optical Society of America A*, 14(9):2144–2151, September 1997.

[Chahl and Srinivasan, 1997b] J. S. Chahl and M. V. Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics*, 36(31):8275–8285, November 1997.

[Goodman, 1968] J. W. Goodman. *Introduction to Fourier Optics*. McGraw Hill, 1968.

[Goshtasby and Gruver, 1993] A. Goshtasby and W. A. Gruver. Design of a single-lens stereo camera system. *Pattern Recognition*, 26(6):923–937, 1993.

[Ishiguro *et al.*, 1992] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *IEEE transactions on pattern analysis and machine intelligence*, 14(2):257–262, 1992.

[Jarvis, 1983] R. A. Jarvis. A perspective on range finding techniques for computer vision. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 5(2):122–139, March 1983.

[Ma and Olsen, 1990] J. Ma and S. I. Olsen. Depth from zooming. *Journal of the Optical Society of America*, 7(10):1883–1890, 1990.

[Nagle and Srinivasan, 1996] M. G. Nagle and M. V. Srinivasan. Structure from motion: Determining range and orientation of surfaces by image interpolation. *Journal of the Optical Society of America A*, 13(1):25–34, January 1996.

[Nagle *et al.*, 1993] M. G. Nagle, M. V. Srinivasan, and P. J. Sobey. Robust depth extraction for mobile robots. In *Intelligent Robots and Computer Vision XII: Active Vision and 3D Methods*, volume 2056, pages 207–218, Boston, Massachusetts, September 1993. SPIE.

[Skifstad and Jain, 1989] K. Skifstad and R. Jain. Range estimation from intensity gradient analysis. *Machine Vision and Applications*, 2:81–102, 1989.

[Srinivasan, 1994a] M. V. Srinivasan. Generalised gradients versus image interpolation: A critical evaluation of two schemes for measurement of image motion. *Australian Journal of Intelligent Information Processing Systems*, 1:41–50, 1994.

[Srinivasan, 1994b] M. V. Srinivasan. An image interpolation technique for the computation of optic flow and egomotion. *Biological Cybernetics*, 71:401–415, 1994.

[Stange *et al.*, 1991] G. Stange, M. Srinivasan, and J. Dalczynski. Rangefinder based on intensity gradient measurement. *Applied Optics*, 30(13):1695–1700, 1991.

[Yakimovsky and Cunningham, 1978] Y. Yakimovsky and R. Cunningham. A system for extracting three-dimensional measurements from a stereo pair of TV cameras. *Computer Graphics ad Image Processing*, 7:195–210, 1978.

[Zhuang *et al.*, 1994] H. Zhuang, R. Sudhakar, and J. Shieh. Depth estimation from a sequence of monocular images with known camera motion. *Robotics and Autonomous Systems*, 13:87–95, 1994.