Active Stereo Based Surface Reconstruction

Subhodev Das and Narendra Ahuja
Coordinated Science Laboratory and Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Abstract

This paper is concerned with the problem of surface reconstruction from stereo images for large scenes having large depth ranges. The passive stereo paradigm is inadequate for this problem because of the need to aim cameras in different directions and to fixate at different objects. We present an active stereo approach in which the scene is systematically scanned and image acquisition and surface reconstruction are integrated using a four-step process. First, a new fixation point is selected from among the nonfixated, low resolution scene parts of current fixation. Second, a reconfiguration of cameras is initiated for refixation. As reconfiguration progresses, the images of the new fixation point gradually deblur and the accuracy of the stereo estimate of the point improves. In the third step, the improved stereo estimate is used to achieve accurate focus and vergence settings of the cameras for fixation. Finally, focus-based depth estimates are obtained at a grid near the fixation point whose density is determined by the local surface slope. These estimates are fused with those obtained from stereo using maximum likelihood (weighted averaging) and are interpolated to the non-grid points.

1 Introduction

This paper is concerned with the problem of surface reconstruction from stereo images for large scenes having large depth ranges. At any stage of such a surface reconstruction process, sharp images can be acquired only for narrow parts of the visual field, capturing a limited depth range. The high resolution parts of the scene, contained within the depth of field of the cameras, are said to constitute a *central visual field*, while the low resolution parts out of the depth of field, and typically away from the image center, are said to belong to the *peripheral visual field* [Das and Ahuja, 1989]. Accurate surface map is extracted for the central visual field by integrating the use of camera focus, camera vergence, and stereo disparity [Abbott and Ahuja, 1988]. When the entire surface of the fixated object has been scanned, the acquired surface map does not smoothly extend, and therefore surface reconstruction must be resumed by fixating on a new object, selected from the periphery of the current visual field. This presents a dilemma since the exact locations and shapes of "new objects" are unknown (otherwise there would be no need for fixation and subsequent surface reconstruction.)

We present an approach to using coarse structural information about the scene in selecting a new fixation point in the peripheral field and acquiring structural information in the vicinity of the selected point at increasing resolution as the cameras reconfigure and aim at the point. The estimates of the surface structure in the vicinity of the new fixation point, both coarse estimates during refixation and accurate estimates after refixation, are provided by multiple sources including camera focus, camera vergence and stereo disparity. Taking into account their reliabilities, these individual estimates are statistically combined to form a final, overall estimate. Thus the scene information available from the individual sources is dynamically integrated.

Section 2 describes in greater detail the background and motivation behind the work reported in this paper. Section 3 presents an algorithm that performs two types of integration. First, it achieves fusion of depth information available from different sources of scene depth to derive accurate central surface maps. Second, it interleaves coarse-to-fine acquisition of stereo images with their analysis for coarse-to-fine surface reconstruction. Section 4 gives details of implementation and the experimental results.

2 Background and Motivation

In this section we summarize the past research related to the work reported in this paper, and the motivations that lead to the development of the approach described in the following sections.

2.1 Background

This paper pursues the basic theme of active, intelligent data acquisition [Bajcsy, 1985, Bajcsy, 1988]. Computational active vision has become more feasible in the recent years with the availability of sophisticated hardware for controlling imaging elements [Ballard, 1989, Burt, 1988, Clark and Ferrier, 1988, Krotkov, 1987]. In their analysis of surface reconstruction from stereo

This research was supported in part by the National Science Foundation under grant IRI-89-11942, Army Research Office under grant DAAL 03-87-K-0006, and State of Illinois Department of Commerce and Community Affairs under grant 90-103.

images, Marr and Poggio [Marr and Poggio, 1979] also point out the role of eye movements in providing large relative image shifts for matching stereo images having large disparities, thus implying the need for active data acquisition. Ballard and Ozcandarli [Ballard and Ozcandarli, 1988] point out that the incorporation of eye movements radically changes (simplifies) many vision computations; for example, the computation of depth near the point of fixation becomes much easier. Aloimonos et al [Aloimonos *et al.*, 1987] show that active control of imaging parameters leads to simpler formulations of many vision problems that are not well behaved in passive vision. Geiger and Yuille [Geiger and Yuille, 1987] describe a framework for using small vergence changes to help disambiguate stereo correspondences. Abbott and Ahuja [Abbott and Ahuja, 1988] demonstrate the efficacy of integrating image acquisition and image analysis for a single object, by interleaving the processes of camera vergence and focusing with those of depth estimation from camera focus and stereo disparity. Shmuel and Werman [Shmuel and Werman, 1990] have considered the related problem of surface map generation from multiple viewpoints; they use iterative Kalman-filtering techniques to predict a new camera pose for maximal reduction of uncertainty in depth information. Some recent studies have considered higher level criteria for fixation (called *attention*), e.g., for recognition [Bolle *et al.*, 1990].

2.2 Motivation

Consider the initial state in which one of the objects in a scene is fixated on. Any parts of the scene in the peripheral visual field appear out of focus, with the degree of blur determined by the distance from the fixation point. Stereo analysis of the out of focus peripheral image regions would result in surface estimates which would be inaccurate due to poor localization of features. This creates an ordering on different parts of the scene such that the earlier a part is in the ordering the better is the accuracy of its surface estimate.

Traditionally, the scope of stereo has been restricted to provide accurate depth estimate from sharp images for the parts of the scene corresponding to the beginning of the ordering. However, stereo can also be used to obtain inaccurate estimates for peripheral objects that occur later in the ordering. In fact, the availability of coarse peripheral maps would make it possible to select a new fixation point on a new object. Once the cameras are fixated at the newly selected object, the resolution of the rest of the objects lying in the direction of the selected object also improves. Therefore, as the finest stereo reconstruction is achieved for the selected object, the accuracy of the surface information available for those other objects which are now closer to the fixation point also improves.

The availability of the coarse depth map for the unmapped parts of the scene has advantages other than the ability to select a new fixation point. While moving from one fixation point to the next, the mechanical reconfiguration of the image planes is not instantaneous. Intermediate images are obtained with decreasing blur which may be continuously stereo analyzed to improve the estimate for the new point. In this process, the computational blurring operation is replaced by instantaneous optical blurring. The number of stereo pairs acquired before fixation is achieved would depend on the amount of image plane reconfiguration required. The improved coarse depth estimate from stereo can help in predicting and expediting the search for the best focus axis setting corresponding to the new fixation point and the camera vergence.

The use of both focus and stereo processes motivates fusion of depth information available from both of them, thus improving the accuracy of the final estimate. Stereo-based estimates have errors determined by the feature location and quantization errors. The focus-based depth estimates on the other hand suffer from the depth of field effect of the lens. Using the uncertainty characteristics of each of these sources, the estimates may be combined using an optimal estimator.

3 Algorithm

In this section we describe an algorithm to achieve the desired integration of multiresolution image acquisition and their coarse-to-fine processing and fusion of depth estimates from different sources in processing the central visual field. To describe the algorithm, consider the state wherein a fine surface map has been constructed for the central visual field along with a coarse map for the peripheral visual field with respect to the current fixation point. Then the algorithm for iteratively extending the surface map consists of the following steps: (1) An unoccluded peripheral point whose selection involves minimum lateral movement of cameras and reconfiguration of their image planes is chosen as the new target point; (2) a sequence of images of increasing resolution is acquired and stereo analyzed, thus obtaining surface maps with increasing accuracy, during the time the cameras verge and focus on the new target point; and (3) focus-based depth estimates are obtained at points along a non-uniform grid within the central visual field, these estimates are fused with estimates from stereo by weighted averaging, and the fused estimates are interpolated to the non-grid points to yield a fine map for the central visual field while a coarse surface map is obtained for the peripheral field. These steps of the algorithm are discussed in the following subsections.

3.1 Target Selection

The extension of the surface map resumes by fixating at another object. The availability of the peripheral surface map makes the selection of a new fixation point possible, albeit with limited accuracy, and thus helps to avoid the need for knowing object depths before they are estimated!

Given an approximate surface map in the peripheral visual field, how should we select a fixation point? In [Abbott and Ahuja, 1988] some criteria were identified for selection of a fixation point which were motivated by known characteristics of fixation in human vision as well as computational considerations. We use similar criteria

here. A target point at position $\mathbf{p}$, in a coordinate system fixed with respect to the camera locations, is chosen from the current periphery so that the following weighted average is minimized:

$$E = a_1 \| \mathbf{p} - \mathbf{p}_{CAM} \| + a_2 \| \mathbf{p} - \mathbf{p}_{POF} \| + a_3 A(\mathbf{p}, \mathbf{p}_{POF}) \quad (1)$$

where $\mathbf{p}_{CAM}$ and $\mathbf{p}_{POF}$ denote the locations of camera reference frame and the current point of fixation, respectively; $\| . \|$ is the Euclidean distance norm; and the function A gives the angular separation between two 3D points in the camera reference frame. Candidate target $\mathbf{p}$ must be visible to both viewpoints, and must lie within camera travel limits.

The first term enforces a near-to-far ordering on fixation points. The second term favors selection of an object close to currently fixated object since the closer it is the more accurate the target location information from the peripheral map is. The third term biases the choice of target to scene points which lie in directions close to that of the current fixation point, preventing large angular movements of the cameras between fixations.

3.2 Target Homing

Once a target point has been selected on a new object, the cameras need to be reconfigured to fixate on the point. This involves changing camera orientations and focus axis settings. The process of performing these changs is called target homing, and is attempted using the largest available focal length (f_{focus}). While still focused at the current fixation point, the change to large focal length causes substantial blurring of the new target point. If the point spread function (p.s.f) of a finite aperture lens is modeled by a 2D Gaussian then the spread parameter σ_l of the Gaussian signifies the degree of optical blurring of a defocused point. The parameter σ_l is proportional to the focal length, aperture and the distance of the defocused point from the fixation point. Let the optical blur of the target point at the beginning of the target homing phase have a $\sigma_l = \sigma_{lf}$. The stereo based depth estimate of the peripheral target point is inaccurate due to the optical blurring ($\sigma_l = \sigma_{ls}$) of the peripheral features in the vicinity of the target point during the previous fixation. In addition, a Laplacian of Gaussian ($\nabla^2 G$) having a spread parameter $\sigma = \sigma_{pfl}$ is used to detect these features that results in location errors of the detected features. The Gaussian expressing the optical and computational blurring effects at a given peripheral point has a spread parameter of $\sigma_t = \sqrt{\sigma^2_{ls} + \sigma^2_{pfl}}$.

As the image planes are gradually reconfigured by changing the focus settings, the new target point becomes less and less blurred; the image sequence acquired during the reconfiguration thus comprises a multiresolution (coarse-to-fine) image sequence of the target area. Each pair of optically blurred images is subsampled, reducing the degree of subsampling as images become less blurred (σ_{lf} decreases). Let $H_i \times H_i$ denote the resolution of the sampled images at the ith stage ($\sigma_{lf} = \sigma^i{}_{lf}$) during reconfiguration:

$$\frac{H_1}{M} = \frac{\sigma_t}{n\sigma^1{}_{lf}} \quad \text{and} \quad \frac{H_i}{H_{i+1}} = \frac{\sigma^{i+1}{}_{lf}}{\sigma^i{}_{lf}} \quad (2)$$

where $f_{focus}/f_{stereo} = n, n > 1$. Since the optically blurred images are obtained continuously, the improvement in the stereo-based depth estimate of the target point from the analysis of two consecutive image pairs is significant only when the difference $\Delta\sigma_{lf} = \sigma^i{}_{lf} - \sigma^{i+1}{}_{lf}$ is significant. Let $\Delta\sigma_T$ be the chosen significant value of $\Delta\sigma_{lf}$. The intermediate images in which the blur of the target point is between $\sigma^i{}_{lf}$ and $\sigma^{i+1}{}_{lf} = \sigma^i{}_{lf} - \Delta\sigma_T$ are skipped for stereo analysis.

Features are detected in the stereo images using Nevatia-Babu operator [Nevatia and Babu, 1980]. The surface estimates derived from an image pair at any stage during camera reconfiguration serve as coarse estimates for surface reconstruction from later images acquired with smaller σ_{lf}. This process of coarse-to-fine image acquisition interleaved with surface reconstruction is continued till σ_{lf} reaches a lower bound on σ, σ_{ctl}. Beyond this stage only coarse-to-fine image acquisition is continued until $\sigma_{lf} = 0$.

3.3 Target Fixation

The target homing stage terminates with the cameras oriented such that the estimated target point location falls at the center of each image. The increasing accuracy of stereo estimate obtained during target homing brings the two cameras approximately in focus and sets up the approximate vergence angle for the cameras. In order to focus the cameras accurately, the depth estimate Z_s of the target point is used to establish an interval of focus axis settings $[p_1, p_2]$ symmetric about an axis setting p_0 corresponding to Z_s ($p_1 < p_0$ and $p_2 > p_0$). This interval which corresponds to the depth of field at p_0 is finely quantized and searched for a peak of the focus criterion function, defined as the total squared gradient over a fixation window centered at the target point. As in [Abbott and Ahuja, 1988], we perturb the camera orientations slightly to maximize sharpness of images and the correlation between the area around the target locations (image centers). The resulting camera configuration is used to initiate surface reconstruction for the new object.

3.4 Fusion and Surface Estimation

Stereo images are acquired with a focal length (f_{stereo}) smaller than the one used for estimating depth from focus (f_{focus}) to increase the field of view. The fixation point is in focus in these images. The parts of the scene that are in sharp focus (corresponding to objects that lie within the depth of field) are segmented out [Das and Ahuja, 1989] to define the central field of view while the defocused regions comprise the peripheral field.

Once a stereo image has been segmented into central and peripheral visual fields, the next step is to obtain depth estimates by fusing the focus-based and stereo-based information. To do this, a 2D grid is constructed within the central visual field. The density of the grid is determined by the local slope computed from the available surface estimate and is thus spatially varying. Wherever no any surface estimate is available (e.g. in the newly uncovered parts of the visual field), a default grid spacing of w_{grid} is used. Focus-based depth esti-

mates are obtained at these grid points by aiming the cameras at the corresponding 3D points.

Stereo reconstruction for the high resolution (using an $N \times N$ grid) central visual field takes place using a small value of σ (σ_{ctl}) for the Laplacian of Gaussian ($\nabla^2 G$) feature detector. This σ gives the best trade-off between localization and stability of the detected zero-crossings (features). The surface reconstruction begins with initial surface estimates obtained in three different ways. Parts of the central visual field have highly accurate estimates obtained during high-zoom target homing. Other parts of the central visual field have only coarse estimates available from the previous fixation at which time these parts belonged to peripheral field. Finally, yet other parts of the central visual field may have entered the visual field during refixation and thus do not have any associated estimates; for these parts, the most recent stereo-based depth estimate of the current fixation point from target homing is used as initial estimate.

At each point on the non-uniform grid, depth estimates Z_f and Z_s are computed from focus and stereo, respectively. These two estimates are considered as observed values of two independent measurements $z_s(\sim N(Z, {\sigma_s}^2))$, $z_f(\sim N(Z, {\sigma_f}^2))$ of the true depth Z at this grid point. The variances are related to the uncertainties in stereo and focus-based estimates. The value of σ_s is estimated from the stereo disparity errors due to $\nabla^2 G(\sigma_{ctl})$ and image plane quantization, and that of σ_f is estimated from the depth of field of the lens observed during the Z_f measurement. The hypothesis that z_s and z_f have the same mean is verified by testing that the random variable

$$x = \frac{z_s - z_f}{\sqrt{{\sigma_s}^2 + {\sigma_f}^2}} \tag{3}$$

is $\sim N(0,1)$. This is done by checking that $|X| \leq X_\alpha$ for $(1-\alpha)$ degree of confidence, where X is the observed value of x. If the hypothesis is not acceptable then one of the estimates is used as the final estimate (e.g. the one with smaller variance, or depth closer to that of the fixation point). If the hypothesis is valid, the best (maximum likelihood) estimate of Z is obtained by weighted averaging of Z_s and Z_f:

$$\hat{Z} = \frac{{\sigma_f}^2 Z_s + {\sigma_s}^2 Z_f}{{\sigma_s}^2 + {\sigma_f}^2}. \tag{4}$$

Such fusion of stereo and focus data has also been reported in [Krotkov and Kories, 1988]. The deviation $\Delta_s = \hat{Z} - Z_s$ at each grid point is interpolated to non-grid points and the interpolated values are used to update the stereo estimates at the non-grid points. The result of stereo reconstruction is a high resolution (*fine*) surface map for the central visual field.

A σ_{pfl} larger than σ_{ctl} is used for the peripheral feature detector to introduce smoothing in addition to that caused by optical blurring so that the number of matchable features is small. In addition to smoothing, the periphery is subsampled using an $M \times M$ grid ($M < N$). The effects of blurring and subsampling significantly degrade the accuracy of stereo and lead to a low resolution (*coarse*) surface map for the peripheral visual field.

4 Implementation and Results

In this section we present details and results of implementing our active stereo algorithm on a dynamic imaging system. The system consists of two Cohu 4815 CCD cameras mounted on a stereo platform and equipped with Vicon V17.5-105M motorized zoom lenses. High-precision stepper-motor rotational units are used to control independent pan, tilt and vergence angles. The imaging system is controlled by a Sun Microsystems 3/160 workstation.

4.1 Implementation Details

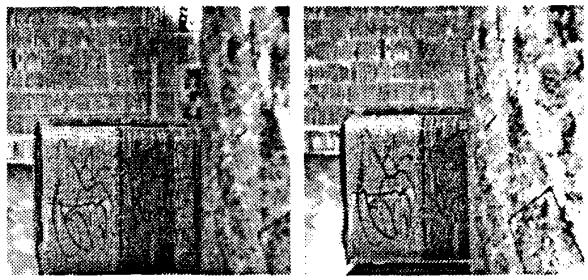
For the left and the right cameras focal lengths (calibrated) of $f_{stereo} = 47.7$ mm and 47.2 mm are used to acquire the stereo images, and $f_{focus} = 105.4$ mm and 101.0 mm (full zoom) are used in the fixation process. The baseline between the cameras is 28 cm. The parameters of (1) are chosen as $a_1 = 0.25$, $a_2 = 0.5$ and $a_3 = 0.25$; $\sigma_{ctl} = 6$ and $N = 256$ for the central visual field; $\sigma_{pfl} = 9$ and $M = 128$ for the peripheral field; $\alpha = 0.05$ and (for $\sim N(0,1)$) $X_\alpha = 1.96$ are used to test the consistency of focus and stereo data; and $\Delta\sigma_T = 3$ is used.

4.2 Experimental Results

The dynamic camera system was made to scan an indoor scene consisting of a vertical barrel (approximately cylindrical) next to a rectangular box, both resting on a flat table top and in front of a rear wall. We present the results after the barrel has been scanned and the cameras are fixated on the box. The stereo images of Figure 1 have the box occupying the central visual field while the barrel and the wall belong to the peripheral field with the barrel being less peripheral (blurred) than the wall. The fine central range map together with the coarse peripheral range map for this fixation is shown for the left viewpoint in Figure 2. A window in Figure 4 marks the newly selected target point on the wall which minimizes (1). The world coordinates of the new target from the coarse stereo depth estimate are (0.106, 0.071, 3.077), all in meters. The measured distance of the target point from the origin of the world coordinates is 3.36 m.

Upon selecting the target the system aims the cameras at it. The focal length of each camera is set to full zoom as required by the fixation process resulting in the optically blurred left and right images of Figure 5, ${\sigma^1}_{lf} = 11$. During the previous fixation ${\sigma_{ls}} = 5.5$ for the new target point and $\sigma_{pfl} = 9$, hence $\sigma_t = \sqrt{{\sigma^2}_{ls} + {\sigma^2}_{pfl}} = 10.5$. These values of ${\sigma^1}_{lf}$ and σ_t when substituted in (2) yield $H_1 = 64$. Nevatia-Babu line extraction algorithm is used to detect features which are matched to obtain the coarse stereo map of Figure 6. The recomputed depth of the target from stereo is 3.174 m. The second pair of optically blurred images that are analyzed are those of Figure 7 for which ${\sigma^2}_{lf} = {\sigma^1}_{lf} - \Delta\sigma_T = 8$ and $H_2 = 128$. The refined depth estimate of the target point is now 3.301 m. The next set of images to be stereo analyzed has ${\sigma^3}_{lf} = 5$. But ${\sigma^3}_{lf} < \sigma_{ctl}$, and the mechanical reconfiguration is therefore continued without stereo analysis until the focus setting corresponding to the depth of

3.301 m has been attained. To compute the focus-based depth estimate of the target, the search interval of focus axis settings is $p_1 = 8499$ and $p_2 = 8967$ (left camera) and $p_1 = 8356$ and $p_2 = 9197$ (right camera). The peak of the focus criterion function is detected at $p_f = 8759$ (left camera) and $p_f = 9075$ (right camera). The focus based depth estimate is 3.287 m, which is used to compute the vergence setting of the cameras. After fixation, the focal length is reduced to $f_s tereo$ to acquire the stereo images which are shown in Figure 8 wherein the wall occupies the central visual field while the barrel and the box belong to the peripheral field. A non-uniform grid is constructed in the central visual field of Figure 8 which corresponds to the projection of a uniform grid on the surface. Then stereo as well as focus-based estimates are obtained at the grid points. These estimates are next fused with the stereo-based estimates as described above. If the two estimates are inconsistent then the one closer in value to the depth of the fixation point is used as the final estimate. The fine map of the wall is added to the composite map in Figure 9 that previously contained only estimates for the barrel and the wall.



(a) (b)

Figure 1: Stereo image pair, (a) left and (b) right, with the box in the central visual field while the back wall continues to occupy the peripheral field.

References

[Abbott and Ahuja, 1988] A. Lynn Abbott and Narendra Ahuja. Surface reconstruction by dynamic integration of focus, camera vergence, and stereo. In *Proc. Second Intl. Conf. on Computer Vision*, pages 532–543, Tarpon Springs, FL, December 1988.

[Aloimonos *et al.*, 1987] John Aloimonos, Isaac Weiss, and Amit Bandyopadhyay. Active vision. In *Proc. First Intl. Conf. on Computer Vision*, pages 35–54, London, UK, June 1987.

[Bajcsy, 1985] Ruzena Bajcsy. Active perception vs. passive perception. In *Proc. Workshop on Computer Vision*, pages 55–59, Bellaire, MI, October 1985.

[Bajcsy, 1988] Ruzena Bajcsy. Perception with feedback. In *Proc. DARPA Image Understanding Workshop*, pages 279–288, Cambridge, MA, April 1988.

[Ballard and Ozcandarli, 1988] Dana H. Ballard and A. Ozcandarli. Eye fixation and early vision: Kinetic depth. In *Proc. Second Intl. Conf. on Computer Vision*, pages 524–531, Tarpon Springs, FL, December 1988.

[Ballard, 1989] Dana H. Ballard. Reference frames for animate vision. In *Proc. 11th IJCAI*, pages 1635–1641, Detroit, MI, August 1989.

[Bolle *et al.*, 1990] Rudd M. Bolle, Andrea Califano, and Rick Kjeldsen. Data and model driven focus of attention. In *Proc. 10th Intl. Conf. on Pattern Recognition*, pages 1–7, Atlantic City, NJ, June 1990.

[Burt, 1988] Peter J. Burt. Algorithms and architectures for smart sensing. In *Proc. DARPA Image Understanding Workshop*, pages 139–153, Cambridge, MA, April 1988.

[Clark and Ferrier, 1988] J. J. Clark and N. J. Ferrier. Modal control of an attentive vision system. In *Proc. Second Intl. Conf. on Computer Vision*, pages 514–523, Tarpon Springs, FL, December 1988.

[Das and Ahuja, 1989] Subhodev Das and Narendra Ahuja. Integrating multiresolution image acquisition and coarse-to-fine surface reconstruction from stereo. In *Proc. IEEE Workshop on Interpretation of 3D Scenes*, pages 9–15, Austin, Texas, November 1989.

[Geiger and Yuille, 1987] Davi Geiger and Alan Yuille. Stereopsis and eye-movement. In *Proc. First Intl. Conf. on Computer Vision*, pages 306–314, London, UK, June 1987.

[Krotkov and Kories, 1988] Eric P. Krotkov and Ralf Kories. Adaptive control of cooperating sensors: Focus and stereo ranging with an agile camera system. In *Proc. IEEE Intl. Conf. on Robotics and Automation*, pages 548–553, Philadelphia, PA, April 1988.

[Krotkov, 1987] Eric P. Krotkov. Exploratory visual sensing for determining spatial layout with an agile stereo camera system. Ph.D. Thesis MS-CIS-87-29, GRASP Laboratory, University of Pennsylvania, Philadelphia, PA, 1987.

[Marr and Poggio, 1979] David Marr and Tomaso Poggio. A computational theory of human stereo vision. In *the Royal Soc. of London, vol. B, no. 204*, pages 301–328, 1979.

[Nevatia and Babu, 1980] Ramakant Nevatia and K. R. Babu. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13:257–269, 1980.

[Shmuel and Werman, 1990] Amir Shmuel and Michael Werman. Active vision: 3d from an image sequence. In *Proc. 10th Intl. Conf. on Pattern Recognition*, pages 48–54, Atlantic City, NJ, June 1990.

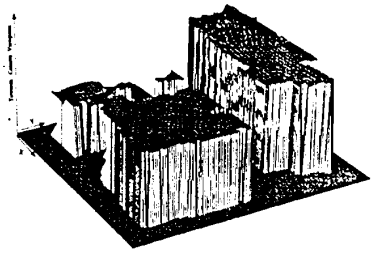


Figure 2: A high resolution range map for the box along with a coarse peripheral map for the wall.

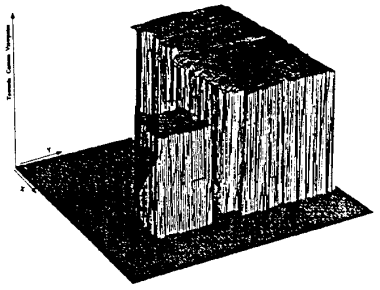


Figure 3: The composite range map is updated by adding the reconstructed surface of the box.

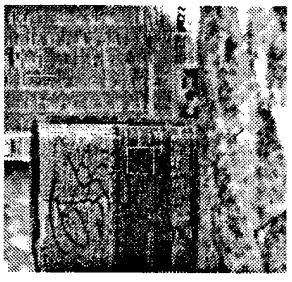
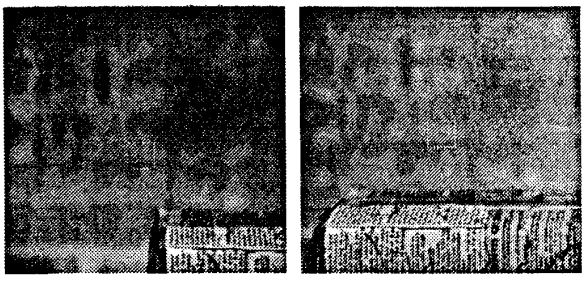


Figure 4: The window marks the new target point on the wall.



(a) (b)

Figure 5: The (a) left and (b) right images as the cameras begin homing on to the target on the wall.

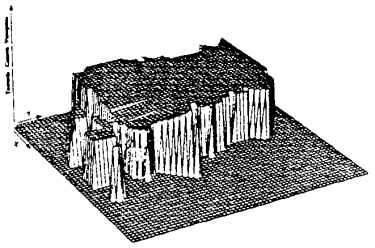
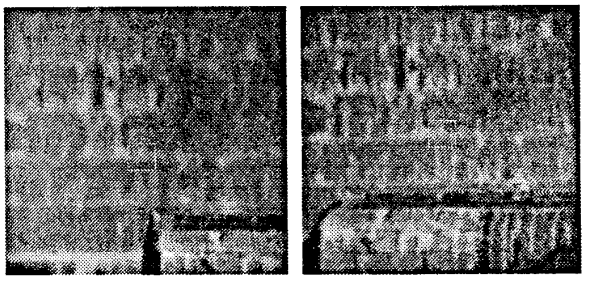
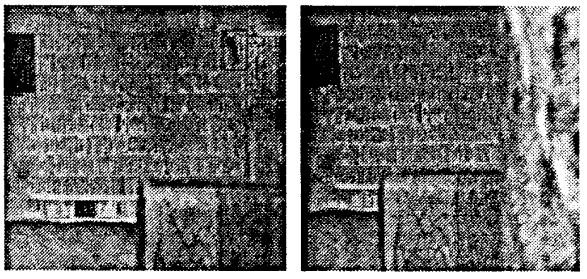


Figure 6: A coarse range map (64 × 64) for the wall around the target point from stereo analysis of the first pair of optically blurred images.



(a) (b)

Figure 7: The (a) left and (b) right images as the image planes are reconfigured to render the target point increasingly sharp.



(a) (b)

Figure 8: Stereo image pair, (a) left and (b) right, when the wall moves into the central visual field. The barrel and the box are now located in the peripheral visual field.

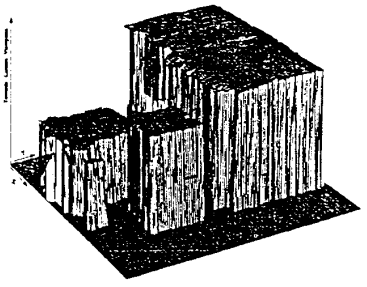


Figure 9: The composite range map after the wall has been reconstructed.