

# Multiresolution Image Acquisition and Surface Reconstruction

Subhudev 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, where it is necessary to aim cameras in different directions and to fixate at different objects. In the past we have reported an approach to acquiring multiresolution surface information. This paper concentrates on the selection of new fixation points from among the nonfixated, low resolution scene parts, and subsequent processing for surface reconstruction. The coarse stereo estimates in the vicinity of the new fixation point are refined as the images of the new fixation point gradually deblur during the process of refixation and are subsequently used to analyze the fixated parts of the scene.

## 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* [8]. Accurate surface map is extracted for the central visual field by integrating the use of camera focus, camera vergence, and stereo disparity [1]. 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.

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.

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.

Section 2 summarizes the past research related to the work reported in this paper. Section 3 presents an algorithm that 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. Section 5 presents concluding remarks.

## 2 Background

This paper pursues the basic theme of active, intelligent data acquisition [3,4]. Computational active vision has become more feasible in the recent years with the availability of sophisticated hardware for controlling imaging elements [1,7,10]. In their analysis of surface reconstruction from stereo images, Marr and Poggio [11] 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 Ozcanarli [5] point out that the incorporation of eye movements radically changes (simplifies) many vision computations. Aloimonos et al [2] 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 [9] describe a framework for using small vergence changes to help disambiguate stereo correspondences. Abbott and Ahuja [1] 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 [13] 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 [6].

## 3 Algorithm

In this section we describe an algorithm to achieve the desired integration of multiresolution image acquisition and their coarse-to-fine processing. 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 re-

spect to the current fixation point. Then the algorithm for iteratively extending the surface map consists of the following steps.

### 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 [1] 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;  $\|\cdot\|$  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.

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 changes 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 the finite aperture lens is modeled by a 2D Gaussian then the spread parameter  $\sigma_l$  of the Gaussian that signifies the degree of optical blurring of the defocused target point is expressed as

$$\sigma_l = k \frac{Af}{Z_{POF}} \left( \frac{|Z - Z_{POF}|}{Z - f} \right) \quad (2)$$

where  $f$  is the focal length and  $A$  is the aperture diameter of the lens,  $Z_{POF}$  and  $Z$  are the depths of the current fixation and the new target points, respectively. The constant of proportionality,  $k$ , is a characteristic of the imaging system. Let the optical blur of the target point at the beginning of the target homing phase have a  $\sigma_l = \sigma_{lf}^1$ , when  $f = f_{focus}$ . The stereo-based depth

estimate of the peripheral target point is inaccurate due to the optical blurring ( $\sigma_l = \sigma_{ls}$ , when  $f = f_{stereo}$ ) 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 the target point has a spread parameter of  $\sigma_t = \sqrt{\sigma_{ls}^2 + \sigma_{pfl}^2}$ .

As the image planes are gradually reconfigured by changing the focus settings and hence  $Z_{POF}$ , 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 ( $\sigma_{lf}$  decreases). Let  $H_i \times H_i$  denote the resolution of the sampled images at the  $i$ th stage ( $\sigma_{lf} = \sigma_{lf}^i$ ) during reconfiguration:

$$\frac{H_1}{M} = \frac{\sigma_t}{n\sigma_{lf}^1} \quad \text{and} \quad \frac{H_i}{H_{i+1}} = \frac{\sigma_{lf}^{i+1}}{\sigma_{lf}^i} \quad (3)$$

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  $\Delta\sigma_{lf} = \sigma_{lf}^i - \sigma_{lf}^{i+1}$  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  $\sigma_{lf}^i$  and  $\sigma_{lf}^{i+1} = \sigma_{lf}^i - \Delta\sigma_T$  are skipped for stereo analysis. This process of coarse-to-fine image acquisition interleaved with surface reconstruction is continued till  $\sigma_{lf}$  reaches a lower bound on  $\sigma, \sigma_{cut}$ . 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 two cameras approximately focused and oriented such that the estimated target point location falls at the center of each image. 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 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 [1], 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 Surface Estimation

Stereo images are acquired with a focal length  $f_{stereo}$  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 are segmented out [8] to define the central field of view while the defocused regions comprise the peripheral field.

Stereo reconstruction for the high resolution (using an  $N \times N$  grid) central visual field takes place using a small value of  $\sigma$  ( $\sigma_{ctl}$ ) for the Laplacian of Gaussian ( $\nabla^2 G$ ) feature detector. 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. The result of stereo reconstruction is a high resolution (*fine*) surface map for the central visual field.

A  $\sigma_{pfl}$  larger than  $\sigma_{ctl}$  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; and  $\Delta\sigma_{lf} = 3$  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. During one of the fixations of the barrel the stereo images of Figure 1 are acquired. Here, the barrel being in focus occupies the central visual field while the box and the back wall constitute the peripheral visual field. 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). A window in Figure 2(b) marks the newly selected target point on the box which minimizes (1).

The coarse stereo depth estimate of the new target is 2.188 m, while the measured distance is 2.18 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 3,  $\sigma_{lf}^1 = 8$ . During the previous fixation  $\sigma_{ls} = 4$  for the new target point and  $\sigma_{pfl} = 9$ , hence  $\sigma_t = \sqrt{\sigma_{ls}^2 + \sigma_{pfl}^2} = 9.9$ . These values of  $\sigma_{lf}^1$  and  $\sigma_t$  when substituted in (3) yield  $H_1 = 64$ . Nevatia-Babu line extraction algorithm [12] is used to detect features which are matched to obtain the coarse stereo map of Figure 4. The recomputed depth the target from stereo is 2.228 m. The next set of images to be stereo analyzed has  $\sigma_{lf}^2 = \sigma_{lf}^1 - \Delta\sigma_{lf} = 5$ . But  $\sigma_{lf}^2 < \sigma_{ctl}$ , and the mechanical reconfiguration is therefore continued without stereo analysis until the focus setting corresponding to the depth of 2.228 m has been attained. To fixate the target, the search interval of focus axis settings is  $p_1 = 5355$  and  $p_2 = 5714$ . The peak of the focus criterion function is detected at  $p_f = 5578$ . The focus based depth estimate is 2.252 m. The stereo images of Figure 5 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 coarse map for the box is now replaced with a fine map as shown in Figure 6(a) which has been added to the composite map in Figure 6(b) that previously contained only estimates for the barrel. In addition, a coarse peripheral map for the wall emerges in Figure 6(a).

## 5 Summary

In this paper we have described our approach to selection of new fixation points during surface reconstruction for large scenes having large depth ranges. The refixation step involves coarse-to-fine mechanical reconfiguration of cameras with gradual deblurring (optical) of the new fixation point during which multiresolution surface reconstruction is performed in parallel with image acquisition. The improved stereo estimates are subsequently used to analyze the newly fixated parts of the scene.

## References

- [1] A. L. Abbott and N. 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.
- [2] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. In *Proc. First Intl. Conf. on Computer Vision*, pages 35-54, London, UK, June 1987.
- [3] R. Bajcsy. Active perception vs. passive perception. In *Proc. Workshop on Computer Vision*, pages 55-59, Belaire, MI, October 1985.
- [4] R. Bajcsy. Perception with feedback. In *Proc. DARPA Image Understanding Workshop*, pages 279-288, Cambridge, MA, April 1988.
- [5] D. H. Ballard and A. Ozcanarli. Eye fixation and early vision: Kinetic depth. In *Proc. Second Intl. Conf. on Computer Vision*, pages 524-531, Tarpon Springs, FL, December 1988.

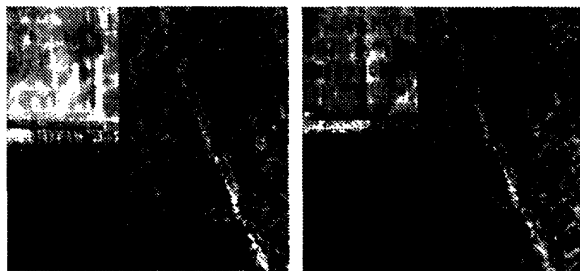


Figure 1: Stereo image pair, (a) left and (b) right, after the fixation of the barrel.

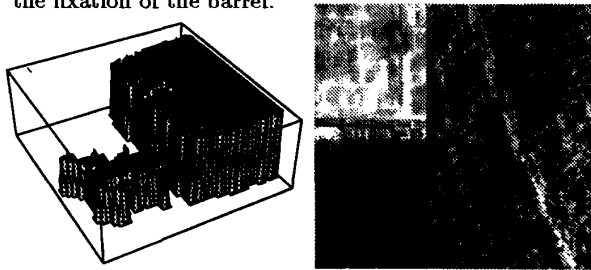


Figure 2: (a) High resolution central range map of the barrel and coarser peripheral range map of the box. (b) A window marks the new target on the box.

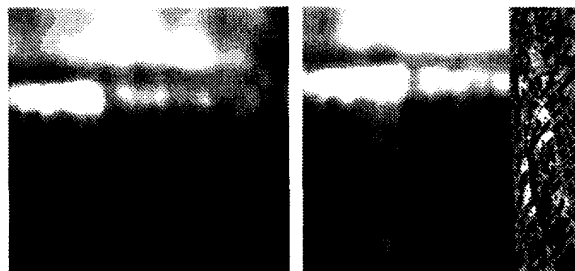


Figure 3: Coarse (a) left and (b) right images at full zoom as the cameras begin homing on the new target.

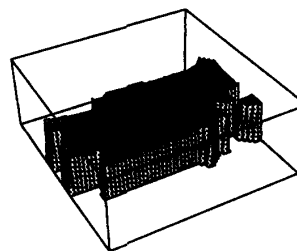


Figure 4: The coarse stereo map in the vicinity of the target point from the optically blurred images (64 x 64).

- [6] R. M. Bolle, A. Califano, and R. Kjeldsen. Data and model driven focus of attention. In *Proc. 10th Intl. Conf. on Pattern Recognition*, pages 1-7, Atlantic City, NJ, June 1990.
- [7] 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.
- [8] S. Das and N. 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.
- [9] D. Geiger and A. Yuille. Stereopsis and eye-movement. In *Proc. First Intl. Conf. on Computer Vision*, pages 306-314, London, UK, June 1987.
- [10] E. 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.
- [11] D. Marr and T. Poggio. A computational theory of human stereo vision. In *the Royal Soc. of London, vol. B, no. 204*, pages 301-328, 1979.
- [12] R. Nevatia and K. R. Babu. Linear feature extraction and description. *Computer Graphics and Image Processing*, 13:257-269, 1980.
- [13] A. Shmuel and M. 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 5: 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.

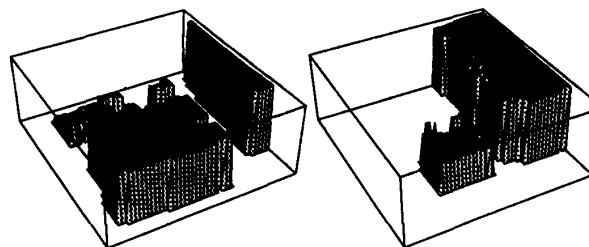


Figure 6: (a) A higher resolution (than in Figure 2) range map for the box that is (b) added to the composite range map.