Range Estimation From Focus Using a Non-frontal Imaging Camera*

Arun Krishnan and **Narendra Ahuja**
Beckman Institute
University Of Illinois
405 North Mathews Ave.,
Urbana IL 61801, U.S.A.
e-mail: arunki@vision.csl.uiuc.edu, ahuja@vision.csl.uiuc.edu

Abstract

This paper is concerned with active sensing of range information from focus. It describes a new type of camera whose image plane is not perpendicular to the optical axis as is standard. This special imaging geometry eliminates the usual focusing need of image plane movement. Camera movement, which is anyway necessary to process large visual fields, integrates panning, focusing, and range estimation. Thus the two standard mechanical actions of focusing and panning are replaced by panning alone. Range estimation is done at the speed of panning. An implementation of the proposed camera design is described and experiments with range estimation are reported.

INTRODUCTION

This paper is concerned with active sensing of range information from focus. It describes a new type of camera which integrates the processes of image acquisition and range estimation. The camera can be viewed as a computational sensor which can perform high speed range estimation over large scenes. Typically, the field of view of a camera is much smaller than the entire visual field of interest. Consequently, the camera must pan to sequentially acquire images of the visual field, a part at a time, and for each part compute range estimates by acquiring and searching images over many image plane locations. Using the proposed approach, range can be computed at the speed of panning the camera.

At the heart of the proposed design is active control of imaging geometry to eliminate the standard mechanical adjustment of image plane location, and further, integration of the only remaining mechanical action of camera panning with focusing and range estimation. Thus, imaging geometry and optics are exploited to replace explicit sequential computation. Since the camera implements a range from focus approach, the resulting estimates have the following characteristics as is true for any such approach [Das and Ahuja, 1990; Das and Ahuja, 1992b]. The scene surfaces of interest must have texture so image sharpness can be measured. The confidence of the estimates improves with the amount of surface texture present. Further, the reliability of estimates is inherently a function of the range to be estimated. However, range estimation for wide scenes using the proposed approach is faster than any traditional range from focus approach, thus eliminating one of the major drawbacks.

The next section describes in detail the pertinence of range estimation from focus, and some problems that characterize previous range from focus approaches and serve as the motivation for the work reported in this paper. The following section presents the new, proposed imaging geometry whose centerpiece is a tilting of the image plane from the standard frontoparallel orientation. It shows how the design achieves the results of search over focus with high computational efficiency. The rest of the paper presents a range from focus algorithm that uses the proposed camera, followed by the results of an experiment demonstrating the feasibility of our method. The last section presents some concluding remarks.

BACKGROUND & MOTIVATION

Range Estimation From Focus and Its Utility

Focus based methods usually obtain a depth estimate of a scene point by mechanically relocating

*The support of the National Science Foundation and Defence Advanced Research Projects Agency under grant IRI-89-02728 and U.S. Army Advance Construction Technology Center under grant DAAL 03-87-K-0006 is gratefully acknowledged.

the image plane, thereby varying the focus distance (v). When the scene point appears in sharp focus, the corresponding u and v values satisfy the standard lens law: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$. The depth value u for the scene point can then be calculated by knowing the values of the focal length and the focus distance [Pentland, 1987; Darrell and Wohn, 1988; Ens and Lawrence, 1991]. To determine when a scene is imaged in sharp focus, several autofocus methods have been proposed in the past [Horn, 1968; Sperling, 1970; Tenebaum, 1971; Jarvis, 1976; Ligthart and Groen, 1982; Schlag *et al.*, 1985; Krotkov *et al.*, 1986; Krotkov, 1986; Darrell and Wohn, 1988; Nayar and Nakagawa, 1990].

Like any other visual cue, range estimation from focus is reliable under some conditions and not so in some other conditions. Therefore, to use the cue appropriately, its shortcomings and strengths must be recognized and the estimation process should be suitably integrated with other processes using different cues, so as to achieve superior estimates under broader conditions of interest [Abbott and Ahuja, 1990; Krotkov, 1989; Das and Ahuja, 1992a]. When accurate depth information is not needed, e.g., for obstacle avoidance during navigation, range estimates from focus or some other cue alone may suffice, even though it may be less accurate than that obtained by an integrated analysis of multiple cues.

Motivation for Proposed Approach

The usual range from focus algorithms involve two mechanical actions, those of panning and for each chosen pan angle finding the best v value. These steps make the algorithms slow. The purpose of the first step is to acquire data over the entire visual field since cameras typically have narrower field of view. This step is therefore essential to construct a range map of the entire scene. The proposed approach is motivated primarily by the desire to eliminate the second step involving mechanical control.

Consider the set of scene points that will be imaged with sharp focus for some constant value of focal length and focus distance. Let us call this set of points the SF surface[1]. For the conventional case where the image is formed on a plane perpendicular to the optical axis, and assuming that the lens has no optical aberrations, this SF surface will be a surface that is approximately planar and normal to the optical axis. The size of SF surface will be a scaled version of the size of the image plane, while its shape will be the same as that of the image plane. Figure 1(a) shows the SF surface for a rectangular image plane.

As the image plane distance from the lens, v, is changed, the SF surface moves away, or toward the camera. As the entire range of v values is traversed, the SF surface sweeps out a cone shaped volume in three-dimensional space, henceforth called the SF cone. The vertex angle of the cone represents the magnification or scaling achieved and is proportional to the f value. Figure 1(b) shows a frustum of the cone.

Only those points of the scene within the SF cone are ever imaged sharply. To increase the size of the imaged scene, the f value used must be increased. Since in practice there is a limit on the usable range of f value, it is not possible to image the entire scene in one viewing. The camera must be panned to repeatedly image different parts of the scene. If the solid angle of the cone is ω, then to image an entire hemisphere one must clearly use at least $\frac{2\Pi}{\omega}$ viewing directions. This is a crude lower bound since it does not take into account the constraints imposed by the packing and tessellability of the hemisphere surface by the shape of the camera visual field.

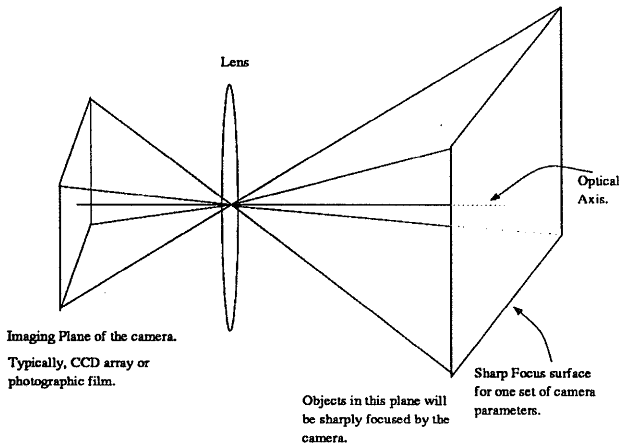
If specialized hardware which can quickly identify focused regions in the image is used, then the time required to obtain the depth estimates is bounded by that required to make all pan angle changes and to process the data acquired for each pan angle.

The goal of the approach proposed in this paper is to select the appropriate v value for each scene point without conducting a dedicated mechanical search over all v values. The next section describes how this is accomplished by slightly changing the camera geometry and exploiting this in conjunction with the pan motion to accomplish the same result as traditionally provided by the two mechanical motions.

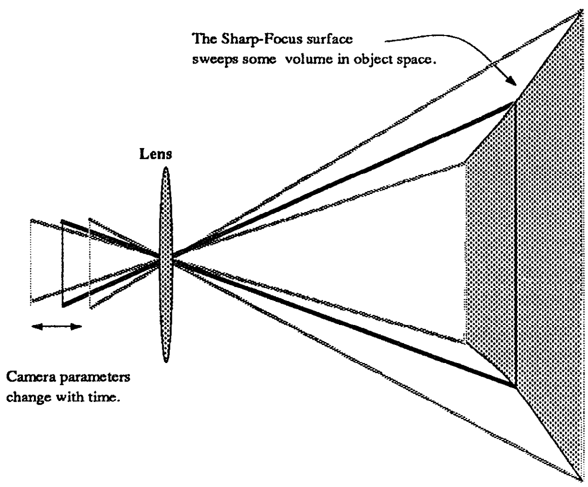
A NON-FRONTAL IMAGING CAMERA

The following observations underlie the proposed approach. In a normal camera, all points on the image plane lie at a fixed distance (v) from the lens. So all scene points are always imaged with a fixed value of v, regardless of where on the image plane they are imaged, i.e., regardless of the camera pan angle. If we instead have an image surface such that the different image surface points are at different distances from the lens, then depending upon where on the imaging surface the image of a scene point is formed (i.e., depending on the pan angle), the imaging parameter v will assume different values. This means that by controlling only the pan angle, we could achieve both goals of the

[1] Actually, the depth-of-field effect will cause the SF surface to be a 3-D volume. We ignore this for the moment, as the arguments being made hold irrespective of whether we have a SF surface, or a SF volume.



(a) SF surface



(b) SF cone

Figure 1: (a) Sharp Focus object surface for the standard planar imaging surface orthogonal to the optical axis. Object points that lie on the SF surface are imaged with the least blur. The location of the SF surface is a function of the camera parameters. (b) A frustum of the cone swept by the SF surface as the value of v is changed. Only those points that lie inside the SF cone can be imaged sharply, and therefore, range-from-focus algorithms can only calculate the range of these points.

traditional mechanical movements, namely, that of changing v values as well as that of scanning the visual field, in an integrated way.

In the rest of this paper, we will consider the simplest case of a nonstandard image surface, namely a plane which has been tilted relative to the standard frontoparallel orientation. Consider the tilted image plane geometry shown in Figure 2(a). For different angles θ, the distance from the lens center to the image plane is different. Consider a point object at an angle θ. The following relation follows from the geometry:

$$| \vec{OC} | = \frac{d \cos \alpha}{\cos(\theta - \alpha)}$$

Since for a tilted image plane, v varies with position, it follows from the lens law that the corresponding SF surface is a surface whose u value also varies with position. The volume swept by the SF surface as the camera is rotated is shown in Figure 2(b).

If the camera turns about the lens center O by an angle ϕ, then the object will now appear at an angle $\theta + \phi$. The new image distance (for the point object) will now be given by the following equation.

$$| \vec{OC'} | = \frac{d \cos \alpha}{\cos(\phi + \theta - \alpha)}$$

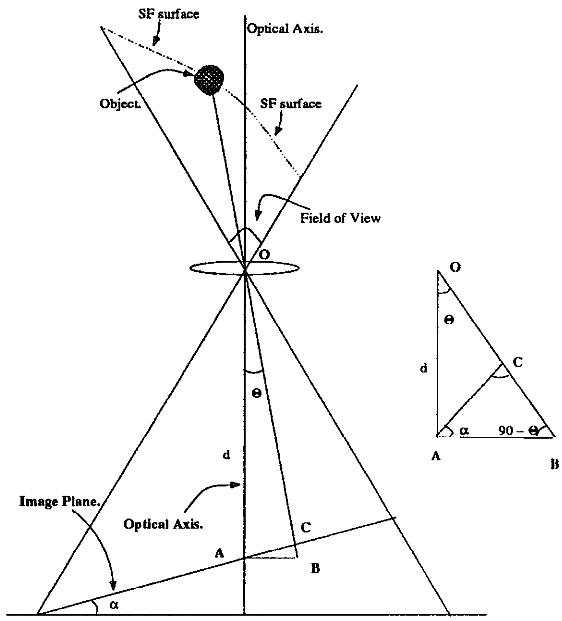
As the angle ϕ changes, the image distance also changes. At some particular angle, the image will appear perfectly focused and as the angle keeps changing, the image will again go out of focus. By identifying the angle ϕ at which any surface appears in sharp focus, we can calculate the image distance, and then from the lens law, the object surface distance.

As the camera rotates about the lens center, new parts of the scene enter the image at the left edge[2] and some previously imaged parts are discarded at the right edge. The entire scene can be imaged and ranged by completely rotating the camera once.

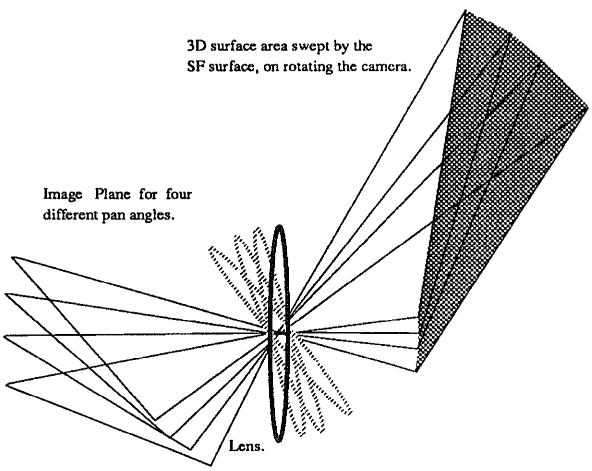
RANGE ESTIMATION ALGORITHM

Let the image plane have N x N pixels and let the range map be a large array of size **N x sN**, where **s >= 1** is a number that depends on how wide a scene is to be imaged. The k^{th} image frame is represented by I_k and the cumulative, environment centered range map with origin at the camera center is represented by R. Every element in the range array is a structure that contains the

[2]Or the right edge, depending upon the direction of the rotation



(a) Tilted Image Surface



(b) SF cone

Figure 2: (a) A point object, initially at an angle of θ, is imaged at point C. The focus distance OC varies as a function of θ. When the camera undergoes a pan motion, θ changes and so does the focus distance. The SF surface is not parallel to the lens and the optical axis is not perpendicular to the SF surface. (b) The 3D volume swept by the proposed SF surface as the non-frontal imaging camera is rotated. For the same rotation, a frontal imaging camera would sweep out an SF cone having a smaller depth.

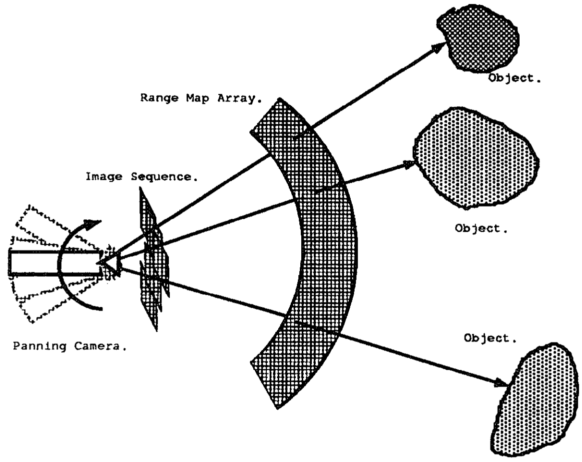


Figure 3: Panning camera, environment fixed range array, and the images obtained at successive pan angles. Each range array element is associated with multiple criterion function values which are computed from different overlapping views. The maximum of the values in any radial direction is the one finally selected for the corresponding range array element, to compute the depth value in that direction.

focus criterion values for different image indices, i.e., for different pan angles. When the stored criterion value shows a maximum, then the index corresponding to the maximum[3] can be used to determine the range for that scene point.

Let the camera start from one side of the scene and pan to the other side. Figure 3 illustrates the geometrical relationships between successive pan angles, pixels of the images obtained, and the range array elements.

Algorithm

Let $j = 0$. $\phi = 0$. Initialize all the arrays and then execute the following steps.

- Capture the j^{th} image I_j.
- Pass the image through a focus criterion filter to yield an array C_j of criterion values.
- For the angle ϕ (which is the angle that the camera has turned from its starting position), calculate the offset into the range map required to align image I_j with the previous images. For example, Pixel $I_j[50][75]$ might correspond to the same object as pixels $I_{j+1}[50][125]$ and $I_{j+2}[50][175]$.

[3] Knowing the index value, we can find out the amount of camera rotation that was needed before the scene point was sharply focused. Using the row and column indices for the range point, and the image index, we can then find out the exact distance from the lens to the image plane (v). We can then use the lens law to calculate the range.

- Check to see if the criterion function for any scene point has crossed the maximum. If so, compute the range for that scene point using the pan angle (and hence v value) for the image with maximum criterion value.
- Rotate the camera by a small amount. Update ϕ and j.
- Repeat the above steps until the entire scene is imaged.

The paper [Krishnan and Ahuja, 1993] contains a pictorial representation of the above algorithm.

EXPERIMENTAL RESULTS

In the experiment we attempt to determine the range of scene points. The scene in experiment 1 consists of, from left to right, a planar surface (range = 73 in), part of the background curtain (range = 132 in), a planar surface (range = 54in) and a planar surface (range = 38 in).

The camera is turned in small steps of 50 units (of the stepper motor), that corresponds to a shift of 15 pixels (in pixel columns) between images. A scene point will thus be present in a maximum of thirty four[4] images. In each image, for the same scene point, the effective distance from the image plane to lens is different. There is a 1-to-1 relationship between the image column number and the distance from lens to image, and therefore, by the lens law, a 1-to-1 relationship between the image column number and the range of the scene point. The column number at which a scene point is imaged with greatest sharpness, is therefore also a measure of the range.

Results Among the focus criterion functions that were tried, the Tennegrad function [Tenebaum, 1971] seemed to have the best performance/speed characteristics. In addition to problems like depth of field, lack of detail, selection of window size etc., that are present in most range-from-focus algorithms, the range map has two problems as described below.

- Consider a scene point, A, that is imaged on pixels, $I_1[230][470]$, $I_2[230][455]$, $I_3[230][440]$ ... Consider also a neighboring scene point B, that is imaged on pixels $I_1[230][471]$, $I_2[230][456]$, $I_3[230][441]$ ... The focus criterion values for point A will peak at a column number that is $470 - n \times 15$ (where $0 \leq n$). If point B is also at the same range as A, then the focus criterion values for point B will peak at a column number that is $471 - n \times 15$, for the same n as that for point A. The peak column number for point A will therefore be 1 less than that of point B. If we have a patch of points that are all at the same distance from the camera, then the peak column numbers obtained will be numbers that change by 1 for neighboring points[5]. The resulting range map therefore shows a local ramping behavior.
- As we mentioned before, a scene point is imaged about 34 times, at different levels of sharpness (or blur). It is very likely that the least blurred image would have been obtained for some camera parameter that corresponds to a value between two input frames.

To reduce these problems, we fit a gaussian to the three focus criterion values around the peak to determine the location of the real maximum. For brevity, we have not included some sample images from the experiments. Figure 4 shows two views of the range disparity values for experiment 1. Parts of the scene where we cannot determine the range disparity values due to a lack of sufficient texture are shown blank. The paper [Krishnan and Ahuja, 1993] contains more experimental results.

SUMMARY AND CONCLUSIONS

In this paper we have shown that using a camera whose image plane is not perpendicular to the optical axis, allows us to determine estimates of range values of object points. We showed that the SF surface, which appears in sharp focus when imaged by our non-frontal imaging camera, is approximately an inclined plane. When the camera's pan angle direction changes, by turning about the lens center, an SF volume is swept out by the SF surface. The points within this volume comprise those for which range can be estimated correctly. We have described an algorithm that determines the range of scene points that lie within the SF volume. We point out some of the shortcomings that are unique to our method. We have also described the results of an experiment that was conducted to prove the feasibility of our method.

References

Abbott, A. Lynn and Ahuja, Narendra 1990. Active surface reconstruction by integrating focus, vergence, stereo, and camera calibration. In *Proc. Third Intl. Conf. Computer Vision.* 489–492.

Darrell, T. and Wohn, K. 1988. Pyramid based depth from focus. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition.* 504–509.

Das, Subhodev and Ahuja, Narendra 1990. Multiresolution image acquisition and surface reconstruction. In *Proc. Third Intl. Conf. Computer Vision.* 485–488.

[4] Roughly $\frac{512}{15}$

[5] Neighbours along vertical columns will not have this problem

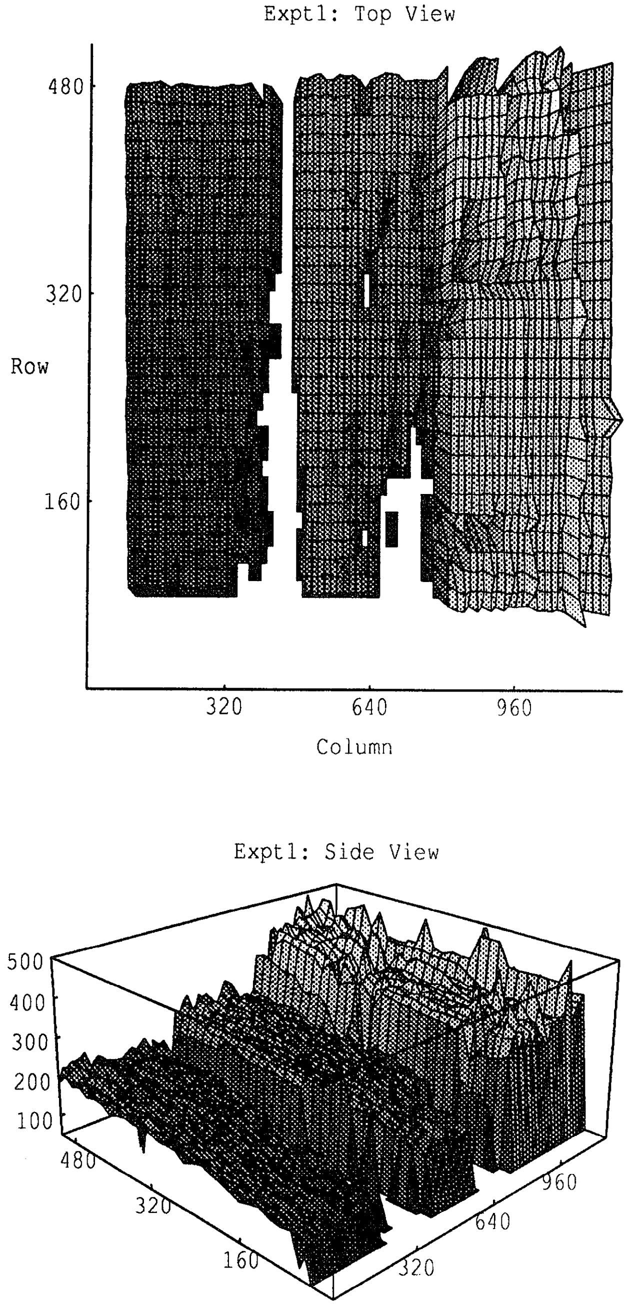


Figure 4: Range disparities for experiment 1. Parts of the scene for which range values could not be calculated are shown blank. The further away a surface is from the camera, the smaller is its height in the range disparity map.

Das, Subhodev and Ahuja, Narendra 1992a. Active surface estimation: Integrating coarse-to-fine image acquisition and estimation from multiple cues. Technical Report CV-92-5-2, Beckman Institute, University of Illinois.

Das, Subhodev and Ahuja, Narendra 1992b. Performance analysis of stereo, vergence, and focus as depth cues for active vision. Technical Report CV-92-6-1, Beckman Institute, University of Illinois.

Ens, John and Lawrence, Peter 1991. A matrix based method for determining depth from focus. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 600–606.

Horn, Berthold Klaus Paul 1968. Focusing. Technical Report 160, MIT Artificial Intelligence Lab, Cambridge, Mass.

Jarvis, R. A. 1976. Focus optimisation criteria for computer image processing. *Microscope* 24:163–180.

Krishnan, Arun and Ahuja, Narendra 1993. Range estimation from focus using a non-frontal imaging camera. In *Proc. DARPA Image Understanding Workshop*.

Krotkov, E. P.; Summers, J.; and Fuma, F. 1986. Computing range with an active camera system. In *Eighth International Conference on Pattern Recognition*. 1156–1158.

Krotkov, E. P. 1986. Focusing. Technical Report MS-CIS-86-22, GRASP Laboratory, University of Pennsylvania.

Krotkov, Eric P. 1989. *Active Computer Vision by Cooperative Focus and Stereo*. New York: Springer-Verlag.

Ligthart, G. and Groen, F. C. A. 1982. A comparison of different autofocus algorithms. In *Proc. Sixth Intl. Conf. Pattern Recognition*. 597–600.

Nayar, Shree K. and Nakagawa, Yasuo 1990. Shape from focus: An effective approach for rough surfaces. In *Proc. IEEE Intl. Conf. Robotics and Automation*. 218–225.

Pentland, Alex Paul 1987. A new sense for depth of field. *IEEE Trans. Pattern Anal. and Machine Intell.* PAMI-9:523–531.

Schlag, J. F.; Sanderson, A. C.; Neuman, C. P.; and Wimberly, F. C. 1985. Implementation of automatic focusing algorithms for a computer vision system with camera control. Technical Report CMU-RI-TR-83-14, Carnegie-Mellon University.

Sperling, G. 1970. Binocular vision: A physical and a neural theory. *Amer. J. Psychology* 83:461–534.

Tenebaum, Jay Martin 1971. *Accomodation in Computer Vision*. Ph.D. Dissertation, Stanford University, Palo Alto, Calif.