International Journal of Computer Vision 20(3), 169–185 (1996)

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
arunki@vision.csl.uiuc.edu
ahuja@vision.csl.uiuc.edu

Received October 28, 1993; Revised November 10, 1994

Abstract. This paper is concerned with active sensing of range information from focus. It describes a new type of camera whose sensor plane is not perpendicular to the optical axis as is standard. This special imaging geometry eliminates sensor plane movement usually necessary for focusing. Camera panning, required for panoramic viewing anyway, in addition enables focusing and range estimation. Thus panning integrates the two standard mechanical actions of focusing and panning, implying range estimation is done at the speed of panning. An implementation of the proposed Non-frontal Imaging Camera (NICAM) design is described. Experiments on range estimation are also presented.

1. 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 panoramic image acquisition and range estimation. The camera can be viewed as a computational sensor which can perform high speed range estimation for 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 comparing images for many sensor plane locations. The proposed camera computes range at the speed of panning.

At the heart of the proposed design is the control of imaging geometry to eliminate the usual mechanical adjustment of sensor plane location, and integrate focusing and range estimation with camera panning. Thus, imaging geometry and optics are exploited to achieve higher computational speed. Since the camera implements a range from focus approach, the resulting estimates have the following characteristics which hold for any such approach (Das and Ahuja, 1990, 1993). 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, the proposed camera makes range estimation from focus practical, by eliminating the drawback of low speed.

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. Section 3 presents the new imaging geometry whose centerpiece is a panning motion about an axis through the optic center in conjunction with tilting of the sensor plane from the standard frontoparallel orientation. It shows how the design achieves focusing with high computational efficiency. Section 4 presents a range from focus algorithm that uses the proposed camera. Section 5 describes a preliminary hardware design of the proposed camera, the implementation of

*The support of the National Science Foundation and Defense 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.

a range estimation algorithm on the camera, and results of experiments demonstrating the camera performance. Section 6 presents concluding remarks.

2. Background and Motivation

This section discusses the characteristics of range estimation from focus, problems with the common range from focus algorithms, and motivation for the proposed approach.

2.1. *Range Estimation from Focus and Its Utility*

Focus based methods usually obtain a depth estimate of a scene point by mechanically relocating the sensor plane, thereby varying the focus distance (v). When the scene point appears in sharp focus, the corresponding u (depth) 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 et al., 1989; Pentland, 1987; Darrell and Wohn, 1988; Ens and Lawrence, 1991).

To determine when a scene is imaged with the least blur, several autofocus methods have been proposed in the past. Horn (1968) describes a Fourier-transform method in which the normalized high-frequency energy from a one-dimensional FFT is used as the criterion function to maximize. Sperling (1970) suggests the squared Laplacian as a measure of image blur which must be minimized. Tenenbaum (1971) uses a thresholded gradient magnitude method in which Sobel operators are used to estimate the gradient. The criterion function used is the sum, over some image window, of the squared gradient magnitudes exceeding a certain threshold. This has also been used by Krotkov (1987) and Krotkov et al. (1986). Jarvis (1976) suggests sharpness measures based on entropy, variance, and gradient. A survey and comparison of several criterion values is presented in Ligthart and Groen (1982). The criterion functions described there make use of such measures as signal power, gray level standard deviation, thresholded pixel counts, and summation of squared gradient in one dimension. Schlag et al. (1985) also implement and compare several auto-focusing methods, including the gradient, Laplacian, and entropy. Darrell and Wohn (1988) describe a depth from focus method that varies the focus distance and uses Laplacian and Gaussian pyramids to obtain the range. Nayar and Nakagawa (1990) use a criterion function where the magnitudes of the second derivatives along individual axes are added.

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. For example, in Abbott and Ahuja (1988, 1993) range estimated from focus is used to identify occlusions and to obtain initial surface estimates for the stereo cue. In Krotkov (1989), range estimates from focus are fused with those from stereo. As another example, consider the case of a scene containing multiple surfaces (Das and Ahuja, 1992). To obtain depth maps of such a scene, objects must be fixated upon, one after another, to acquire suitable image data for each of them. When an object is being analyzed, all the other objects in the vicinity of the current object appear blurred. Further, the variation in the amount of blur observed over any succession of changes in vergence angles helps determine the locations of the nearby objects, with the error in estimation for a particular object given by the magnitude of blur which is proportional to the relative depth of the object. Thus, focus based range estimates help lead the fixation process. Initially, the focus may provide only minimal information about the next object location, e.g., whether it is closer or farther than the current object. This information is used to suitably change the geometric and optical imaging parameters. As the fixation point moves towards the object and the optical parameters change, the estimation accuracy of the nearby object improves. At any stage, an estimate is available which is based on all the information derived from all cues. 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.

2.2. *Motivation for Proposed Approach*

To process large scenes, the usual range from focus algorithms must involve two mechanical (and hence slow) actions, those of panning and for each chosen pan angle finding the best v value. The purpose of the first action is to acquire data for the entire visual field since cameras typically have narrower fields of view.

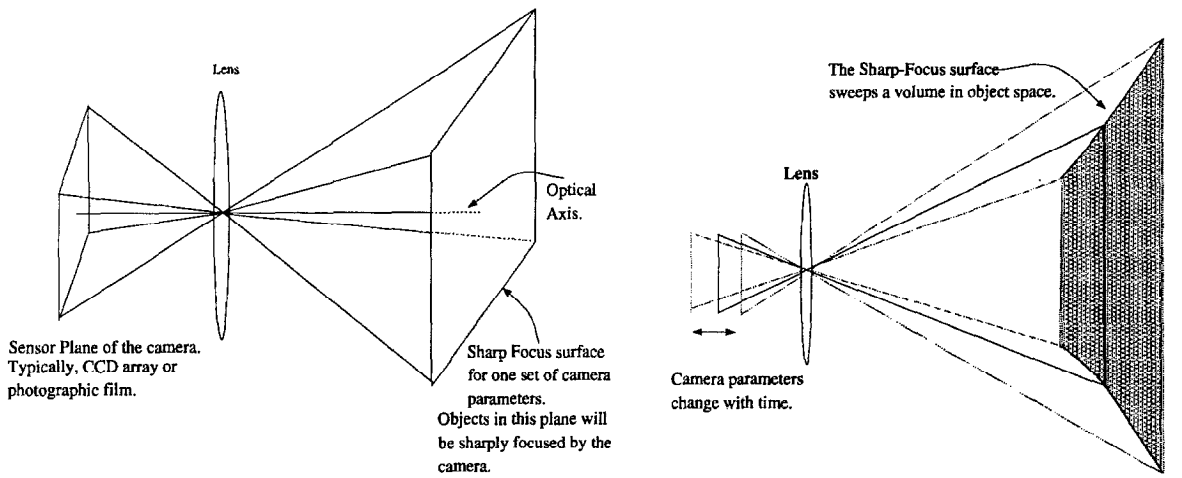


Figure 1. (a) Sharp Focus object surface for the standard planar sensor surface orthogonal to the optical axis. Object points that lie on the SF surface are imaged with the least blur. The location of the SF surface is a function of the camera parameters. (b) A frustum of the cone swept by the SF surface 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.

This mechanical action is therefore essential. The proposed approach is motivated primarily by the desire to eliminate the second action to achieve higher speed (Krishnan and Ahuja, 1993a, b).

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 sensor plane, while its shape will be the same as that of the sensor plane. The magnification or scaling achieved is proportional to the f value. Figure 1(a) shows the SF surface for a rectangular sensor plane.

As the sensor 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. 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 camera must be panned to 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 $\lceil \frac{2\Pi}{\omega} \rceil$ 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 the focused image among those acquired using different v values can be identified in negligible time, for example through the use of specialized hardware, then the time required to obtain the depth estimates is bounded by that required to make all pan angle changes and to acquire the images obtained using different v values for each pan angle.

The goal of the approach proposed in this paper is to estimate the v value yielding the sharpest image of a 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. The result with only one mechanical motion is the same as traditionally provided by both mechanical motions.

3. A Non-Frontal Imaging Camera

The following observations underlie the proposed approach. In a normal camera, all points on the sensor 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 sensor plane they are imaged, i.e., regardless of the camera pan angle. If we instead have a sensor surface such that the different

sensor surface points are at different distances from the lens, then depending upon where on the sensor surface the image of a scene point is formed (i.e., depending on the camera 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 traditional mechanical movements, namely, that of changing v values as well as that of scanning the visual field, in an integrated way. This non-standard configuration of a camera which images each scene point at different sensor surface points which are at various distances from the lens is called a *non-frontal camera* henceforth.

One method of achieving a non-frontal camera is to rotate the lens system in a standard camera about the lens center while keeping the sensor surface fixed. This will cause the optic axis to also rotate and intersect with the sensor surface at different locations and angles. The effective v for points on the sensor surface will therefore change relative to their values for the initial lens orientation, with some v values decreasing and with some increasing. At sensor surface locations which are at large angles to the optic axis, the possible disadvantage with this design is that optical aberations caused by the sensor surface no longer being in the para-axial region will be large. An alternative method of achieving a non-frontal camera that always has the sensor surface in the para-axial region is to tilt the sensor surface relative to the optic axis and form images of the same scene point at different locations along the sensor surface. In the rest of this paper, we will discuss a specific example of such a Non-frontal Imaging Camera (NICAM). We will consider the simplest case of a nonstandard sensor surface, namely a plane which has been tilted relative to the standard frontoparallel orientation. First, in Section 3.1, we will briefly describe this proposed camera geometry. Then we will formally show that this modified camera has an optical transfer function which in fact behaves as expected from the geometry. That is, the optical transfer function varies with image location but this variation is captured completely by the function for the frontoparallel case while using the v value associated with the image location under consideration.

3.1. *Non-Frontal Imaging Geometry*

Consider the two dimensional cross-section of the tilted sensor plane geometry as shown in Fig. 2. Consider an object point at an angle θ from the optical axis. For

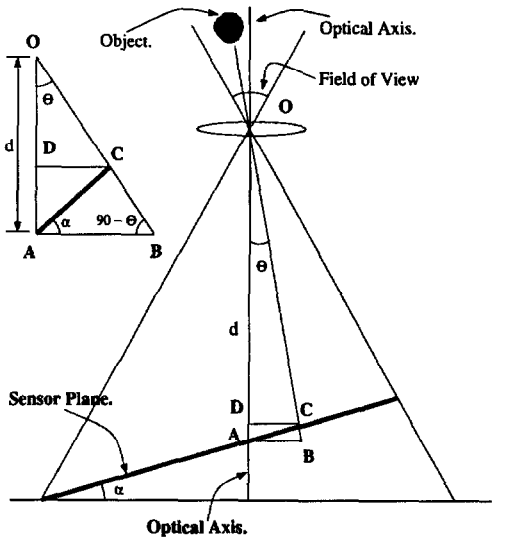


Figure 2. Tilted sensor surface. The sensor plane's normal makes an angle of α w.r.t. the optical axis. Initially, an object point at an angle θ from the optical axis is imaged at point C which is at the focus distance $|\vec{OD}|$. As the camera undergoes a pan motion, θ changes and so does the focus distance.

different angles θ, the distances $|\vec{OC}|$ and $|\vec{OD}|$ from the lens center to the sensor plane are different and are given by

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

Since for a tilted sensor plane, v varies linearly with position along the plane, it follows from the lens law (derived in the next subsection) that the corresponding SF surface is a plane along which the u value mirrors the v variation. The SF surface is shown in Fig. 3(a). The volume swept by the SF surface as the camera is rotated is shown in Fig. 3(b).

If the camera turns about the lens center O by an angle ϕ, then the same object point considered earlier will now be seen at an angle $\theta+\phi$ from the new optical axis. The new image distance for the object point will be given by the equation:

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

As the angle ϕ increases, the image distance also increases. At some particular angle, the image will appear perfectly focused and as the angle further increases, the image will again go out of focus. By identifying the

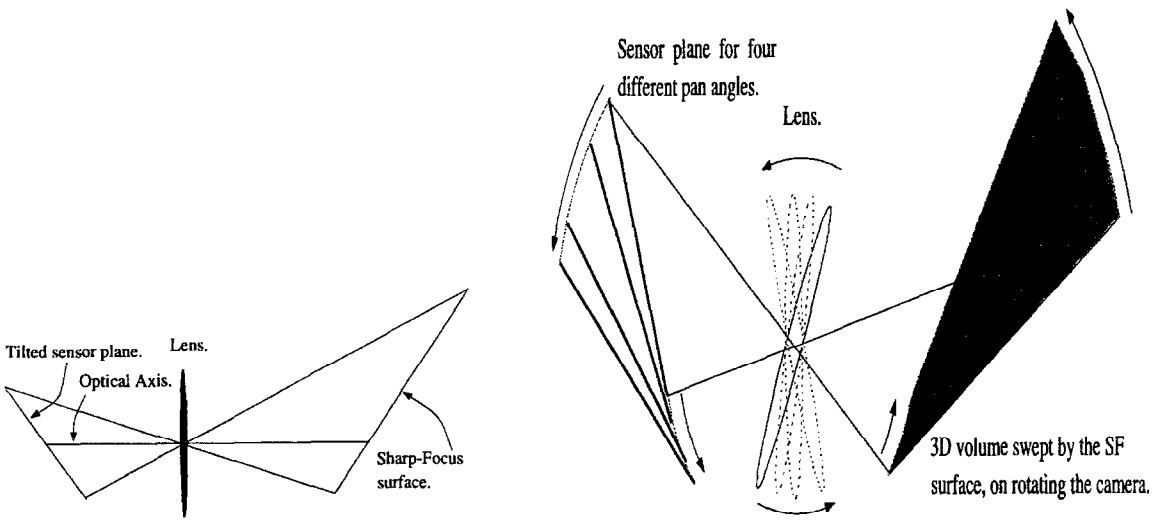


Figure 3. (a) The SF surface for the proposed camera with a tilted sensor plane. The SF surface is not parallel to the lens and the optical axis is not perpendicular to the SF surface. (b) The 3D volume swept by the proposed SF surface as the non-frontal camera is rotated about an axis perpendicular to the optical axis and passing through the lens center (perpendicular to the page in the above figure). For the same rotation, a frontal camera would sweep out a SF cone having a smaller depth.

angle ϕ at which the object point appears in sharp focus, we can calculate the focus distance, and then from the lens law, the object distance.

As the camera rotates about the lens center to increase ϕ, new parts of the scene enter the image at the left edge (or the right edge, depending upon the direction of the rotation) and some previously imaged parts exit at the right edge. The entire 360 degree panoramic scene can be imaged and range estimated by completely rotating the camera once.

We derive the optical transfer function for the modified camera in the next subsection.

3.2. *Derivation of Lens Law*

In this subsection we show that the sharp-focus surface for the modified camera as given by the optical transfer function is a plane inclined to the optical axis[2]. To calculate the optical transfer function, we split the process into the following parts: transformation from a point source on the object to the lens; transformation from one side of the lens to the other side; and finally the transformation from the lens to the tilted sensor plane.

Before we derive the lens law, we shall first derive the transformation that occurs in the propagation of the wave from one plane to another plane that is separated from and inclined to the first plane. We then use this transformation to derive the lens law for the modified camera. The following derivation is an adaptation of the derivation for usual frontal imaging described by Goodman (1968).

Effect of Propagation from One Plane to Another Plane. In this subsection, we consider the effect of propagation of light from one plane, to another plane that is at a mean distance of z, and at an angle of θ as shown in Fig. 4(a).

Using the Hygens-Fresnel principle, the field amplitude at point (x_0, y_0), $\mathbf{U}(x_0, y_0)$, can be written in terms of the field amplitude at point (x_1, y_1), $\mathbf{U}(x_1, y_1)$ as follows.

$$\mathbf{U}(x_0, y_0) = \int\int_{-\infty}^{\infty} \mathbf{h}(x_0, y_0; x_1, y_1)\mathbf{U}(x_1, y_1)\, dx_1\, dy_1$$

where

$$\mathbf{h}(x_0, y_0; x_1, y_1) = \frac{1}{j\lambda} \frac{\exp(jkr_{01})}{r_{01}} \cos(n, r_{01})$$

We will assume that $\mathbf{U}(x_1, y_1)$ is identically zero outside the aperture Σ. Further, we will also assume that the distance z between the aperture and the sensor plane is much greater than the maximum linear dimension of the aperture Σ, and that the region of interest in the sensor plane has a linear dimension much smaller than z. Then the obliquity factor can be approximated by $\cos(n, r_{01}) \cong 1$.

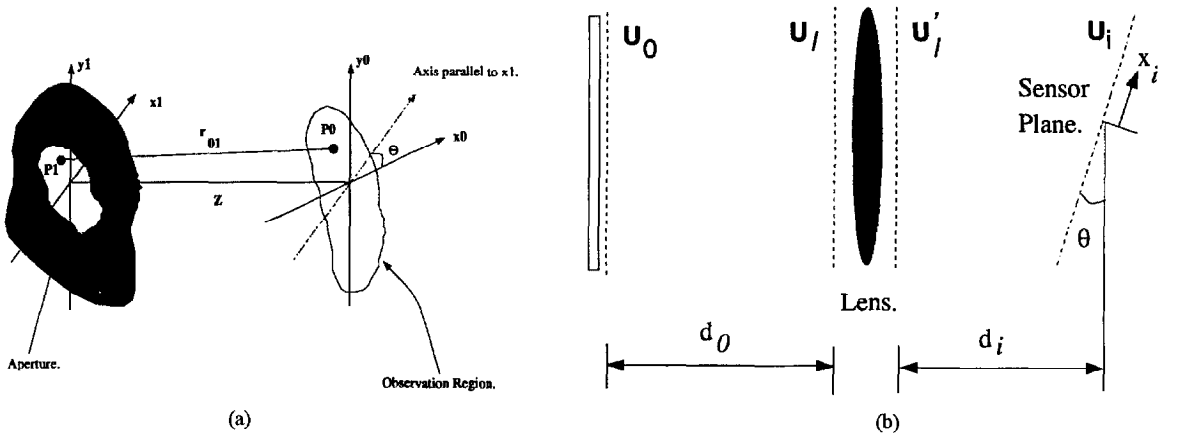


Figure 4. (a) Geometry for image formation (adapted from Goodman, 1968). A planar object with a complex field of $\mathbf{U}_0(x_0, y_0)$ is placed at a distance of d_0 from a positive lens. At a mean distance of d_i, we observe the image field $\mathbf{U}_i(x_i, y_i)$ on a tilted plane. $\mathbf{U}_l(x, y)$ and $\mathbf{U}'_l(x, y)$ represent the complex fields to the left and to the right of the lens, respectively. (b) Diffraction geometry (adapted from Goodman, 1968). The axes y_0 and y_1 are parallel while axes x_0 and x_1 make an angle of θ with one another. The distance between the origins of the two coordinate frames is Z.

Another approximation we make is for distance r_{01}, given by

$$r_{01} \cong (z + x_0 \sin(\theta)) \times \left[1 + \frac{1}{2}\left(\frac{x_0\cos(\theta) - x_1}{z + x_0\sin(\theta)}\right)^2 + \frac{1}{2}\left(\frac{y_0 - y_1}{z + x_0\sin(\theta)}\right)^2\right]$$

Using the above approximations for r_{01} and $\cos(n, r_{01})$, we get

$$\mathbf{U}(x_0, y_0) = \frac{1}{j\lambda}\frac{\exp(jk(z + x_0\sin(\theta)))}{z + x_0\sin(\theta)} \times \int\int_{-\infty}^{\infty} \exp\left[\frac{jk([x_0\cos(\theta) - x_1]^2 + [y_0 - y_1]^2)}{2\lambda(z + x_0\sin(\theta))}\right] \times \mathbf{U}(x_1, y_1)\, dx_1\, dy_1$$

The above equation gives the transformation that occurs on propagation from one plane to another inclined plane. In the next subsection we shall use this result to characterize the transformation from the plane just after the lens to the sensor plane.

Impulse Response of a Positive Lens. In Fig. 4(b) we show a positive lens and the geometry for image formation. d_0 is the distance of the object from the lens, while d_i is the distance of the sensor plane from the lens along the optical axis. The sensor plane makes an angle of $90 - \theta$ with the optical axis. x_i is the distance measure along the x coordinate of the sensor plane. $\mathbf{U}_0(x_0, y_0)$ represents the complex field immediately to the right of the object. $\mathbf{U}_l(x, y)$ and $\mathbf{U}'_l(x, y)$ represent the complex fields to the left and to the right of the lens respectively. $\mathbf{U}_i(x_i, y_i)$ is the complex field that appears on the sensor plane.

The imaged field can be expressed as a function of the object field by the following equation.

$$\mathbf{U}_i(x_i, y_i) = \int\int_{-\infty}^{\infty} \mathbf{h}(x_i, y_i; x_0, y_0)\mathbf{U}_0(x_0, y_0)\, dx_0\, dy_0$$

where $\mathbf{h}(x_i, y_i; x_0, y_0)$ is the field amplitude produced at coordinates (x_i, y_i) by a unit-amplitude point source located at object coordinates (x_0, y_0).

If the object is a point source, then the field incident on the lens, $\mathbf{U}_l(x, y)$, will be a spherical wave diverging from point (x_0, y_0). This can be approximated by

$$\mathbf{U}_l(x, y) = \frac{1}{j\lambda d_0}\exp\left[\frac{jk}{2d_0}[(x - x_0)^2 + (y - y_0)^2]\right]$$

Passage through the lens causes $\mathbf{U}_l(x, y)$ to be multiplied by the pupil function $\mathbf{P}(x, y)$, and the transfer function of the lens. This can be written as

$$\mathbf{U}'_l(x, y) = \mathbf{U}_l(x, y)\mathbf{P}(x, y) \times \exp\left[\frac{-jk}{2f}(x^2 + y^2)\right]$$

Finally, the propagation by a distance of d_i to an inclined sensor plane causes the transformation described in the previous subsection. Substituting for all parts of

the optical transfer function yields the impulse function. For points that satisfy $[-\frac{1}{f}+\frac{1}{d_0}+\frac{1}{d_i+x_i\sin(\theta)}=0]$, the impulse response reduces to

$$\mathbf{h}(x_i, y_i; x_0, y_0) \cong \frac{1}{\lambda^2 d_0(d_i + x_i\sin(\theta))}\int\int_{-\infty}^{\infty}\mathbf{P}(x, y) \times \exp\left[\frac{-jkx\cos(\theta)}{d_i + x_i\sin(\theta)}[x_0M + x_i]\right] \times \exp\left[\frac{-jky\cos(\theta)}{d_i + x_i\sin(\theta)}[y_0M + y_i]\right]dx\,dy$$

where $M = \frac{d_i+x_i\sin(\theta)}{d_0\cos(\theta)}$ is the magnification, and phase terms that do not affect the intensity of the illumination on the sensor plane have been ignored.

For a well-focused optical system, we want the image $\mathbf{U}_i(x_i, y_i)$ to be as similar as possible to the object field $\mathbf{U}_0(x_0, y_0)$. That is, the impulse response should be close to

$$\mathbf{h}(x_i, y_i; x_0, y_0) \cong K\delta(x_i + M_{x_0,y_i} + M_{y_0})$$

From the above equations we see that for points that satisfy $[-\frac{1}{f}+\frac{1}{d_0}+\frac{1}{d_i+x_i\sin(\theta)}=0]$, the impulse response is close to what is desired (within the bounds of the distortion caused by the Franhaufer diffraction) for a well-focused system.

$$\frac{1}{d_0} + \frac{1}{d_i + x_i\sin(\theta)} = \frac{1}{f}$$

is therefore the lens law for the modified camera.

4. Range Estimation Algorithm for Proposed Camera

In this section we describe an algorithm to estimate range at scene points using a NICAM. First we review the factors that affect the performance of any range from focus algorithm.

4.1. *Performance of Shape from Focus Methods*

An object patch is never imaged without any degradation due to diffraction and optical aberrations. Given a set of images, each taken with some variation in camera parameters, shape-from-focus algorithms determine the image having the least amount of degradation. We can then use the known camera parameters to determine the range.

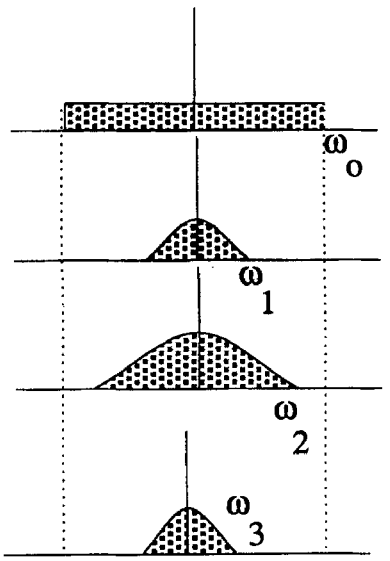
Irrespective of the algorithm used, the estimated range has the following sources of error.

- *Depth of Field (DOF).* For given camera parameters, the DOF will vary as a function of object range. At large values of range, the DOF will be larger and so will be the uncertainty in estimated range.
- *Object Frequency Content.* Blurring or the degradation in imaging can be modeled by a low pass filter. Assuming ideal optics but a finite aperture camera, the diffractive blurring will be in the form of Airy patterns. Defocusing caused by a misplaced sensor plane is modeled in geometric optics by a pill-box model of optical transfer function. This is also a form of low pass filtering ($\frac{J(\rho)}{\rho}$). Therefore, shape-from-focus algorithms must identify the image that has been least degraded. In other words, the image should have been filtered by a low pass filter with the highest cutoff frequency. Clearly, this must be achieved when the frequency content of the object patch is not known, but the frequency contents of the images are available.

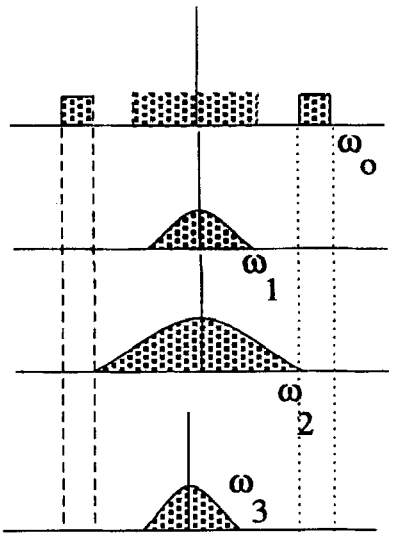
Different amounts of blurring correspond to different cutoff frequencies of the low pass filter. Therefore, for objects having a broad frequency range, the shape-from-focus methods identify the least blurred image by selecting the filtered image with the highest frequency content. However a narrower or more complicated object spectrum may cause difficulties.

Consider three images of an object, image 1, image 2, and image 3, that have been blurred by different amounts. Let image 2 be the least blurred of the three. Let the object patch have frequencies no higher than ω_0. The three low pass filters (corresponding to the different degrees of blur) have cutoff frequencies that satisfy, $\omega_1 < \omega_2$ and $\omega_3 < \omega_2$. Figure 5 shows the three possible scenarios that can occur.

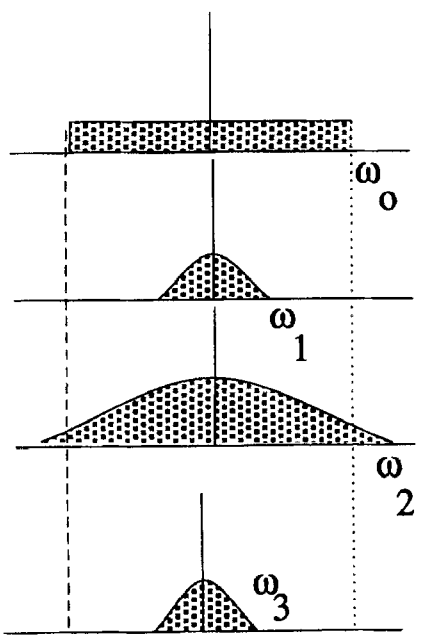
1. Case I: $\omega_0 > \{\omega_1, \omega_2, \omega_3\}$.
   If the frequency content of the object patch is evenly distributed over the entire range (up to ω_0), then by comparing the spectra of the three images, we can easily decide that image 2 has least degradation. This is because image 2 will be the only image with non-zero frequencies in the range $\max[\omega_1, \omega_3]$ to ω_2.
   But if we have an object patch that has only low frequency bands (0 to $\min[\omega_1, \omega_3]$) and high



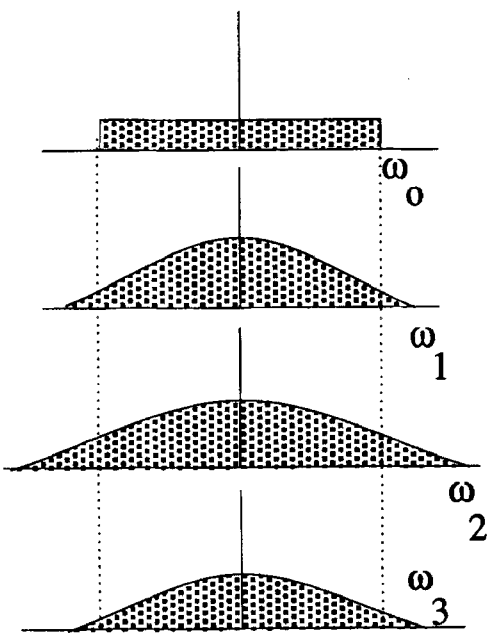
(a) Case Ia: $\omega_0 > \omega_2 > \omega_1, \omega_3$



(b) Case Ib: $\omega_0 > \omega_2 > \omega_1, \omega_3$ but the object patch has no frequencies between Min(ω_1, ω_3) and ω_2



(c) Case II: $\omega_2 > \omega_0 > \omega_1, \omega_3$



(d) Case III: $\omega_1, \omega_2, \omega_3 > \omega_0$

Figure 5. Object spatial frequency and different blurring kernels. Let ω_0 be the extent of the spatial frequencies in the object patch. ω_2, ω_1, ω_3 denote the highest frequencies that are observed for the least blurred image and the two next-to-least blurred images, respectively.

frequency bands (ω_2 to ω_0), then all three images will have similar frequency contents. It is not possible to identify image 2 as having the least degradation.

2. Case II: $\omega_2 > \omega_0 > \{\omega_1, \omega_3\}$.
   If the object patch has frequencies in the range $\max[\omega_1, \omega_3]$ to ω_0 then only image 2 will have those bands present and could thus be identified.
3. Case III: $\{\omega_1, \omega_2, \omega_3\} > \omega_0$.
   For this scenario all three images will have similar spectral cutoffs. Isolating image 2 will therefore not be possible based on the frequency content alone.

- *Size of the Object/Image Patches.* To obtain a more accurate measure of the frequency content, we need to increase the size of the patches used. But as the patch size increases, we cannot be sure that all the points in the patch are at the same range. In fact, some points in the image patch might actually be from different objects, across an occlusion boundary.

4.2. *A Range from Focus Algorithm for NICAM*

Let the sensor plane have $N \times N$ pixels and let the range map be a large array of size $\mathbf{N} \times \mathbf{sN}$, where $\mathbf{s} \geq 1$ is a number that depends on how wide a scene is to be imaged during one sweep of the camera. The kth image frame is represented by I_k and the cumulative, environment centered criterion array with origin at the camera center is represented by R. Every element in the criterion array is a structure that contains the 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] is used to determine the range for that scene point.

Figures 6 and 7 illustrate the geometrical relationships between successive pan angles, pixels of the acquired images, and the criterion array elements as the camera pans from one side of the scene to the other side.

Algorithm. Let $j = 0$. $\phi = 0$. Initialize R and then execute the following steps.

- *Step 1.* Capture the jth image I_j.
- *Step 2.* Pass the image through a focus criterion filter to yield an array C_j of criterion values.
- *Step 3.* For the angle ϕ (which is the angle that the camera has turned from its starting position),

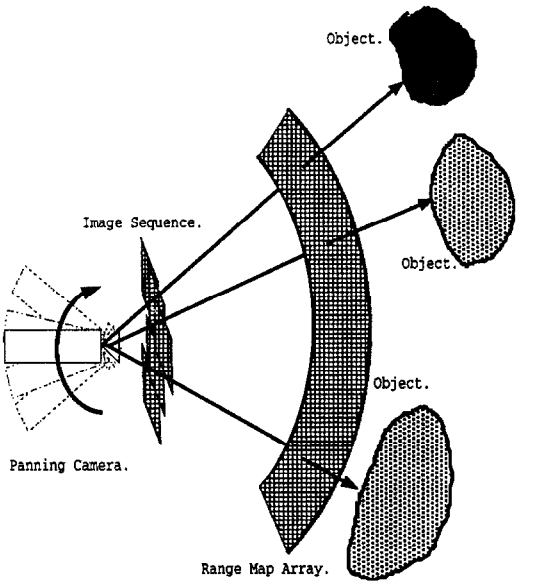


Figure 6. Panning camera, environment centered range array, and the images obtained at successive pan angles. Each range array element is associated with multiple criterion function values which are computed from different overlapping views of the scene. The maximum of the values in any radial direction is selected for the corresponding range array element, to compute the depth value in that direction.

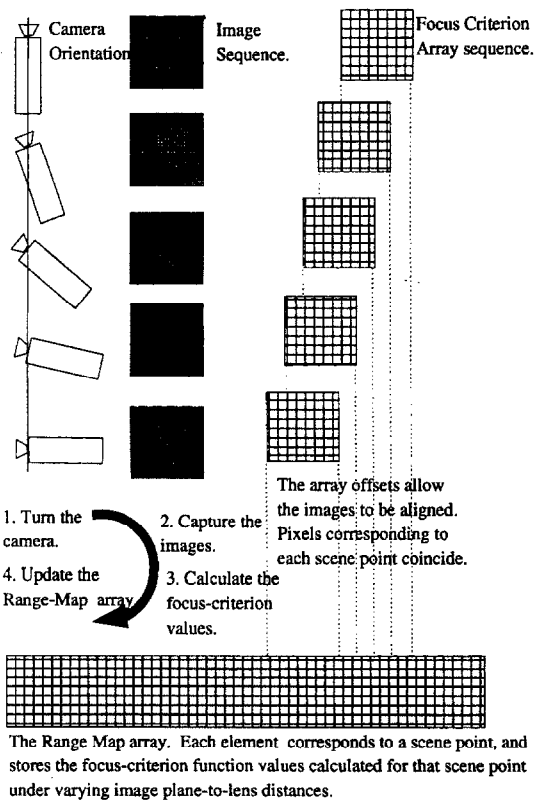


Figure 7. Steps involved in obtaining range from focus.

calculate the offsets into the range map required to align image I_j with the previous images such that overlapping pixels correspond to the same scene point. For example, Pixel $I_j[50][75]$ might correspond to the same object location as pixels $I_{j+1}[50][125]$ and $I_{j+2}[50][175]$.

- *Step 4.* 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) associated with the image having maximum criterion value. Record this value in the range array location corresponding to the scene point's direction from the lens center.
- *Step 5.* Rotate the camera to the next pan angle. Update ϕ and j.
- *Step 6.* Repeat the above steps until the entire scene is imaged.

4.3. *Comparison with Traditional Algorithms*

Traditional range from focus methods using frontal cameras concentrate on scenes which can all be imaged in a single frame. A ranging method was proposed in Wilkox (1993) where a scanning mirror was placed in front of a stationary non-frontal camera. As the scanning mirror cannot be placed at the lens center, there will be problems with correspondence of scene points over different image frames[4].

To be able to compare traditional range from focus algorithms using frontal cameras with the proposed algorithm using a NICAM, we need to first extend traditional algorithms to handle panoramic scenes.

Frontal Camera. Consider the two dimensional SF cone cross-section of a frontal camera which has a sensor plane of length $2l$ units. Let the sensor surface translate from a distance of v_1 from the lens center to v_2, $v_2 \leq v_1$. When the sensor surface is at position v_1, the angle subtended by the SF surface at the lens center is

$$\theta_1 = 2 \arctan\left[\frac{l}{v_1}\right]$$

and when at position v_2, the subtended angle is

$$\theta_2 = 2 \arctan\left[\frac{l}{v_2}\right]$$

If the camera pans about the lens center, as $\theta_2 \geq \theta_1$, to cover the entire angle of 2π around the camera we will need to take image sets at $\lceil \frac{2\pi}{2 \arctan[\frac{l}{v_1}]} \rceil$ pan angles. For each pan angle, the algorithm would have to translate the sensor plane in order to estimate the ranges of the object points within the field of view. Figure 8 shows that the SF cones for adjoining camera pan values will have overlapping regions, i.e., regions where the range is calculated twice.

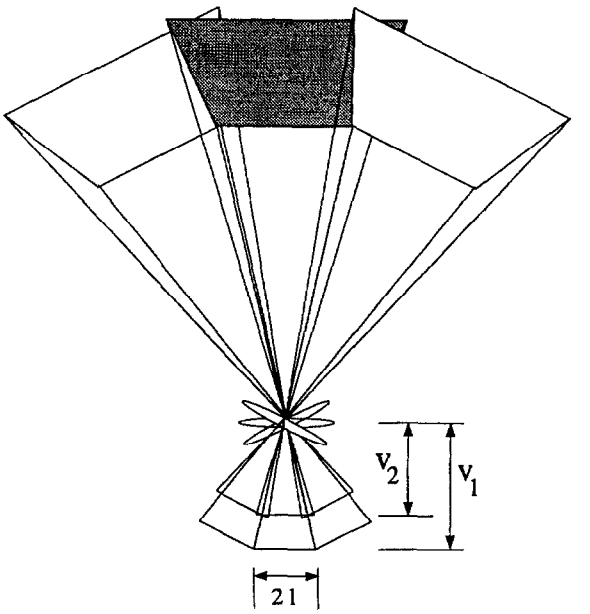


Figure 8. Panning a frontal camera about the lens center. For each pan angle, the sensor plane of length $2l$ is translated between the two focus adjustment limits (v_1 to v_2) to create a SF cone.

Comparison. Figure 9 shows the schematic of the SF surfaces using the traditional frontal camera and the proposed non-frontal camera. The shaded region in the figure represents the space swept by the SF surfaces. Following are some points of difference between the traditional range from focus algorithms used for panoramic scenes and NICAM:

- Both methods need panning the camera, although the traditional method has an extra mechanical motion wherein the sensor plane is moved towards and away from the camera.
- Panning the frontal camera about the lens center causes overlap of adjoining SF cones, and therefore some wasted computation.
- In many frontal cameras the focus adjustment moves the lens assembly and not the sensor plane. This causes the image position of scene points to be dependent on the scene point's range.

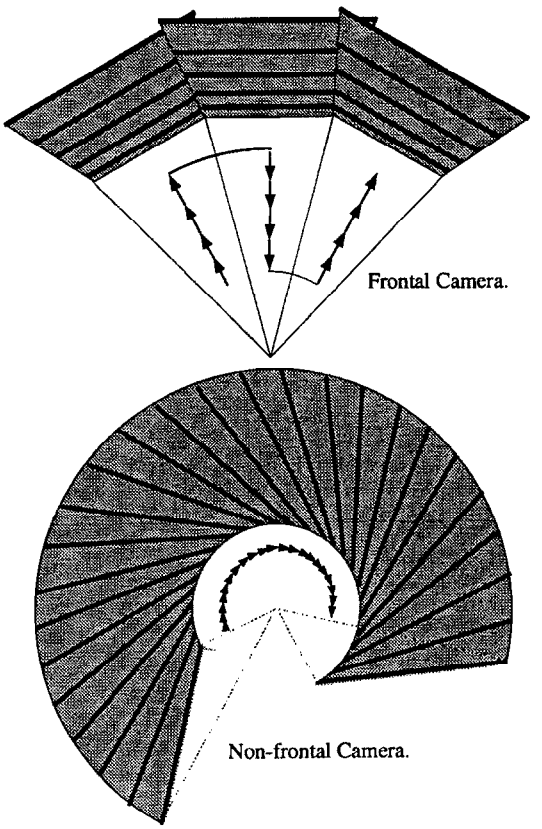


Figure 9. SF cones for the frontal and non-frontal sensor plane cameras. Thick lines indicate the position of the SF surface at different times. Shaded regions indicate the SF cones. Arrows show the sequence of mechanical motions. The frontal camera has two mechanical motions, whereas the non-frontal camera has only a pan motion.

Some common features of traditional algorithms and the proposed algorithm using the NICAM are the following:

- The translational sensor plane motion in traditional algorithms also induces a zoom effect which is the same for each pixel. However, the zoom value changes with changing v value and must be compensated for by the algorithm. A non-frontal camera involves a linearly varying zoom effect across the sensor plane due to the linearly changing focus distance. This leads to a warping effect which makes horizontal parallel lines in the scene appear to converge in the image. Dewarping is therefore necessary. The amount of dewarping is a function of horizontal image location and the sensor plane tilt. For a given tilt, the warping can be compensated for while updating the range may array (in Step 3 of the algorithm). Both methods thus require the calculation of the focus criterion function and compensating for the warp.
- Depth of field increases with increasing range for both frontal and non-frontal cameras. Either moving the sensor plane in uniform steps from v_1 to v_2 in a frontal camera or changing the pan angle in uniform steps in a non-frontal camera, will cause the spacing between the SF surfaces to be small at close range values and increase with increasing range (Krishnan and Ahuja, 1994). This ensures that there is not much overlap between neighbouring SF surfaces due to the depth of field.

5. Camera Realization, Algorithm Implementation and Experimental Results

A CCD camera was modified such that its sensor plane tilt was controllable. The camera was then mounted on a rotation platform, such that the axis of rotation of the platform, passed through the lens center. Figure 10 shows a schematic of the modified camera. Not shown in the figure are four linear stages which can be only manually controlled. Two micrometer stages control the sensor plane tilt motor inside the camera. The third micrometer stage controls the height of the CCD board from the bottom of the camera (to align it with the lens). The fourth linear stage controls the position of the camera in relation to the camera pivot axis. Rough calibration of the camera was done by first adjusting the sensor plane tilt to zero and finding the point of

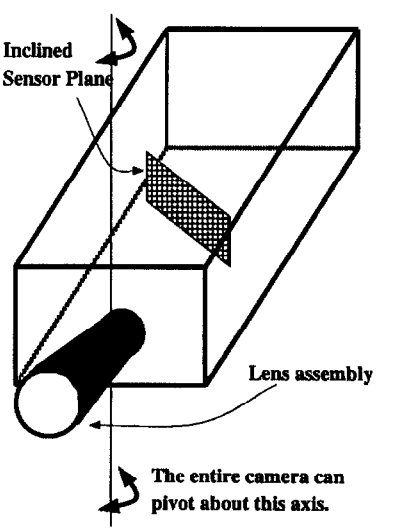


Figure 10. Schematic of the modified camera.

intersection of the optical axis and the sensor plane. This was done by finding the center of expansion of image points, due to changes in the lens's magnification. The intersection point was then brought to the image center by adjusting the micrometer linear stages.

5.1. Experimental Results

The following experiments were done to check the performance of the camera.

Illustration 1: Two similar lenses of focal length 35 mm were mounted on two cameras. One camera was the modified camera (called NICAM henceforth) and the other was a regular camera. The lens apertures were adjusted for ambient lighting. A grid pattern was placed vertically, perpendicular to the optical axis of each lens. The image of the grid pattern observed by each camera was examined separately.

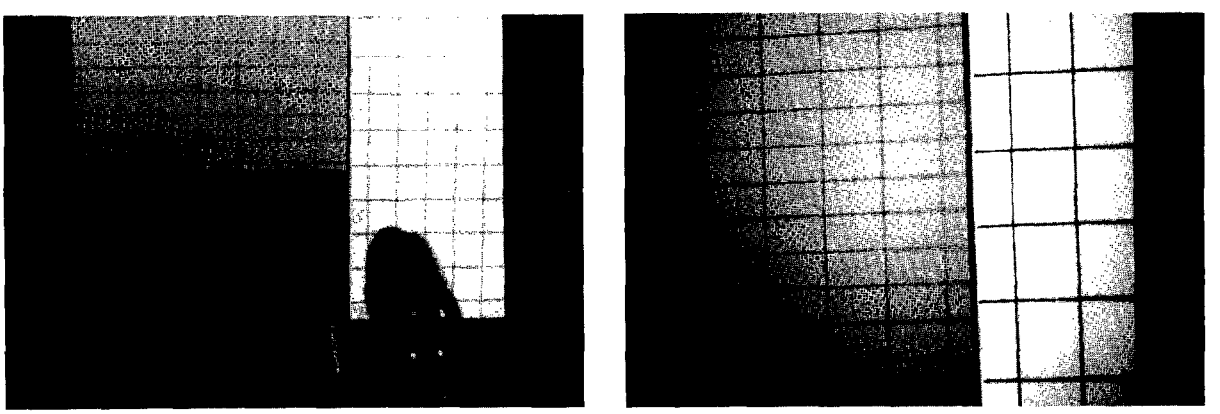
Result: As expected, the NICAM was unfocused in some regions and focused in others. By turning the focus adjustment on the lens, we could focus the grid pattern on different parts of the sensor plane. Along image columns, the grid pattern was identically blurred. Along image rows, the blurring varied across the image such that there was one part that had the least blurring (corresponding to the region under sharpest focus).

The regular camera could not be made to exhibit the behavior of the NICAM. Turning the focus adjustment on the lens would either cause the entire imaged region to be in sharp focus or would cause the entire image to be blurred.

Illustration 2: In this experiment we placed two grid patterns (one square and the other rectangular) parallel to each other and about 45 inches apart in range. The NICAM was fitted with a 35 mm lens and was placed about 42 inches from the near grid pattern such that the camera's optical axis was perpendicular to both the grids. The grid patterns were moved in their planes such that the left side of the image corresponded to the far grid pattern and the right side of the image corresponded to the near grid pattern. The lens's focus adjustment was left at an intermediate setting and was not changed during the experiment. The lens's aperture was also left at a value suitable for the ambient lighting conditions. Figure 11(a) shows the overview of the setup.

Result: The left side of the image exhibited a part that had the least blur. This corresponded to a focused image of part of the far grid. At the same time, the right side of the image too had a part that was under sharp focus. This focused region corresponded to part of the near grid pattern. The modified camera could therefore have two different regions, at different depths, in sharp focus concurrently. Figure 11(b) shows an image seen through the modified camera.

This experiment was repeated by moving the near grid pattern farther away from the camera. The near grid was also moved to the left with each backward shift. The far grid was not moved at any time. The experiment showed that the focused region of the near grid was imaged closer and closer to the focused image of the far grid. Figure 12(a) shows a schematic of the above experiment and result.



(a) (b)

Figure 11. (a) Overview of the scene for Illustration 1. There are two grid patterns, about 45 inches apart in range. (b) Image through the modified camera. There are focused regions on both grids. For a frontal camera we can get focused regions on only one grid.

Conclusions: Let us suppose that the NICAM is panned across the scene, from left to right. Also let the rotation axis be the vertical line passing through the lens's principal point. We have seen from Illustration 2 that the left side of the camera focuses regions that are about 85 inches in range while the right side of the camera focuses regions that are about 40 inches in range. Therefore by panning the camera, we will have all regions, that lie between 40 and 85 inches from the camera, in sharp focus, each region in focus for some camera pan angle.

Experiments 1 and 2: In these experiments, 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 = 54 in) and a planar surface (range = 38 in). The scene in Experiment 2 consists of, from left to right, a planar surface (range = 70 in), a planar surface (range = 50 in), and a face of a box (at a depth of 35 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. So, the scene point that is imaged on pixel $I_1[230][470]$ in the first image, is also imaged on pixel $I_2[230][455]$ in the second image, pixel $I_3[230][440]$ in the third image, etc.

A scene point will thus be present in a maximum of thirty four[5] images. In each image, the effective distance from the lens to the sensor plane location where a fixed scene point is imaged is different. There is a 1-to-1 relationship between the image column number and the distance from lens to sensor element. The column number at which a scene point is imaged with greatest sharpness, is therefore also a measure of the range of the scene point which is given by the lens law.

Each frame was processed in sequence with the focus criterion values calculated for all pixels. The rangemap data structure was updated with the criterion values and then the next frame was processed. After all the images for a particular scene point are obtained (a maximum of 34 frames), the column number where the focus criterion value peaked was found out. This process of determining the peaks is carried out in conjunction with processing of new incoming frames and updating the rangemap data structure.

Results: Among the focus criterion functions that were tried, the Tenengrad function seemed to have the best performance/speed characteristics. The ranging accuracy is the same as that for traditional range from focus methods. In addition to the problems described in Section 4.1, the range map has the following two other problems:

- Consider a scene point A, that is imaged on pixels, $I_1[230][470]$, $I_2[230][455]$, $I_3[230][440]$, etc. Consider also a neighboring scene point B, that is imaged on pixels $I_1[230][471]$, $I_2[230][456]$, $I_3[230][441]$, etc. 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[6]. 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. The shape of the criterion function depends on the number of samples taken and the spacing between the samples. The spacing between the samples is related to the pan angle increments. This problem also occurs in traditional range from focus methods where the sample spacing is related to the movement of the sensor plane (Xiong and Shafer, 1993).

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. Figure 12(b) shows some of the images from the image sequence used for Experiments 1 and 2. Because the sensor plane is fixed relative to the camera, there is a one-to-one correspondence between the column number of a pixel, and the distance of the pixel to the lens (v). The column number in the image where a particular scene point appeared in sharpest focus (where it's focus criterion function was maximum) is thus a measure of the range of the scene point. The column number is therefore the *range disparity* value for the scene point. Figure 13 shows two views of the range disparity values calculated for Experiments 1 and 2. Parts of the scene where we cannot determine the range values primarily due to lack of sufficient detail are shown blank.

6. Summary and Conclusions

In this paper we have shown that panning a camera whose sensor plane is not perpendicular to the optical axis, allows us to estimate range values of object points. We derived the lens law for the proposed camera and showed that using the exact distance between a sensor pixel and the lens (as the image distance v) instead of the perpendicular distance between the sensor plane in the familiar lens equation for the frontal sensor plane case yields the lens law for the modified camera. The SF-surface is thus an inclined plane. When the camera's pan angle direction changes by turning about the lens center, a 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 the shortcoming of range-from-focus algorithms in general and also some that are unique to our method. We have also described the results of some experiments that were conducted to demonstrate the feasibility of our method.

A by-product of the range estimation algorithm is the intensity value of each scene point at it's best focus. We can thus create images of panoramic scenes using a non-frontal imaging camera in which all objects are in sharp focus regardless of their locations (Krishnan and Ahuja, 1993d). A consequence of knowing the range and intensity value for the scene points is the capability to produce stereo pairs of images showing all scene points in sharp focus from data acquired by a single NICAM (Krishnan and Ahuja, 1993c).

Other spatially variant sensors that might be more appropriate for certain tasks could be devised. Two examples of modifications that would yield such sensors are:

- The sensor surface consists of planar strips that are perpendicular to the optical axis, but are situated at different depths from the lens. This amounts to replacing the tilted sensor plane by a staircase approximation of it.
- The sensor surface consists of a central, planar part that is perpendicular to the optical axis, and a

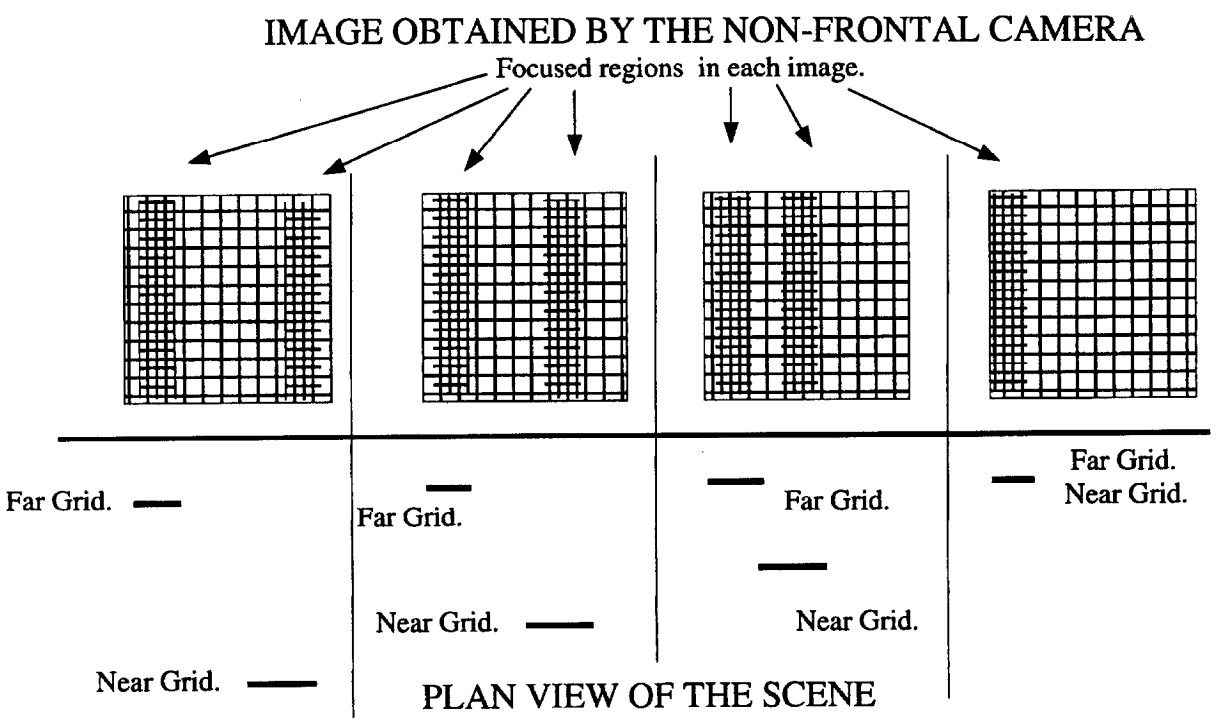


Figure 12a. As the grid nearer to the camera is moved toward the far grid (forward and to the left), the image region that is in sharp focus also moves to the left. The top part of the figure shows schematics of the image obtained using the non-frontal camera. The bottom part of the figure shows the plan view of the location of the near and far grid for each image.

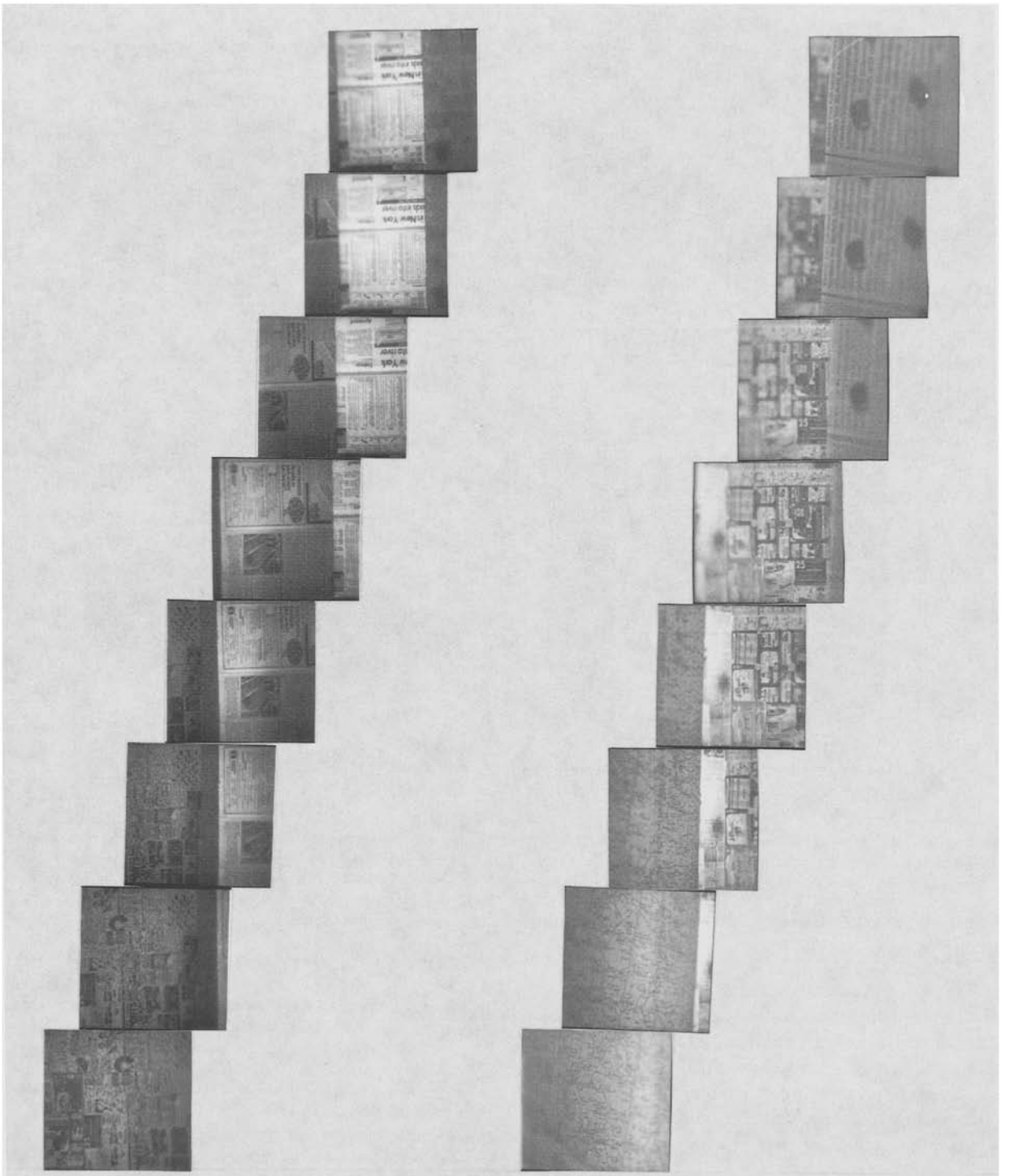


Figure 12b. The images on the left show samples of the image sequence taken for Experiment 1. The images are stacked in the figure such that vertical lines pass through the same scene points in all the images. The images on the right are for Experiment 2.

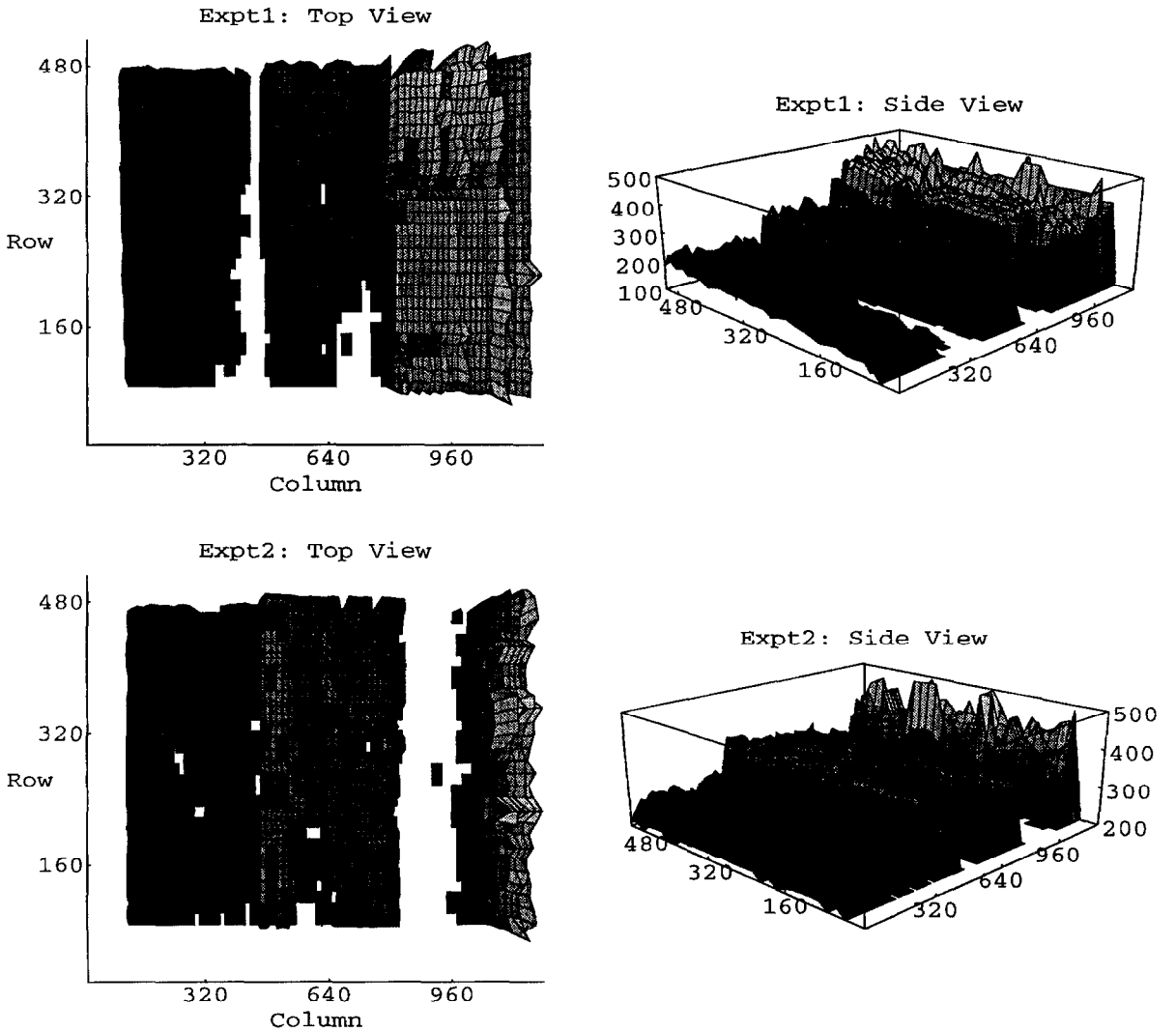


Figure 13. A 512 row × 1200 column portion of the calculated range disparities for Experiments 1 and 2. The further away a surface is from the camera, the darker and smaller is it's height in the range disparity map shown above. Parts of the scene for which range values could not be calculated due to lack of sufficient texture are showm as blank.

peripheral part that slopes away at an angle to the optical axis.

We have considered the use of NICAM for naturally lit scenes. Those parts of the scene having low texture are difficult to analyze. In situations where lighting can be controlled, it is possible to create artificial texture on the scene surfaces by structured illumination. This would allow focusing and therefore range estimation for all visible surfaces with or without texture.

Notes

1. Actually, the depth-of-field effect with 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.
2. Usually referred to in Photogrammetry as Scheimpflug's condition.
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 sensor plane (v). We can then use the lens law to calculate the range.
4. Frame to frame motion of the scene points will depend on the amount of rotation of the mirror, as well as the range of the scene points. This is in essence similar to the stereo correspondence issue.
5. Roughly $\frac{512}{15}$.
6. Neighbors along vertical columns will not have this problem.

References

Abbott, A.L. and Ahuja, N. 1988. Surface reconstruction by dynamic integration of focus, camera vergence, and stereo. In *Proc. Second Intl. Conf. Computer Vision*, Tarpon Springs, Fla, pp. 532–543.

Ahuja, N. and Abbott, A.L. 1993. Active stereo-integrating disparity, vergence, focus, aperture, and calibration for surface estimation. *IEEE Trans. Pattern Anal. Machine. Intell.*, 15(10):1007–1029.

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

Das, S. and Ahuja, N. 1990. Multiresolution image acquisition and surface reconstruction. In *Proc. Third Intl. Conf. Computer Vision*, Osaka, Japan, pp. 485–488.

Das, S. and Ahuja, N. 1992. 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, S. and Ahuja, N. 1993. A comparative study of stereo, vergence, and focus as depth cues for active vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 194–199, New York City, NY.

Ens, J. and Lawrence, P. 1991. A matrix based method for determining depth from focus. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Maui, Hawaii, pp. 600–606.

Goodman, J.W. 1968. *Introduction to Fourier Optics*. McGraw-Hill: Physical and Quantum Electronics Series.

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

Klaus, B. and Horn, P. 1968. Focusing, Technical Report 160, MIT Artificial Intelligence Lab., Cambridge, Mass.

Krishnan, A. and Ahuja, N. 1993a. Range estimation from focus using a non-frontal imaging camera. In *Proceedings of the DARPA Image Understanding Workshop*, Washington, D.C, pp. 959–965.

Krishnan, A. and Ahuja, N. 1993b. Range estimation from focus using a non-frontal imaging camera. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, Washington, D.C. pp. 830–835,

Krishnan, A. and Ahuja, N. 1993c. Stereo display of large scenes from monocular images using a novel non-frontal camera. In *Proceedings of ACCV-93 Asian Conference on Computer Vision*, Osaka, Japan, pp. 133–136.

Krishnan, A. and Ahuja, N. 1993d. Use of a non-frontal camera for extended depth of field in wide scenes. In *Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision XII: Active Vision and 3D Methods*, Boston MA, pp. 62–72.

Krishnan, A. and Ahuja, N. 1994. Depth of field for tilted sensor plane. Technical Report UIUC-BI-AI-RCV-94-08, Beckman Institute, University of Illinois.

Krotkov, E.P. 1987. Focusing. *International Journal of Computer Vision*, 1(3):223–237.

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

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

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

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

Pentland, A., Darrell, T., Turk, M., and Huang, W. 1989. A simple, real-time range camera. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 256–261.

Pentland, A.P. 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.

Tenenbaum, J.M. 1971. Accomodation in computer vision. Ph.D. Thesis, Stanford University, Palo Alto, Calif.

Wilcox, B. 1993. Personal communication on tilted sensor plane camera with scanning mirror.

Xiong, Y. and Shafer, S.A. 1993. Depth from focusing and defocusing. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, New York City, NY, pp. 68–73.