Image Understanding Workshop, 1994

Obtaining Focused Images 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, e-mail: ahuja@vision.csl.uiuc.edu

Abstract

This paper is concerned with obtaining focused images of a scene using a non-frontal camera. The non-frontal camera used here has a sensor surface which is planar, can be dynamically translated and oriented, and can be rotated with respect to the optic center. Up to three object points in a scene can be focused simulaneously using the translation and orientation degrees of freedom of the sensor plane. A scene of arbitrary size and depth can be focused using the rotational degree of freedom to obtain a panoramic image, showing up to 360 degree view of the scene. The focusing ability directly leads to range estimates. The range and focused intensities can be combined to generate 3D stereo displays of the scene.

1 Introduction

This paper describes methods for obtaining focused images of scenes of arbitrary size and depth using a new type of camera introduced in [5] called non-frontal imaging camera. Such cameras have sensor surfaces which can be dynamically configured with respect to the optic center, with multiple degrees of freedom. For example, the non-frontal camera described in [5] has one translational degree and one rotational degree of freedom.

In this paper we are concerned with the use of the non-frontal camera of Figure 1(a) to obtain focused images. Standard cameras only have one focus adjustment which effectively changes the focus distance (v) between the lens and the sensor plane. Further, in standard cameras all points on the sensor plane have the same v value. In contrast, non-frontal cameras have v values that vary across the sensor surface. This property of non-frontal cameras is used in Section 2 to focus up to three arbitrary points simultaneously. If the scene points of interest number more than three, or if the points are spread out in a wide scene, the rotation of the sensor plane about the optic center, or camera panning, is exploited which is described in Section 3. The focusing ability of the non-frontal camera in this case derives from the fact that when the non-frontal camera 1(a) is panned about the lens center, a fixed scene point is imaged in successive frames (pan angles) with different values of v. There will be one such value of v, and equivalently, one such image frame, for which the scene point is imaged at the sharpest focus. This value of v is achieved through (manual or motorized) mechanical adjustments in standard cameras, for one object at a time. Analysis of the image sequence generated by a non-frontal camera panning about its lens center can therefore yield a composite focused image for objects at all depths in an arbitrarily wide scene. A panoramic focused image of up to 360 degree view of a scene can be obtained by selecting the corresponding bounds for pan angles. Section 4 discusses combining the focused intensity estimates with the range estimates derived from the focusing process to synthesize a stereo display of the scene from the monocular image sequence acquired by a single panning non-frontal camera.

2 Focused Single Frame Images.

This section describes methods for obtaining a focused image of objects visible in a single view using a non-frontal camera. The sensor surface's capability for dynamic reorientation is used for this purpose. A standard camera can only simultaneously focus on scene points that are all at the same range from the camera or that all lie within the depth of field. With the non-frontal camera, up to three points (objects), not all within the depth of field of the standard camera, can be imaged in sharp focus.

*The support of the Advanced Research Projects Agency under grant N00014-93-1-1167 is gratefully acknowledged.

One Point: In this simple case, the fixed distance to the object point determines the sensor plane distance v. For standard as well as non-frontal cameras, the desired value of v could be obtained by translating the sensor plane. A non-frontal camera also has the option of using a sensor surface tilt control to achieve the desired v value.

Two Points: Standard cameras cannot image two or more scene points that are not all at the same range (within the depth of field). Two scene points specify two points behind the lens that the sensor plane should pass through. With position control and tilt control, the sensor plane can be positioned to pass through any two arbitrary points (within the motion limits of the stages), and thus to image the two points in focus.

Three Points: Tilt control about two orthogonal axes along with position control allows the sensor plane to achieve arbitrary position and orientation inside the camera. Three scene points define three fixed points behind the lens that the sensor plane must pass through. Therefore by maximizing the sharpness at each point, the camera can sharply focus three arbitrary points simultaneously.

Four or more Points: Imaging more than three scene points in sharp focus simultaneously would require warping the sensor plane and is difficult at this time. Alternatively, one could maximize the sharpness of all points combined. This may be achieved by controlling the position and orientation of the sensor plane such that a chosen function of the image blurs at the image points is minimized. In conjunction with a model of the image blur as a function of the displacement of the sensor plane with respect to the true image location, the sensor plane can be configured to minimize the mean squared distance from the true image locations of the points.

A solution that does not require warping the sensor plane and that also does not compromise on the focused image quality is to reconstruct the focused image from a series of images that have individual scene points in sharp focus. A series of images, each taken with different sensor plane positions (different values of v), can be analyzed to determine sharply focused regions, and then the focused regions from each image can be combined to obtain a composite focused image of the scene [11, 4].

If a standard camera with a frontal, translating sensor plane is used, objects in different scene directions will have to be focused by (mechanically) translating the sensor plane. Using a non-frontal camera in which the sensor plane is allowed to rotate about the optic center, the panning motion suffices to focus each scene point as described below.

3 Obtaining Focused Images by Panning a Non-frontal Camera

Consider a non-frontal camera with a sensor plane that is not perpendicular to the optical axis. This causes different sensor pixels to be at different distances from the lens. The non-frontal camera will image with sharp focus a volume of object space that is also tilted with respect to the optical axis. So in any one image frame, there will be scene points in sharp focus that are at different depths from the camera. If the entire camera is now rotated in small steps about a vertical axis through the lens center, then the entire scene will be sequentially imaged by a plurality of frames. By choosing the angle of rotation between successive frames to be sufficiently small, each scene point will be imaged in multiple frames, but by different sensor elements (at different locations on the sensor plane) in different frames. Since the sensor elements are located at different distances from the lens center, the scene point will therefore be imaged with different levels of blur, the sharpest focus image occurring for some rotation angle, and progressively blurred images as the camera rotates further away. To determine when a scene point has been imaged with sharp focus, standard focus criterion functions that estimate the high frequency content can be used [3, 8, 12, 9].

Along with the focused image, the v value corresponding to the focused image of a scene point also yields its range from the camera, using the lens law as done in the range from focus methods [10, 1, 2, 6].

Before we give the algorithm, two factors must be discussed that distinguish the images obtained from those obtained using the standard camera. First, the variation in the registered brightness of the same scene point from different directions (pan angles) must be taken into account (due to varying distance between sensor plane points and lens center). Second, the varying degrees of perspective warping at different locations on the non-frontal sensor plane must be considered.

Irradiance dependency on tilt angle: The pixel brightness (E) in standard frontal cameras depends on the area of the lens aperture (A_{lens}), the scene radiance (L), the mean distance between the lens and the sensor plane (λ) and the pixel location (x, y) as shown below

$$E = A_{lens} \times L \frac{\lambda^2}{[x^2 + y^2 + \lambda^2]^2}$$

It can be shown that for the non-frontal camera with sensor plane tilt of α, the expression for pixel brightness becomes:

$$E = A_{lens} \times L \frac{\cos\alpha \left[\lambda^2 - \lambda x \sin\alpha\right]}{\left[x^2 + y^2 + \lambda^2 - 2\lambda x \sin\alpha\right]^2} \quad (1)$$

Image Warp: The non-frontal camera with a tilted sensor plane gives a oblique perspective view of the scene. It can be expressed in the form,

$$\begin{aligned} x &= \frac{-\lambda X}{[Z - X]\cos\alpha - X\sin\alpha} \\ y &= \frac{-\lambda Y \cos\alpha}{[Z - X]\cos\alpha - X\sin\alpha} \end{aligned}$$

where x and y are the sensor plane coordinates of the image of a point located at (X, Y, Z). The sensor plane is tilted at an angle α with respect to the optical axis. It can be shown that if the sensor plane were not tilted, then a scene point that imaged at (x, y) on the tilted sensor plane will now be imaged at (x', y') given by

$$x' = \frac{x\lambda\cos\alpha}{\lambda - x\cos\alpha} \quad (2)$$

$$y' = \frac{y\lambda}{\lambda - x\cos\alpha}. \quad (3)$$

The above equations can be used to dewarp the image obtained from a non-frontal camera.

3.1 Algorithm

This section presents an algorithm to obtain focused images as well as range estimates of a scene. Let the sensor plane have N x N pixels and let the focus map and range map be large arrays of size N x sN, where s >= 1 is a number that depends on how wide a scene is to be imaged during one sweep of the camera. The k^{th} image frame is represented by I_k and the cumulative, environment centered criterion array with origin at the camera center is represented by **C**. 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 is used to determine the range value for the scene point as well as the intensity for that scene point in the panoramic composite image.

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

- Capture the j^{th} image I_j.
- Compensate for the intensity variation due to sensor plane tilt and position on the sensor plane using Equation 1.
- Dewarp the intensity rectified image using Equations 2 and 3.
- Pass the dewarped 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 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 $I_{j+2}[50][175]$.
- 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. Also fill in the composite focus map with the scene point's intensity in the dewarped image that corresponds to the focus criterion maximum.
- Rotate the camera to the next pan angle. Update ϕ and j.
- Repeat the above steps until the entire scene is imaged.

4 Stereo Display

The algorithm outlined in Section 3.1 yields two types of information about the points in the scene:

- The pixel intensity of the object point, as it would appear in a sharply focused image.
- The 3D location of the object point.

These two separate attributes can be combined in a straightforward manner to create a stereo display[7]. Each of the stereo images generated registers scene points at the image locations determined by the 3D location (estimated during panning) and intensity values taken from the sharpest focused image (estimated during panning). Figure 1(b) shows the data flow diagram. Thus, a stereo viewable 3D display of the scene can be produced from the image sequence acquired by a single panning non-frontal camera.

References

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

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

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

[4] K. Kaneda, S. Ishida, A. Ishida, and E. Nakamae. Image processing and synthesis for extended depth of field of optical microscopes. *Visual Computer*, 8(5-6):351–360, June 1992.

[5] Arun Krishnan and Narendra Ahuja. Range estimation from focus using a non-frontal imaging camera. In *Proceedings of the DARPA Image Understanding Workshop*, pages 959–965, Washington D.C., April 1993.

[6] Arun Krishnan and Narendra Ahuja. Range estimation from focus using a non-frontal imaging camera. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 830–835, Washington D.C., July 11-15 1993.

[7] Arun Krishnan and Narendra Ahuja. Stereo display of large scenes from monocular images using a novel non-frontal camera. In *Proceedings of ACCV-93 Asian Conference on Computer Vision*, pages 133–136, Osaka Japan, November 23-25 1993.

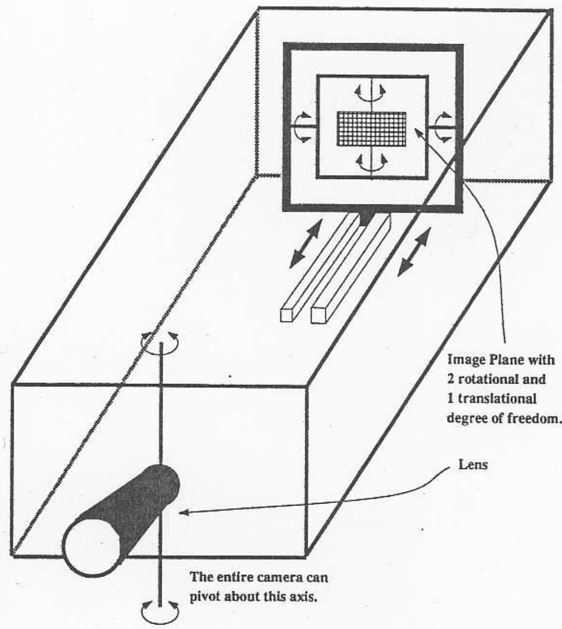
[8] E. P. Krotkov. Focusing. Technical Report MS-CIS-86-22, GRASP Laboratory, University of Pennsylvania, April 1986.

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

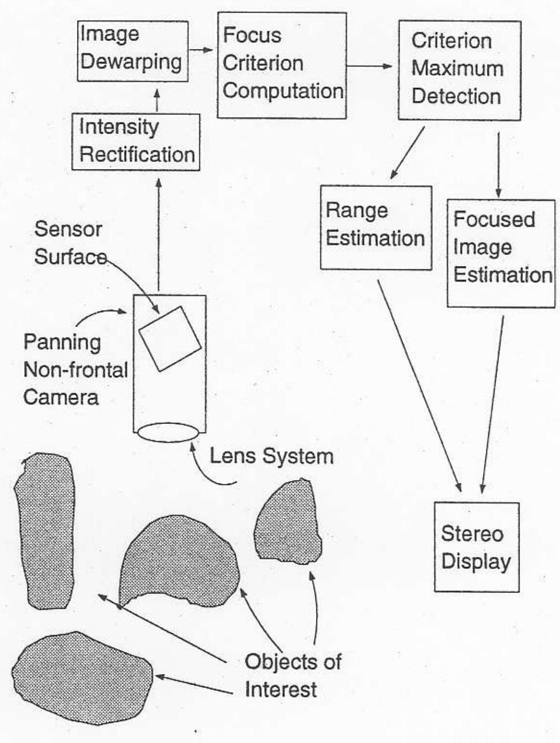
[10] Alex Paul Pentland. A new sense for depth of field. *IEEE Trans. Pattern Anal. and Machine Intell.*, PAMI-9:523–531, 1987.

[11] R. J. Pieper and A. Kopel. Image processing for extended depth of field. *Applied Optics*, 22(10):1449–1453, May 1983.

[12] Jay Martin Tenebaum. *Accomodation in Computer Vision*. PhD thesis, Stanford University, Palo Alto, Calif., 1971.



(a) Non-frontal Imaging Camera.



(b) Data flow Diagram

Figure 1: (a) Non-frontal camera with one rotational degree of freedom about the optical center and one translational and two rotational degrees of freedom for the sensor plane. (b) Dataflow diagram to compute the stereo display.