Appears in Asian Conference on Computer Vision (ACCV'93).

Stereo Display of Large Scenes from Monocular Images Using a Novel Non-frontal 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 a stereo display of a panoramic scene. It uses a novel non-frontal 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 while panning the camera to view panoramic scenes. Camera movement thus integrates panning, focusing and range estimation. A focused image of the large scene is then combined with the range information to obtain a stereo display. Some results in obtaining a stereo display of a panoramic scene are also shown.

1 Introduction

This paper describes a method for obtaining a stereo display from a monocular image sequence of a scene obtained using a novel non-frontal camera. The stereo display is generated from range and reflectance (image intensity) estimates of points in large scenes obtained by the camera. Large scenes are defined as scenes that are deep and wide (panoramic). To image such scenes, a standard camera will have to be panned in order to image all objects/surfaces of interest and focused for each pan angle in order to obtain sharp images. This paper first describes the motivation behind the non-frontal camera, and then presents an algorithm to obtain a focused image of a large scene. Section 4 outlines a method to estimate the range of scene points and Section 5 discusses combining range and intensity estimates to synthesize a stereo display.

2 Motivation Behind the Non-frontal Camera.

The following observations underlie the proposed non-frontal camera. 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 and the location of the object), the imaging parameter v will assume different values. So for fixed scene points, the v value for that scene point will vary as a function of the pan angle. This means that by controlling just 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 [5].

Since in common cameras and lens assemblies, change of focus (v) is usually done by either a manual control or a mechanical control of the focus ring in the lens system, integration of panning with focusing leads to increased operational speed. Further, when the image sensor plane is displaced within the lens system of a standard camera, some elements may undergo translational motion or a screw type rotational motion with the final effect being that the distance of the sensor plane (fixed to the camera) from the lens center may change. Thus focusing action in standard cameras can sometimes also have the side effect of changing the distance of the viewed object from the lens center which is avoided by the above integration.

3 Obtaining Focused Images.

Obtaining focused images of even one frame is difficult when the visible object points lie at different depths from the camera. For a given position of the sensor plane from the lens, only parts of the scene that lie within the depth of field will appear in sharp focus. One way to compose a focused image of a scene with objects at different depths is to first obtain a series of images, taken with

*The support of the National Science Foundation and Defence Advanced Research Projects Agency under grant IRI-89-02728 is gratefully acknowledged.

different sensor plane positions (different values of v). Each image can be analyzed to determine sharply focused regions, and then the focused regions from each image can be put together to obtain a final focused image of the scene [11, 4]. For scenes that are wider than the field-of-view of a camera, image sequences will have to be taken with varying values of v for every pan angle until all parts of the scene are covered. A non-frontal camera simplifies this process considerably [6].

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 corresponding to each frame 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 camera rotation, 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, 12, 7, 13, 9].

Algorithm Let the image plane have N x N pixels and let the focus 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 desired, cumulative, focused image is represented by R. Every element in the focus 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 can be used to determine the focused image for that scene point. Let the camera start from one side of the scene and pan to the other side. Figure 1(a) illustrates the geometrical relationships between successive pan angles, pixels of the images obtained, and the focus array elements.

4 Range Estimation

Range from focus algorithms use the lens law to determine the depth when the sharp image location and focal length are known [10, 1, 2, 8]. For the non-frontal camera, the sensor element to lens distance varies across the sensor elements.

The same image sequence as used to compose the focused panoramic image (Sec 3) can be used to determine the depth map [5]. As the sensor plane tilt remains constant throughout the camera pan, the distance from the lens to the different pixel elements also remains a constant. The sensor element to lens distance associated with the sharpest image of a point yields the depth of the scene point using the lens law. A simple calibration procedure can thus create a look-up table relating row and column numbers of the sensor elements yielding sharpest focus to range values.

5 Stereo Display

The algorithms outlined in Sections 2 and 3 yield 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 by creating a stereo display. Each of the stereo images generated contains scene points at the image locations determined by their 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.

5.1 Implementation and Results

The stereo display was created and observed using the library routines on a Silicon-Graphics machine. The inputs are the arrays that contain range information (and thus 3D locations) and intensity information for the scene points at the corresponding image locations. Stereo disparity values are computed for a given pair of viewpoints.

Figure 2(a) shows the range disparity values for one image sequence and Figure 2(b) shows the computed focused image. The scene 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).

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 Eleventh National Conference on Artificial Intelligence*, pages 830–835, Washington D.C., July 11-15 1993.

[6] Arun Krishnan and Narendra Ahuja. Use of a non-frontal camera for extended depth of field in wide scenes. In *Proceedings of SPIE Intelligent Robots and Computer Vision XII: Algorithms and techniques*, Boston MA, September 1993.

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

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

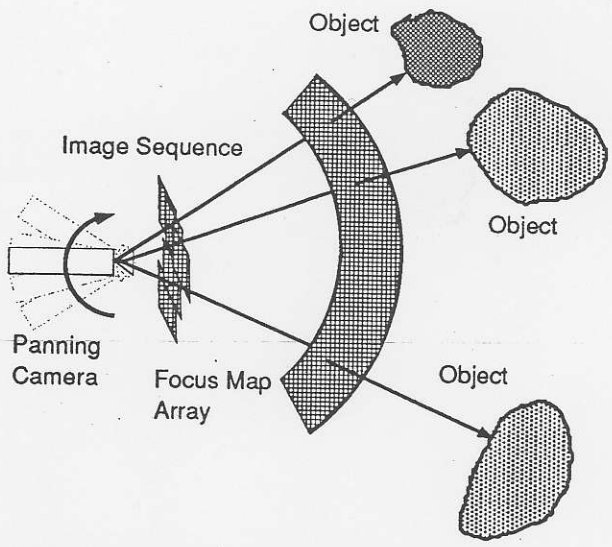
[9] G. Ligthart and F. C. A. Groen. A comparison of different autofocus algorithms. In *Proc. Sixth Intl. Conf. Pattern Recognition*, pages 597–600, Oct. 1982.

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

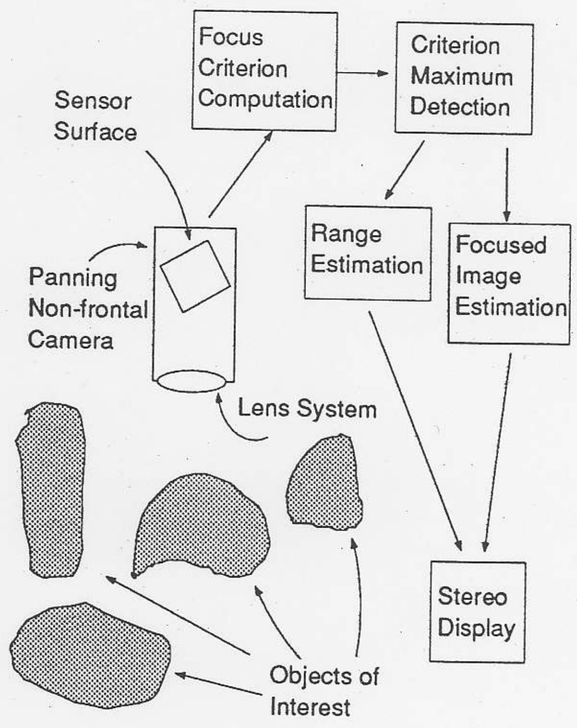
[11] R. J. Pieper and A. Kopel. Image processing for extended depth of field. *Applied Optics*, 22(10):1449–1453, May 1983.

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

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

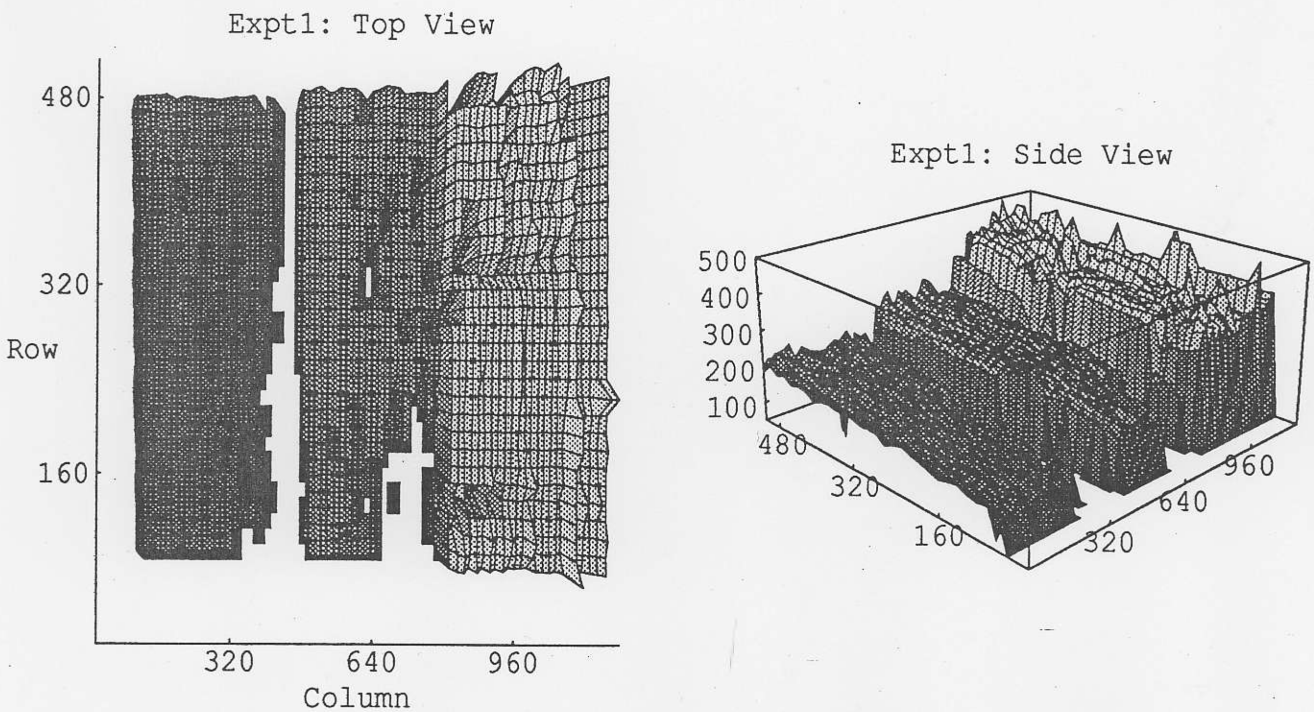


(a) Panning camera, focus array, and the images obtained at successive pan angles.

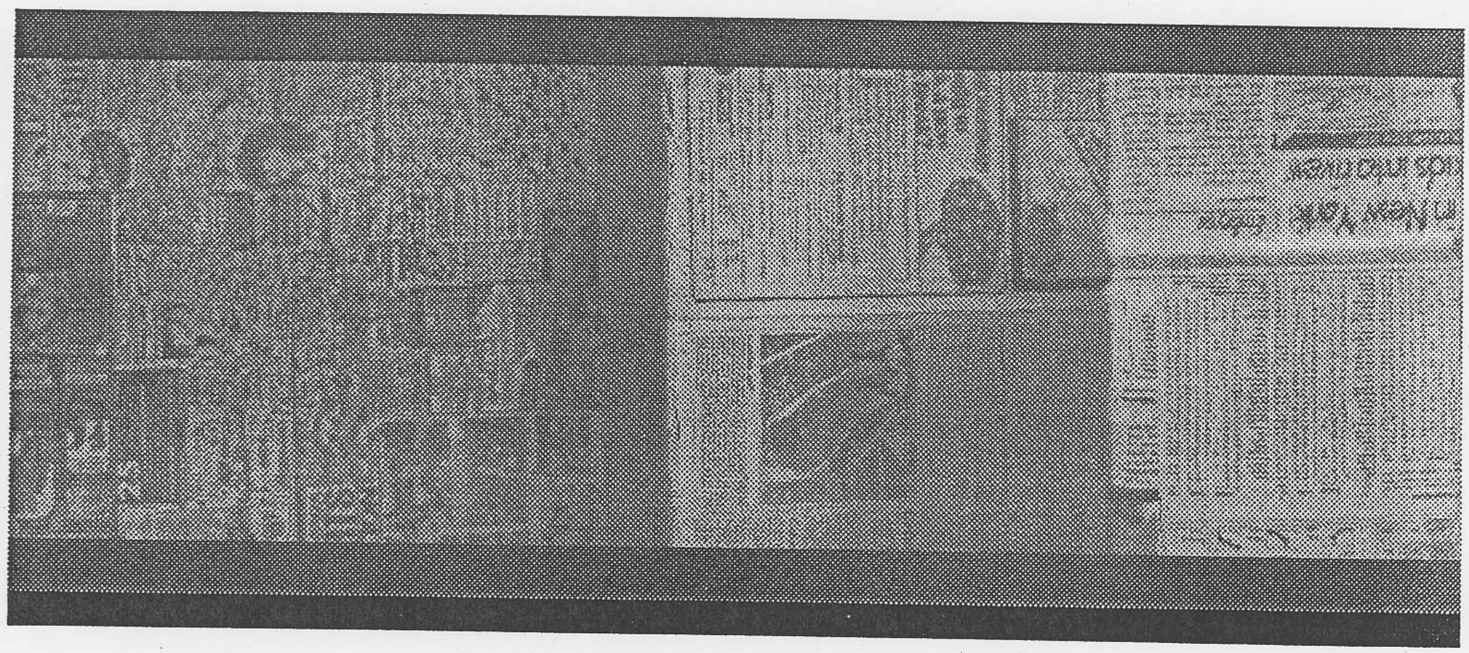


(b) Data flow Diagram

Figure 1: (a) Each focus 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 focus array element, to compute the image intensity value in that direction. (b) Dataflow diagram to compute the stereo display.



(a) Range Disparity values



(b) Focused Image

Figure 2: (a) Range disparities and (b) Focused image. The further away a surface is from the camera, the smaller is its 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 shown blank. The corresponding focused image is also shown.