A New Method for Shape from Focus

Murali Subbarao   Tae Choi

Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, New York 11794-2350

Abstract

A new shape-from-focus method is described which is based on a new concept named Focused Image Surface *(FIS). FIS of an object is defined as the surface formed by the set of points at which the object points are focused by a camera lens. According to paraxial-geometric optics, there is a one-to-one correspondence between the shape of an object and the shape of its FIS. Therefore, the problem of shape recovery can be posed as the problem of determining the shape of the FIS. From the shape of FIS the shape of the object is easily obtained. In this paper the shape of the FIS is determined by searching for a shape which maximizes a focus measure. In contrast with previous literature where the focus measure is computed over the planar image detector of the camera, here the focus measure is computed over the FIS. This results in more accurate shape recovery than the traditional methods. Also, using FIS, a more accurate focused image can be reconstructed from a sequence of images than is possible with traditional methods. The new method has been implemented on an actual camera system, and the results of shape recovery and focused image reconstruction are presented.*

1 Introduction

The image of a scene formed by an optical system such as a lens contains both *photometric* and *geometric* information about the scene. Brightness or radiance and color of objects in the scene are part of photometric information whereas distance and shape of objects are part of geometric information. Recovering this information from a set of images sensed by a camera is an important problem in computer vision. Shape-From-Focus (SFF) methods provide one solution to the problem.

For an aberration-free convex lens, (i) the radiance at a point in the scene is proportional to the irradiance at its *focused image* [6], and (ii) the position of the point in the scene and the position of its focused image are related by the *lens formula*

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{1}$$

Figure 1: Image formed by a convex lens.

where f is the focal length, u is the distance of the object from the lens plane, and v is the distance of the focused image from the lens plane (see Figure 1). Given the irradiance and the position of the focused image of a point, its radiance and position in the scene are uniquely determined. In fact the positions of a point-object and its image are *interchangeable,* i.e. the image of the image is the object itself. Now, if we think of an object surface in front of the lens to be comprised of a set of points, then the focused images of these points define another surface behind the lens (see Figure 1). We define this surface to be the *Focused Image Surface* (FIS) and the the image irradiance on this surface to be the *focused image.* There is a *one to one correspondence* between FIS and the object surface. The geometry (i.e. the shape) and the radiance distribution of the object surface is uniquely determined by the FIS and the focused image.

In this paper we are concerned with the principles and computational methods for recovering the geometry and the radiance of an object from its *sensed image.* Note that a sensed image is in general quite different from the focused image of an object. In computer vision, the sensors are usually planar image detectors such as CCD arrays. Therefore, for curved objects, only some parts of the image will be focused whereas other parts will be blurred. A sensed image will be the focused image only when the shape of the sensor and the shape of FIS match.

In traditional SFF methods (e.g. [5, **?**, **?**, **?**, 10, 18, **?**]) a sequence of images are obtained by continuously varying one or both of the following camera parameters: (i) distance between the lens and image detector, and (ii) the focal length. For each image in the sequence, a sharpness measure or focus measure is computed at each pixel using a small (about 20×20) image neighborhood around the pixel. At each pixel, that image frame among the image sequence which gives a maximum sharpness measure is determined. The grey level (which is proportional to image irradiance) of the pixel in the image frame thus determined is taken to be the grey level of the focused image for that pixel. The camera parameter values for this image frame are used to compute the distance of the object point corresponding to the pixel. A simple measure of sharpness of an image $g(x, y)$ is its grey level variance. Measures based on the energy of derivatives of images are however better suited ([18]).

The traditional SFF methods do not yield accurate shape or depth-map of objects. The main reason for this is that a focus measure is defined and computed over image frames sensed by planar image detectors. The focus measure at each pixel in an image frame is computed using a small window around the pixel. This corresponds to a piecewise constant approximation of the object shape in the window. Because of this approximation, the focused image reconstructed from the image sequence will be an approximation to the actual focused image.

The fundamental contribution of this paper is the idea that focus measures should be computed over the FIS using pixels lying on the FIS in the image sequence rather than over image frames where the pixels lie on a plane. Maximization of focus measures computed over FIS avoids the piecewise constant approximation of object shape found in the traditional SFF methods. The computational implementation of this idea involves two steps. The first step is essentially to estimate an approximate FIS using one of the traditional SFF methods. The second step is to refine this approximate estimate by searching for an FIS shape which maximizes a focus measure computed over pixels lying on the FIS. The search is local and therefore computationally efficient. At present, our implementation corresponds to a piecewise planar (or linear) approximation of object shape as opposed to piecewise constant approximation. However, our implementation algorithm can be easily extended to higher order approximation at the cost of additional computation.

Our SFF algorithm has been implemented on a prototype camera system named Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). A brief description of SPARCS architecture is included. A number of experiments were carried out using SPARCS to evaluate our SFF algorithm. The experiments and their results are described. The experimental results show that our algorithm performs well.

In this paper we are mainly concerned with SFF methods which give dense and accurate depth-maps, and which do not require a a detailed knowledge of the camera characteristics. These methods require a sequence of image frames (about 10 to 30) recorded with different camera parameter settings. However, there are methods [11, 19, 20, 2] which do not require a sequence of images, but only a few (about 2 or 3) acquired with different camera parameter values. These methods are very fast (about 10 times), but less accurate (their best performance gives an RMS error which is twice that of the SFF methods). These fast methods are known as Depth-from-Defocus (DFD) methods whereas the SFF methods considered here are known as Depth-from-Focus (DFF) methods. Clearly, DFD methods can be used first to obtain a rough estimate of shape and then DFF or SFF methods can be used to refine the rough estimate to obtain a more accurate estimate of shape.

We first consider the case of recording the image sequence by moving the image detector (or lens) along the optical axis of the lens. The results of this case can be easily extended to that of obtaining the image sequence by adjusting the focal length of the lens. When the image detector of a camera is moved from one end to the other, typically the focus measure in an image window gradually increases, reaches a maximum at the FIS, and then decreases gradually thereafter. The problem then is to find the image detector position at which the focus measure is a maximum. This is essentially a search of the image detector position space.

2 Relation between object surface and FIS

Figure 2 shows a right handed Object Space Coordinate System (OSCS) (X, Y, Z) with its origin O at the optical center of a convex lens L. The visible surface of any object in the object space (scene) can be expressed as $Z = Z(X, Y)$. We assume that

$$Z(X, Y) \geq f \tag{2}$$

because any object point closer than focal length f will produce only a virtual image but no real image.

The image formed by a convex lens is inverted with respect to the object. Therefore, for convenience, we define a left handed Image Space Coordinate System (ISCS) (x, y, z) (see Fig. 2) with its origin at O. The axes of the ISCS point in the direction exactly opposite to that of the corresponding axes of the OSCS. For any point (X, Y, Z) in OSCS, let its focused image in the ISCS be (x, y, z). The points (X, Y, Z) and (x, y, z) form a conjugate pair of points. Using the properties of similar triangles, it is easy to show that

$$\frac{x}{z} = \frac{X}{Z} \quad \textit{and} \quad \frac{y}{z} = \frac{Y}{Z} \tag{3}$$

For an object surface $Z(X, Y)$, the corresponding Focused Image Surface or FIS can be denoted by $z(x, y)$. It follows from the lens formula 1 that

$$\frac{1}{f} = \frac{1}{Z(X, Y)} + \frac{1}{z(x, y)} \tag{4}$$

The above relation can be used to obtain the shape of an object from its FIS. It is helpful to recall here a few well-known results. A point at $Z = \infty$ comes to focus at $z = f$, a point at $Z = 2f$ comes to focus at $z = 2f$, and a point at $Z = f$ comes to focus at $z = \infty$.

Although the above relation between the shape of an object $Z(X, Y)$ and its FIS $z(x, y)$ is quite simple, the following theorem is particularly interesting.

Theorem 1 *The Focused Image Surface of a planar object is also planar.*

Proof Let

$$Z = PX + QY + R \tag{5}$$

be a planar object where P and Q are respectively the slopes of the surface along X and Y axes respectively and R is the intercept along the Z axis. We have

$$Z - PX - QY = R \tag{6}$$

$$\rightarrow 1 - P\frac{X}{Z} - Q\frac{Y}{Z} = \frac{R}{Z} \tag{7}$$

(8)

Using relations **??**, the above equation can be written as

$$1 - P\frac{x}{z} - Q\frac{y}{z} = R\left(\frac{1}{f} - \frac{1}{z}\right) \tag{9}$$

Rearranging terms in the above relation, we obtain

$$z = px + qy + r \tag{10}$$

where

$$c = \frac{f}{R - f}, \quad p = -cP \quad q = -cQ \quad and \quad r = cR \tag{11}$$

Therefore we have shown that the FIS of a planar object $Z(X, Y) = PX + QY + R$ is also planar and is given by $z(x, y) = px + qy + r$. The relation between the object surface parameters P, Q, R and FIS parameters is given by 11.

The fact that the FIS of a planar object is also planar can be used to conclude the following: (i) the FIS of a polyhedral object is also polyhedral (with the same number of planar faces), (ii) the FIS of a straight line is also a straight line, and (iii) the FIS of a surface which can be generated by sweeping a straight line in 3D object space is also a surface which can be generated by sweeping a straight line in 3D image space. The first result follows directly from the theorem. The second result can be proved by noting that the intersection of two planes in object space is a straight line and the FIS of the straight line is the intersection of the FISs of the two intersecting planes which are themselves planar. The third result follows from the second result. A consequence of the third result is that the FIS of a cone is a "distorted cone", and the FIS of a cylinder is a "distorted cylinder".

An important significance of the above theorem in SFF methods is that if the shape of an object can be approximated well by a piecewise planar surface then the shape of the corresponding FIS can also be approximated well by a piecewise planar surface.

3 Focus Measure

There are many focus measures which perform well when used in SFF methods [?, 10, ?]. Any one of these could be used in our algorithm described next. Two simple examples of reasonably good focus measures are grey level variance and energy of Laplacian of the image. In our implementation we have chosen a focus measure named M_2' which was proposed in [?]. One way to compute M_2' of an image is to first low-pass filter the image by convolving with a small Gaussian mask, and then computing the sum of the square of gradient magnitudes at each pixel. We have chosen M_2' because it exhibits many desirable properties such as monotonicity, unimodality, and large gradient in the presence of noise, aliasing, and side-lobe effects of camera modulation transfer function.

Before the focus measure of an image is calculated, the image is first normalized with respect to brightness. This is done by dividing the grey level of each pixel by the mean grey level of the whole image.

4 Algorithm

Conceptually, our Shape-from-Focus (SFF) algorithm can be described as follows. The image detector is moved to $z = f$. A sequence of images $g(i,j,k)$ are recorded by moving the image detector to positions $z_i = z_0 + (i-1)\delta$ where δ is a small displacement, for $i = 1, 2, \cdots, I$, $j = 1, 2, \cdots, J$, and $k = 1, 2, \cdots, K$. Usually, $z_0 = f$. J and K are the number of rows and columns respectively in each image frame and I is the number of image frames (see Figure 3). We can think of this image sequence as an image volume. In this image volume, our problem is to find the set of pixels which lie on the focused image surface (FIS) of the object. For surfaces with moderate or low slopes, for any given row j and column k, there is only one pixel which lies on the FIS. The image frame number i to which this pixel belongs depends on (j,k) and therefore it can be expressed as a function $i(j,k)$. The grey level of this pixel is $g(i(j,k),j,k)$. The relation between the row number j and the y coordinate is $y = (j - j_c)d$ where $j_c = J/2$ is the row index of the center row and d is the distance between two rows of pixels on the image detector array. Similarly, the relation between the x coordinate and the column index k is $x = (k - k_c)d$ where $k_c = K/2$ is the column index of the center column and d is the distance between two columns of pixels on the image detector array.

The shape of the FIS can be determined from the function $i(j,k)$ which gives the frame number of the pixel lying on the FIS for any given (j,k). The focused image $F(j,k)$ of the object is obtained from the image sequence and the function $i(j,k)$ as

$$F(j,k) = g(i(j,k),j,k) \tag{12}$$

In order to find the function $i(j,k)$ which specifies the FIS, we use the fact that the focus measure of $F(j,k)$ (or $g(i(j,k),j,k)$) is a maximum over all possible functions. Since a search for a function is computationally expensive, a two step procedure is used in our implementation.

In the first step of the algorithm, a rough estimate of FIS is estimated using a traditional SFF method as follows. A small set (about 10) of N image frames g_n at regular intervals of I/N are selected from the original image sequence g_i. For each selected image frame, focus measures are computed in small image windows of size $M \times M$ (value of M is about 20). Usually overlapping windows are used at intervals of about $M/4$. Non-overlapping windows may also be used. For each window, the image frame for which the focus measure is a maximum over all the image frames is determined. This gives a very rough estimate of the FIS. This estimate can be improved through an interpolation scheme. For each window, centered around the image frame for which the focus measure was a maximum, a curve is fitted to the focus measures across image frames and the location of the maximum of the curve is computed. Usually a quadratic or cubic fit is used. A Gaussian function fit is used by some researchers. Having obtained a rough estimate of the FIS, an approximate estimate of the slope of the FIS can be obtained by computing partial derivatives along x and y directions.

In the second step of our algorithm, the initial estimate of FIS is refined as follows. In this step, the entire original image sequence g_i containing I image frames is used. For every window in which the FIS was estimated in the first step, a small cubic volume (about the size of $M/4 \times M/4 \times M/4$) image space is considered in the image sequence. The volume is centered at the initial estimate of FIS in that window. Now in this volume, a search is made for a planar surface which is closest to the actual FIS by maximizing the focus measure computed over the planar surface. The initial estimates of position and orientation of the FIS are used as starting values during the search. A brute-force or a simple gradient ascent search can be used. Thus an accurate piece-wise planar approximation to the actual FIS can be obtained. At the cost of additional computation, it also possible to search and obtain piece-wise quadratic and higher order approximations to the FIS.

5 SPARCS

The SFF method described here was implemented on a camera system named Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). SPARCS was built by us in our laboratory over the last few years. A block diagram of the system is shown in Figure ??. SPARCS consists of a SONY XC-77 CCD camera and an Olympus 35-70mm motorized lens. Images from the camera are captured by a frame grabber board (Quickcapture DT2953 of Data Translation). The frame grabber board resides in an IBM PS/2 (model 70) personal computer. The images taken by the frame grabber are processed in the PS/2 computer.

The lens system consists of multiple lenses and focusing is done by moving the front lens forward and backward. To facilitate computer control of the lens movement there is a stepper motor with 97 steps, numbered 0 to 96. Step number 0 corresponds to focusing an object at distance infinity and step number 96 corresponds to focusing a nearby object, at a distance of about 50cm from the lens. The motor is controlled by a microprocessor, which can communicate with the IBM PS/2 through a digital I/O board (Contec mPIO24/24). Pictures taken by the camera can be displayed in real time on a color monitor (SONY PVM-1342 Q). The images acquired and stored in the IBM PS/2 can be transferred to a SUN workstation. The camera settings used in the experiments were: focal length = 35mm, aperture diameter = 35/4 mm, and camera gain control = +6dB.

6 Experiments

The SFF algorithm described above was implemented on the SPARCS camera system. Here we present the results for two objects: (i) a slanted planar object (Figure ??), and (ii) a cone object of length 79 inches and diameter 15 inches (Figure ??). The illumination for the two objects was about 600 lux. Image size was 360×360. In order to reduce electronic noise, for a fixed lens position five image frames were time averaged. The image sequence contained 97 image frames, one for each lens step position of the stepper motor. The absolute displacement between two consecutive image frames was about 0.03 mm and the distance between pixels was about 0.013 mm. The window size for computing focus measures was 20×20.

An initial estimate of FIS was obtained by computing M_2' over 9 image frames equally spaced apart (about 10 frames apart) in the original image sequence. The position of the maximum of focus measure was first improved by a quadratic interpolation scheme using three points centered at the maximum point. An typical plot of the 9 values along with the position of maximum position obtained using interpolation is shown in Figure 4.

Figure **??** shows the results for the slanted planar object. Figs. 5(a) to 5(d) show the image frames recorded when the lens position was at motor steps 20, 40, 60, and 80. In each of these frames, only one part of the image is focused whereas the other parts are blurred by varying degrees. This is particularly noticeable in Fig. 5(d) where the closer part of the object on the left is focused whereas the blur increases gradually towards right as the object distance increases. The shape or depth-map recovered by our SFF algorithm is shown in Fig. 5(e). The results in this case are close to the actual shape except in regions where there is insufficient contrast. The reconstructed focused image of the object is shown in Fig. 5(f). We see that all parts of the image are sharp focus.

Figure 6 is similar to Figure 5 except that the results in this case are for the cone object. In Fig. 6(e) we see that the recovered shape of the cone has a blunt tip rather than a sharp tip. This is due to the piece-wise planar approximation. Except in areas where there is insufficient grey level variance, the shape recovered is good. Figure 6(f) shows the reconstructed focused image of the cone object. In comparison with the image frames shown in Fig. 6(a) to 6(d), the reconstructed image appears focused every where.

7 Conclusion

References

[1] M. Born and E. Wolf, *Principles of Optics,* Pergamon Press, Oxford, Sixth Edition, 1980.

[2] J. Enns and P. Lawrence, "A Matrix Based Method for Determining Depth from Focus", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1991.

[3] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics,* John Wiley & Sons, New York, 1978.

[4] E. Hecht, *Optics,* Addison-Wesley Publishing Co., 1987.

[5] B. K. P. Horn, "Focusing", Artificial Intelligence Memo No. 160, MIT, May 1968.

[6] B. K. P. Horn, *Robot Vision,* McGraw-Hill Book Company, 1986.

[7] R. A. Jarvis, "Focus optimization criteria for computer image processing", Microscope 24, pp.163, 1976.

[8] E. Krotkov, "Focusing", *International Journal of Computer Vision*, 1, pp. 223-237, 1987.

[9] G. Ligthart and F. Groen, "A Comparison of Different Autofocus Algorithms", *International Conference on Pattern Recognition*, pp. 597-600,1982.

[10] S. Nayar, "Shape from Focus System for Rough Surfaces", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp. 302-308, June 1992.

[11] A. P. Pentland, "A new sense for depth of field", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 4, pp. 523-531, July 1987.

[12] J. F. Schlag, A. C. Sanderson, C. P. Neuman, and F. C. Wimberly, "Implementation of automatic focusing algorithms for a computer vision system with camera control", CMU-RI-TR-83-14, Robotics Institute, Carnegie-Mellon University, 1983.

[13] W.F. Schreiber, *Fundamentals of Electronic Imaging Systems*, Springer-Verlag,1986.

[14] M. Subbarao, "Parallel depth recovery by changing camera parameters", *Second International Conference on Computer Vision,* Florida, USA, pp. 149-155, December 1988.

[15] M. Subbarao, "Efficient depth recovery through inverse optics", Editor: H. Freeman, *Machine Vision for Inspection and Measurement,* Academic press, Boston, pp. 101-126, 1989.

[16] M. Subbarao, "Computational methods and electronic camera apparatus for determining distance of objects, rapid autofocusing, and obtaining improved focus images", U.S. patent application serial number 07/373,996, June 1989 (pending).

[17] M. Subbarao, "Determining distance from defocused images of simple objects", Tech. Report No. 89.07.20, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

[18] M. Subbarao, T. Choi, and A. Nikzad, "Focusing Techniques", *OE/BOSTON '92, SPIE conference*, Boston, November 1992.

[19] M. Subbarao, N. Agarwal, and G. Surya, "Application of Spatial-Domain Convolution/DeconvolutionTransform for Determining Distance from Image Defocus", *OE/BOSTON '92, SPIE conference*, Boston, November 1992.

[20] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing: A Practical Approach", Tech. Report No. 92.01.17, CVL, Dept. of EE, SUNY, Stony Brook, NY 11794-2350, 1992. (An abridged version of this appears in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, pp.773-776 , June 15-18,1992.

[21] J. M. Tenenbaum, "Accommodation in Computer Vision", Ph.D. Dissertation, Stanford University, November 1970.