Depth from Defocus by Changing Camera Aperture: A Spatial Domain Approach

Gopal Surya Murali Subbarao
Department of Electrical Engineering
State University of New York
Stony Brook, NY 11794-2350
email : gopal@roger.ee.sunysb.edu

ABSTRACT

This paper describes the application of a new Spatial-Domain Convolution/ Deconvolution transform (S transform) for determining distance of objects and rapid autofocusing of camera systems using image defocus. The method known as STMAP involves simple local operations *on only two images taken with different aperture diameters and can be easily implemented in parallel. Both images can be arbitrarily blurred and neither of them needs to be a focused image taken with a pin-hole camera. STMAP has been implemented on an actual camera system named SPARCS. Experiments on the performance of STMAP and their results on real-world objects are presented. The results indicate that STMAP is useful in practical applications. The utility of the method is demonstrated for rapid autofocusing of electronic cameras. STMAP is computationally more efficient than other methods and the results are comparable to a Fourier transform based approach [15]. When combined with a Depth-from-Focus method or a stereo ranging method, STMAP can reduce the computations by about one order of magnitude.*

CATEGORY

Shape and Object Representation

Depth from Defocus by Changing Camera Aperture: A Spatial Domain Approach

ABSTRACT

This paper describes the application of a new Spatial-Domain Convolution/ Deconvolution transform (S transform) for determining distance of objects and rapid autofocusing of camera systems using image defocus. The method known as STMAP involves simple local operations *on only two images taken with different aperture diameters and can be easily implemented in parallel. Both images can be arbitrarily blurred and neither of them needs to be a focused image taken with a pin-hole camera. STMAP has been implemented on an actual camera system named SPARCS. Experiments on the performance of STMAP and their results on real-world objects are presented. The results indicate that STMAP is useful in practical applications. The utility of the method is demonstrated for rapid autofocusing of electronic cameras. STMAP is computationally more efficient than other methods and the results are comparable to a Fourier transform based approach [15]. When combined with a Depth-from-Focus method or a stereo ranging method, STMAP can reduce the computations by about one order of magnitude.*

1 Introduction

Determining distance of objects using image defocus information has been investigated by many researchers. Most of the previous work deals with methods involving processing a large number of images [3, 6, 7, 17], or simple objects such as edges [2, 11, 18], or camera systems with restricted class of point spread functions [9, 12], or computing Fourier coefficients of images [15]. Recently some researchers [5, 9, 15] have proposed methods for finding distance of an object which does not involve focusing the object. They take the level of defocus of the object into account in determining distance. This approach known as *Depth-from-Defocus* (DFD) does not involve searching for any camera parameter values that maximize a focus measure of the image. Therefore the DFD methods require processing only a few images (two or three), irrespective of whether the objects are focused or not.

Pentland [9] proposed a DFD method for finding the distance of an object which required acquiring the focused image of the object. The focused image was obtained by setting the aperture diameter to be very small (pin-hole) and then acquiring an image. A very small aperture has two problems: (i) it increases diffraction effects thus distorting the acquired image, and (ii) it increases the exposure period of the camera. Subbarao and Wei [15] have proposed a DFD method based on a Fourier Domain approach. The performance of their method is comparable to the method presented here. Enns and Lawrence [5] have proposed a method based on a spatial domain analysis of two blurred images. It is a matrix based method and employs an iterative regularization approach in the presence of noise.

A new transform named Spatial-Domain Convolution/Deconvolution Transform or S Transform was introduced recently [14]. S transform provides a direct spatial domain method for both convolution and deconvolution of images. It is useful in the restoration of defocused images, and in determining distance of objects from a camera system. The S Transform has been successfully applied for distance estimation by using two blurred images taken with different sets of camera parameters [16]. The method reported in [16], called the S Transform method (STM), required two blurred images of an object taken with two different lens positions. The results obtained for this method on a large number of experiments (about 350) yielded an RMS error of about 2.25 % in lens position. In addition to being novel, STM has one important advantage over existing methods in that it involves only local operations on images. Therefore it can yield denser depth-maps, and can be easily implemented in parallel. Although it was mentioned that the two blurred images could be obtained by changing any of the camera parameters (such as aperture, wavelength of light, focal length etc.,) the experimental results were obtained only by moving the lens.

In this paper we describe the implementation of the S Transform method by changing the aperture diameter (we shall call it STMAP). As before, it is applicable to arbitrary objects and does not impose restrictions on the point spread function of camera systems. Further, the method does not involve the computation of Fourier coefficients of images. STMAP has been implemented on an actual camera system. The results indicate that STMAP is useful in practical applications such as passive ranging in robotic vision and rapid autofocusing of camera systems.

In Section 2 we briefly review the camera model and image defocus model used in this paper. In Section 3 we give some examples of the S transform for third order polynomials. Section 4 describes the theory of STMAP. Implementation of STMAP, experiments, and their results are presented in the subsequent sections.

2 Camera model

Image formation in a simple camera is shown in Fig. 1. Let P be a point on a visible surface in the scene and p be its focussed image. Let the light energy incident on the camera aperture in one exposure period of the camera be 1 unit. The relation between the positions of P and p is given by the lens formula,

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

where u is the object distance and v is the image distance. If P is not in focus it gives rise to a blurred image. According to geometric optics, the blurred image of P has the same shape as the lens aperture but scaled by a factor. Let the blurred image of the point P be $h(x, y)$. Clearly $h(x, y)$ is the response of the camera to a point source and hence $h(x, y)$ is the point spread function.

Usually camera systems have a circular aperture. In this case the blurred image of a point on the image detector is circular in shape and is called the *blur circle.* Let R be the radius of the blur circle and D be the diameter of the lens aperture, and s be the distance from the lens to the image detector plane (Fig. 1). Also let q be the scaling factor defined $q = 2R/D$. In Fig. 1, from similar triangles we have

$$R = q\frac{D}{2} = s\frac{D}{2}\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right] \tag{2}$$

Note that q and therefore R can be either positive or negative depending on whether $s \geq v$ or $s < v$. In the former case the image detector plane is behind the focused image of P and in the latter case it is in front of the focused image of P. In either case the magnitude of R corresponds to the actual radius of the blur circle. According to geometric optics, the intensity within the blur circle is approximately constant. If we assume the camera to be a lossless system (i.e., no light energy is absorbed by the camera system) then

$$\int\int h(x, y)\ dxdy = 1 \tag{3}$$

because the light energy incident on the lens is taken to be one unit. So we get

$$h_1(x, y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $h_1(x, y)$ is the point spread function of the camera, derived using geometric optics. Taking diffraction and non-idealities of lenses into account, an alternative model has been suggested for the intensity distribution given by a two dimensional Gaussian [9, 12],

$$h_2(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}} \tag{5}$$

where σ is the spread parameter such that

$$\sigma = k\ R \tag{6}$$

where k is a constant ($k > 0$) of proportionality characteristic of a given camera. Except when σ is very small (in which case diffraction effects dominate), in most practical cases

$$k = \frac{1}{\sqrt{2}} \tag{7}$$

is a good approximation [11, 13].

If the radius R is a constant over some region on the image plane, the camera acts as a linear shift invariant system. This is justified because the camera parameters s, D and f all remain the same. Therefore the observed image $g(x, y)$ is the result of convolving the corresponding focused image $f(x, y)$ with the camera's point spread function $h(x, y)$, i.e.,

$$g(x, y) = h(x, y) * f(x, y) \tag{8}$$

where $*$ denotes the convolution operation.

The point spread functions h_1 and h_2 defined above are only two specific examples used to clarify our method. In order to deal with other forms of point spread functions, we use the spread parameter σ_h to characterize them where σ_h is the standard deviation of the distribution of any function h. Using the polar coordinate system it can be shown [11] that the spread parameter σ_{h_1} corresponding to h_1 is $R/\sqrt{2}$. Therefore from equation (2) we have

$$\sigma_{h_1} = mu^{-1} + c \tag{9}$$

where

$$m = -\frac{Ds}{2\sqrt{2}} \text{ and } c = \frac{Ds}{2\sqrt{2}}\left[\frac{1}{f} - \frac{1}{s}\right] \tag{10}$$

We see that for a given camera setting (i.e., for a given value of the camera parameters s, f, D) the spread parameter σ_{h_1} depends linearly on inverse distance u^{-1}. Similarly it can be shown that the spread parameter σ_{h_2} of h_2 is σ. Therefore from equations (6), (7) and (2) we again obtain

$$\sigma_{h_2} = mu^{-1} + c. \tag{11}$$

In fact it has been shown in [13] that even for an arbitrarily shaped aperture, σ_h is linearly related to inverse distance u^{-1}.

In a practical camera system, if two images $g_1(x, y)$ and $g_2(x, y)$ are taken with different aperture diameters, then the mean image brightness will change even though nothing has changed in the scene. In order to compare the blur in images g_1 and g_2 in a correct and consistent manner, they must first be normalized with respect to mean brightness. It is carried out by dividing the image brightness at every point by the mean brightness of the image.

3 S Transform

A detailed discussion of S transform can be found in [14]. In this section we summarize some results relevant to STMAP.

Let $f(x, y)$ be a two variable cubic polynomial defined by

$$f(x, y) = \sum_{m=0}^{3} \sum_{n=0}^{3-m} a_{m,n} x^m y^n \tag{12}$$

where $a_{m,n}$ are the polynomial coefficients.

Let $h(x, y)$ be a rotationally symmetric point spread function. The moments of the point spread function are defined by

$$h_{m,n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^m y^n h(x, y) dx dy \tag{13}$$

Now consider the convolution of the image $f(x, y)$ and the point spread function $h(x, y)$

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - \zeta, y - \eta) h(\zeta, \eta)\ d\zeta d\eta \tag{14}$$

Since f is a cubic polynomial, it can be expressed in a Taylor series as

$$\begin{aligned}
f(x - \zeta, y - \eta) &= \sum_{0 \le m+n \le 3} \frac{(-\zeta)^m}{m!} \frac{(-\eta)^n}{n!} f^{m,n}(x, y) && (15) \\
\text{where} \qquad f^{m,n}(x, y) &\equiv \frac{\partial^m}{\partial x^m} \frac{\partial^n}{\partial y^n} f(x, y) && (16) \\
\Rightarrow g(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{0 \le m+n \le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x, y) \zeta^m \eta^n h(\zeta, \eta) d\zeta d\eta && (17) \\
&= \sum_{0 \le m+n \le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x, y) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \zeta^m \eta^n h(\zeta, \eta) d\zeta d\eta && (18) \\
&= \sum_{0 \le m+n \le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x, y) h_{m,n} && (19)
\end{aligned}$$

Equation (19) expresses the convolution of a function $f(x, y)$ with another function $h(x, y)$ as a summation involving the derivatives of $f(x, y)$ and moments of $h(x, y)$. This corresponds to the *forward S-Transform.* Now let us use this formula to derive a deconvolution formula. Since $h(x, y)$ is circularly symmetric it can be shown that

$$h_{0,1} = h_{1,0} = h_{1,1} = h_{0,3} = h_{3,0} = h_{2,1} = h_{1,2} = 0 \text{ and } h_{2,0} = h_{0,2} \tag{20}$$

Also from equation (3)

$$h_{0,0} = 1 \tag{21}$$

Therefore we obtain

$$g(x,y) = f(x,y) + \frac{h_{2,0}}{2}\left(f^{2,0}(x,y) + f^{0,2}(x,y)\right) \tag{22}$$

$$\Rightarrow f(x,y) = g(x,y) - \frac{h_{2,0}}{2}\left(f^{2,0}(x,y) + f^{0,2}(x,y)\right) \tag{23}$$

Applying $\frac{\partial^2}{\partial x^2}$ to the above expression on either side

$$f^{2,0}(x,y) = g^{2,0}(x,y) \tag{24}$$

Similarly applying $\frac{\partial^2}{\partial y^2}$ we get

$$f^{0,2}(x,y) = g^{0,2}(x,y) \tag{25}$$

$$\begin{aligned}
\text{Therefore } f(x,y) &= g(x,y) - \frac{h_{2,0}}{2}\left(g^{2,0}(x,y) + g^{0,2}(x,y)\right) \\
&= g(x,y) - \frac{h_{2,0}}{2} \bigtriangledown^2 g(x,y)
\end{aligned} \tag{26}$$

where $\bigtriangledown^2$ is the Laplacian operator. Equation (26) is a deconvolution formula. It expresses the original function $f(x,y)$ in terms of the convolved function $g(x,y)$, its derivatives and the moments of the point spread function. In the general case this corresponds to the *Inverse S-Transform* [14]. In the following section we describe the application of this formula to the problem of distance estimation from blurred images.

4 Determining Distance

STMAP is based on approximating the image function $f(x,y)$ as a polynomial function in small regions. (Note: At least in principle, for any given analytic finction, a polynomial arbitrarily close to it exists, according to a famous theorem by Weirstrass, as is well known in Approximation Theory.) In our application, usually a third order polynomial approximation in image neighborhoods of size about 9×9 pixels suffices. This assumption is specified by

$$f(x,y) = \sum_{0 \le m+n \le 3} a_{m,n} x^m y^n \quad \begin{cases} A1 \le x \le A2 \\ B1 \le y \le B2 \end{cases} \tag{27}$$

From the definition of moments and σ_h we have $h_{0,2} = h_{2,0} = \sigma_h^2/2$. So

$$f(x,y) = g(x,y) - \frac{\sigma_h^2}{4} \ \nabla^2 g(x,y) \tag{28}$$

Let us consider a blurred image $g_1(x,y)$ of the object and let σ_1 be the corresponding spread parameter. Now, the original (focused) image $f(x,y)$ can be expressed in terms of $g_1(x,y)$ as

$$f(x,y) = g_1(x,y) - \frac{\sigma_1^2}{4} \ \nabla^2 g_1(x,y) \tag{29}$$

Similarly $f(x,y)$ can also be expressed in terms of a second blurred image $g_2(x,y)$ as

$$f(x,y) = g_2(x,y) - \frac{\sigma_2^2}{4} \ \nabla^2 g_2(x,y) \tag{30}$$

Equating the right hand sides of equations (29) and (30) we have

$$g_1(x,y) - \frac{\sigma_1^2}{4} \ \nabla^2 g_1(x,y) = g_2(x,y) - \frac{\sigma_2^2}{4} \ \nabla^2 g_2(x,y) \tag{31}$$

If R_1 and R_2 are the radii of blur circles corresponding to two aperture diameters D_1 and D_2 respectively, then from equation (2) it can be seen that

$$\frac{R_1}{R_2} = \frac{D_1}{D_2} \tag{32}$$

And since $\sigma = \frac{R}{\sqrt{2}}$ we have

$$\frac{\sigma_1}{\sigma_2} = \frac{D_1}{D_2} \tag{33}$$

Therefore

$$\sigma_1 = A\sigma_2 \text{ where } A = \frac{D_1}{D_2} \tag{34}$$

Using this result in equation (31) we get

$$\sigma_2 = \sqrt{\frac{4(g_1 - g_2)}{(A^2 \nabla^2 g_1 - \nabla^2 g_2)}} \tag{35}$$

In the above equation the variables x and y have been left out for notational convenience. The Laplacians $\nabla^2 g_1$ and $\nabla^2 g_2$ are computed from the two observed images g_1 and g_2. We can then calculate the value of σ_2 from equation (35). The distance u of the object is then obtained from equation (11).

However due to noise, the two focused images derived from the two blurred images may not be exactly the same and equation (31) may not be valid. In order to make the method robust in the presence of noise, a small variation was used, which is described below. We have from equation (31)

$$g_1 - g_2 = \frac{1}{4}\ (\sigma_1^2 - \sigma_2^2)\ \ \nabla^2 g_1 \tag{36}$$

where we have made use of the fact that for a third order polynomial $\nabla^2 g_1 = \nabla^2 g_2$. Squaring and integrating over a small region around the point (x, y) we get

$$\int\int (g_1 - g_2)^2 dxdy = \frac{1}{16}(\sigma_1^2 - \sigma_2^2)^2\ \int\int (\nabla^2 g_1)^2\ dxdy \tag{37}$$

which can be expressed as

$$\begin{aligned} (\sigma_1^2 - \sigma_2^2)^2 &= G^2 \\ \Rightarrow \sigma_1^2 - \sigma_2^2 &= \pm G \qquad (38) \\ \text{where } G^2 = 16\ \frac{\int\int (g_1 - g_2)^2 dxdy}{\int\int (\nabla^2 g_1)^2\ dxdy} & \qquad (39) \end{aligned}$$

Combining equations (38) and (34) we get

$$\begin{aligned} A^2\sigma_2^2 - \sigma_2^2 &= \pm G \\ \Rightarrow \sigma_2 &= \sqrt{\frac{\pm G}{A^2 - 1}} \qquad (40) \end{aligned}$$

The distance u of the object is then obtained from equation (11). Ideally it should be possible to compute the value of σ_2 at one single point (x, y) in the image and obtain an estimate of distance from it. But because the third degree polynomial approximation may not be valid at all points of the image, it is required to compute the values of σ_2 at many points in a region. When σ_2 is calculated for many points, each point may give rise to a different value of σ_2. In order to pick the correct value of σ_2 a histogram of number of pixels verses σ_2 is generated and the mode of the distribution is chosen to be the correct value of σ_2. It can be interpreted as if each pixel in a region votes for a value of σ_2 and that value of σ_2 which receives the maximum number of votes is chosen as the correct value of σ_2. However due to non idealities of the lens system a correction may have to be applied to the value of u so obtained. A more effective method would be to use a "look-up table" from which it would be possible to obtain the distance of the object corresponding to any value of σ_2. In fact our implementation of this method uses this kind of a look-up table.

5 Implementation

STMAP described above was implemented on a camera system named Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). SPARCS was built over the last few years in the Computer Vision Laboratory at the Department of Electrical Engineering, State University of New York, Stony Brook. A block diagram of the system is shown in Figure 2. 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 focal length of the lens can be varied manually from about 35mm to 70mm. The F-number which is defined as the ratio of the focal length f to aperture diameter D can also be set manually to 4, 8, 22 etc.,. The lens system consists of multiple lenses and focusing is done by moving the front lens forward and backward. The lens can be moved either manually or under computer control. 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 50 cm 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.

As mentioned earlier, there are 97 step positions for the stepper motor, and therefore there are 97 distinct focusing positions of the lens. For each position of the lens, there corresponds a unique distance such that an object placed at that distance from the camera would be in focus. For convenience we shall use lens step positions to specify and measure distances. For example, if the distance of an object is said to be step 30, we mean that the distance of the object is such that if the lens is moved to step number 30, the object would be in best focus.

5.1 Camera Calibration

The first step in the implementation of STMAP is calibration. During calibration a table of σ versus object distance is generated. In the present set of experiments, only one camera parameter, the aperture diameter D was varied. All other parameters (such as focal length f and lens position s) were constant. The lens step was chosen to be zero (focusing at ∞). The steps involved in the calibration are given in the flow chart of Figure 10. We place an object at a given step distance and take two pictures $g_1(x, y)$ and $g_2(x, y)$ with different aperture settings. In our case the two apertures correspond to F-Number 4 and F-Number 8 (this corresponds to about 35/4 mm and 35/8 mm diameters respectively). Once we have the two blurred images $g_1(x, y)$ and $g_2(x, y)$ the next step is to smooth them and obtain their Laplacians $\bigtriangledown^2 g_1$ and $\bigtriangledown^2 g_2$. Smoothing and Laplacian estimation were done using the smoothed differentiation filters proposed by Meer and Weiss [8]. With the knowledge of camera parameters, the value of σ_2 can be obtained for each point (x, y) in the image using equation (40). This computation is repeated for all pixels in the image. The allowable range of σ_2 is divided into about 200 divisions and with every pixel giving rise to a value of σ_2, a histogram of number of pixels versus σ_2 is generated. The histogram is smoothed using the Parzen Window method with window size 5. The mode of the histogram is chosen to be the correct value of σ_2. Thus for one particular image and one particular object distance we have obtained the value of σ_2. This value is recorded and the procedure is repeated for all object distances corresponding to step 10 through step 95, with increments of 5 steps. Similar procedure is repeated with different objects. The values of sigma thus obtained for lens step 0 are shown in Table 2. Another sigma table was generated by keeping the lens position at 60, which is shown in Table 3.

In each row of Tables 2 and 3, the mean value of the three recorded entries was calculated. The results are plotted in Figure 4. In this Figure, the plot with label “step0.sigma” corresponds to Table 2 and the plot with label “step60.sigma” corresponds to Table 3. These plots essentially constitute our look-up tables. In Figure 4 the X-axis corresponds to object distances (in lens step numbers) and the Y-axis corresponds to the sigma values. As can be seen from Table 2, when the object distance is closer than step 60 (i.e., when the step number is greater than 60) , there is a large variation in sigma from object to object. Hence it will be necessary to use the calibration data of Table 3. There is a two fold ambiguity in the plot of Table 3 because we are squaring and integrating in equation (37). However we use only one side of the folded curve (for step numbers greater than 60). For step numbers less than 60, the plot "step0.sigma" is used.

The program can be run in the Calibration Mode during which the user just places an object facing the camera and inputs the actual distance. The sigma values are computed and saved in a file.

5.2 Distance Estimation

Once calibration is done, the program can be run in the depth estimation mode. It is similar to the calibration mode, as far as obtaining the sigma values are concerned. That is, we take two blurred images of an object with two different aperture settings and calculate the values of σ_2. We then have to look up the calibration table to find the object distance. If the calculated sigma value lies in between two successive values of the calibration table, linear interpolation is used to determine the object distance. Once the object distance is known the step number to which the lens has to be moved in order to focus the image, is easily calculated. The lens is then moved to make the image appear focused (we noticed that in our camera system there is an assembly error and we need to offset the lens positions calculated by the program by about 12 steps).

Experiments were conducted on four different objects at room illumination (about 200-300 Lux) and ten different objects at 400 Lux illumination. All these images have been saved as a database. Some of the objects in the database are shown in Figure 9. For each object the experiment was repeated by moving the object to different distances from step 10 to step 95 in steps of 5. Thus the total number of experiments is $18 \times 14 = 252$. The results are tabulated in Tables 4 and 5. The mean values of the results and the standard deviations are also shown in these tables. The mean values of these results are plotted in Figure 5 and Figure 6. The results seem to be well within the noise margin. The overall RMS error is about 2.25 steps out of 97 steps.

Figure 3 is a plot of the reciprocal of the object distance $1/u$ versus the lens step number. The relationship is almost linear and can be expressed as

$$1/u = ax + b \tag{41}$$

where x specifies lens position. For our camera, the lens position is specified in terms of a motor step number where each step corresponds to a displacement of about 0.03mm. The RMS errors mentioned above are for the lens position and it gives a good indication of the performance of the method for application in rapid autofocusing of cameras. In order to compute the error in terms of object distance, we have to consider the error differentials in equation (41).

$$\begin{aligned} |\delta(1/u)| &= a|\delta x| & (42)\\ \Rightarrow |\frac{\delta u}{u}| &= a|\delta x|u & (43)\\ \Rightarrow |\delta u| &= a|\delta x|u^2 & (44)\end{aligned}$$

From the above relations we see that the relative (percentage error) $|\frac{\delta u}{u}|$ in actual distance u increases linearly with distance, and the absolute error $|\delta u|$ in actual distance increases quadratically with distance. Based on the lens data of Table 1, for our camera system the constant $a = 0.0172$.

We have also implemented a Depth-from-Focus method (DFF) on our camera system based on a focus measure proposed in [17]. The DFF methods usually take a number of images (about 10-12) and search for the sharpest focus position by maximizing some focus measure. Many different focus measures have been proposed and their performances are nearly the same. Since DFF methods involve exhaustive search for the focused position, we assume that the results that can be obtained by a DFD method (which takes just 2-3 images) can at best be equal to a DFF method. Hence we shall call the results obtained by the DFF method as DFF.BST and it serves as a benchmark to evaluate STMAP. A number of experiments were performed with the DFF method using the same objects used for the STMAP experiments and yielded an RMS error of 1.52 steps.

Setting $|\delta x|$ to be the RMS error of 1.52 steps and 2.25 steps respectively, a plot of relative error $|\frac{\delta u}{u}|$ is shown in Figure 7 and a plot of the absolute error is shown in Figure 8. In Figure 7 we see that for STMAP the percentage error in distance at 0.6 meter is about 2.3% and increases linearly to about 20% at 5 meter distance. This compares well with the best possible error of about 1.6%

at 0.6 meter and increasing linearly to about 12.5 % at 5 meter distance that is obtained with the DFF method. Figure 8 shows that for STMAP, absolute error increases quadratically from 1.3 cms at 0.6 meter to about 1.0 meter at 5 meters distance. The corresponding numbers for the DFF method are 1 cm at 0.6 meter and about 0.6 meter at 5 meters distance.

The camera settings used in the experiment were (i)Focal Length = 35mm. (ii) F- Numbers = 4 and 8. (iii) Camera Gain Control = +6dB.

6 Conclusions

We have described the theory and implementation of a new DFD method for determining depth from image defocus information. The method uses two blurred images of an object taken with two different aperture settings. Both the aperture settings can be arbitrarily chosen and neither of them needs to be a pin-hole. The method known as STMAP has been successfully demonstrated on an actual camera system built by us. Experimental results indicate that STMAP is useful for passive ranging and rapid autofocusing. The ranging accuracy is high for nearby objects and decreases with increasing distance. This method can be combined with a Depth-from-Focus method to reduce the percentage error by a factor of about 2 at the additional cost of acquiring and processing a few (about 3) more images. The combination will then be a good trade off between speed and accuracy.

In comparison with the stereo method of ranging, the STM methods do not suffer from the *correspondence* problem, but it is in general less accurate than stereo vision. Therefore the STM methods can be used to get a rough estimate of distance which can then be used by a stereo algorithm to determine more accurate distance. The computation associated with establishing correspondence is reduced due to the availability of a rough estimate of distance.

The distance of "plain" objects such as white walls which do not exhibit reflectance variation under uniform illumination cannot be determined by the STM methods. However a random illumination pattern can be projected onto such objects to make them "textured". The STM methods can then be used.

Most existing camera systems (including our camera) are designed to maximize the depth-of-field since the goal is to obtain a "good" image of the scene for viewing by humans. However this minimizes the accuracy when ranging is concerned, since maximizing depth of field reduces the difference in blur between objects at different distances. Therefore, STM methods can be made much more accurate by designing cameras with small depth of field for the purpose of ranging.

Acknowledgements: The support of this research by the National Science Foundation and Olympus Optical Corporation is gratefully acknowledged.

References

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

[2] P. Grossman, "Depth from focus", *Pattern Recognition Letters* 5, pp. 63–69, Jan. 1987.

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

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

[5] 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.

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

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

[8] P. Meer and I. Weiss, *Smoothed differentiation filters for images,* Tech. Report No. CS-TR-2194, Center for Automation Research, University of Maryland, College Park, MD 20742-3411.

[9] 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.

[10] A. Rosenfeld, and A. C. Kak, *Digital Picture Processing,* Vol. I . Academic Press, 1982.

[11] M. Subbarao, and G. Natarajan, "Depth recovery from blurred edges", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, pp. 498-503, June 1988.

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

[13] 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.

[14] M. Subbarao, "Spatial-Domain Convolution/Deconvolution Transform ", Tech. Report No. 91.07.03, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

[15] M. Subbarao, and T. Wei, "Depth from Defocus and Rapid Autofocusing : A practical Approach", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, Illinois, June 1992

[16] M. Subbarao, and G. Surya, "Application of Spatial-Domain Convolution /Deconvolution Transform for Determining Distance from Image Defocus", SPIE conference, Boston, Nov. 1992.

[17] M. Subbarao, T. S. Choi and A. Nikzad, "Focusing Techniques", SPIE conference, Boston, Nov. 1992.

[18] S. Lai, C. Fu, "A Generalized Depth Estimation Algorithm with a Single Image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, NO. 4, April 1992, pp 405-411.

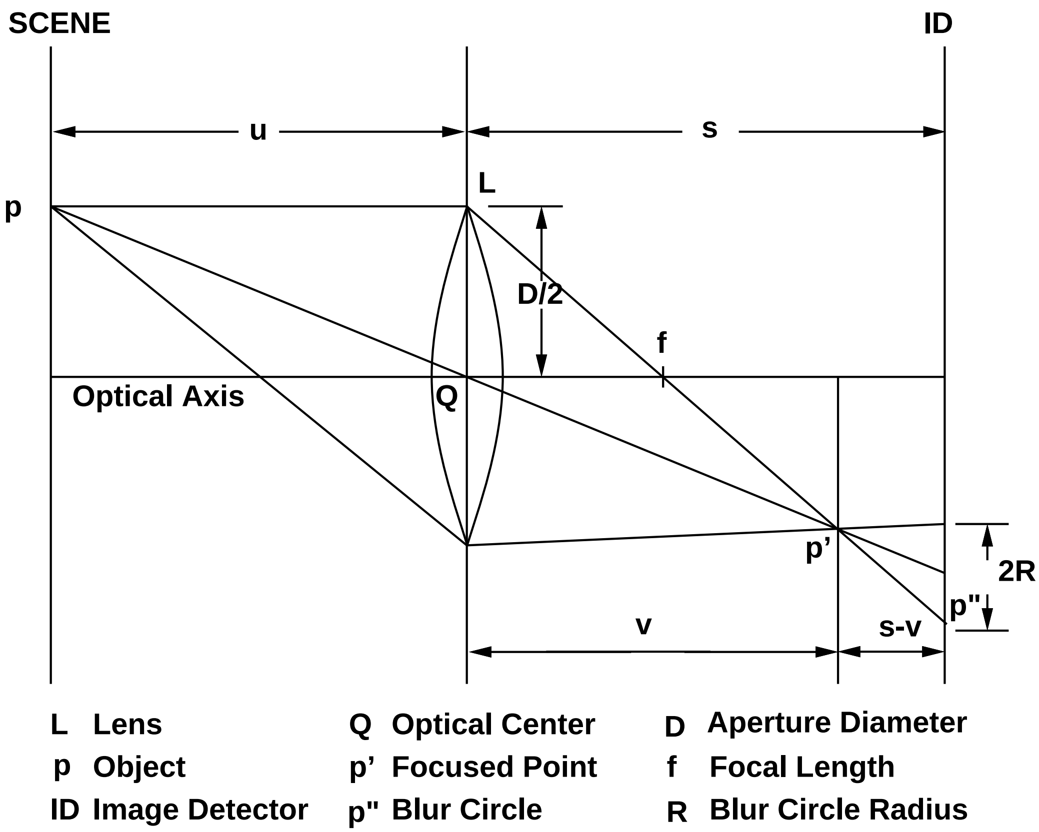
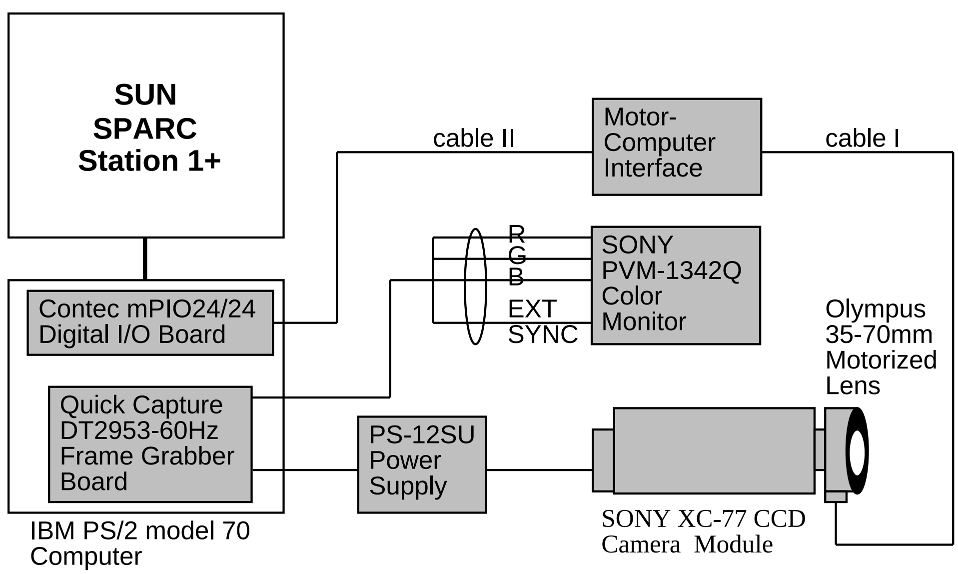


Fig.1 Image Formation in a Simple Camera System



Stonybrook Passive Autofocusing and Ranging Camera System- SPARCS - is a prototype camera system developed at the Computer Vision Laboratory for experimental research in robotic vision, State University of New York at Stony Brook.

Fig.2 Block Diagram of SPARCS

Lens Step	0	5	10	15	20	25	30	35	40	45
Distance(m)	∞	5.300	3.750	2.850	2.500	1.930	1.720	1.465	1.320	1.170
Lens Step	50	55	60	65	70	75	80	85	90	95
Distance(m)	1.080	0.965	0.900	0.822	0.770	0.715	0.670	0.628	0.595	0.560

Table 1. Lens step Vs Best Focused Distance (Lens Data)

Dist.	Sigma Values			
Step	tg	fa	gs	mean
10	2.2	2.3	2.1	2.2
15	2.5	2.6	2.5	2.53
20	2.8	2.9	2.8	2.83
25	3.3	3.4	3.3	3.33
30	3.8	3.8	3.7	3.77
35	4.2	4.2	4.2	4.2
40	4.8	4.7	4.5	4.67
45	5.1	4.9	5.0	5.0
50	5.5	5.5	5.2	5.4
55	5.9	5.7	5.4	5.67
60	6.2	6.1	5.7	6.0
65	6.5	6.4	5.7	6.2
70	6.7	6.6	6.3	6.53
75	7.0	7.0	6.2	6.73
80	7.3	7.2	6.7	7.07
85	7.6	7.1	6.6	7.10
90	8.2	7.1	6.4	7.23
95	8.4	7.9	6.2	7.5

Table 2. Sigma with lens step 0

Dist.	Sigma Values			
Step	tg	fa	gs	mean
10	3.6	3.7	3.5	3.6
15	3.3	3.4	3.3	3.33
20	3.0	3.1	3.0	3.03
25	2.6	2.6	2.6	2.6
30	2.2	2.2	2.2	2.2
35	1.8	1.8	1.8	1.8
40	1.4	1.4	1.4	1.4
45	1.1	1.3	1.0	1.13
50	1.0	1.0	0.8	0.93
55	1.3	1.5	1.0	1.27
60	1.5	1.8	1.4	1.57
65	1.9	1.8	1.9	1.87
70	2.2	2.2	2.3	2.23
75	2.7	2.6	2.8	2.7
80	3.3	3.1	3.2	3.2
85	3.6	3.5	3.6	3.57
90	4.1	4.0	4.1	4.07
95	4.5	4.4	4.4	4.43

Table 3. Sigma with lens step 60

Actual Dist.		Estimated Dist. (step)					
Step	meters	Tgr	Face	GS	Edge	mean	std. dev.
10	3.750	10	11	10	13	11.00	1.22
15	2.850	14	16	14	17	15.25	1.30
20	2.500	19	20	19	21	19.75	0.83
25	1.930	24	25	24	24	24.25	0.43
30	1.720	30	30	29	31	30.00	0.71
35	1.465	35	35	35	37	35.50	0.87
40	1.320	41	40	38	40	39.75	1.09
45	1.170	46	43	45	47	45.25	1.48
50	1.080	51	51	47	50	49.75	1.64
55	0.965	55	58	51	63	56.75	4.38
60	0.900	58	63	57	65	60.75	3.34
65	0.822	65	63	65	70	65.75	2.59
70	0.770	69	69	70	75	70.75	2.49
75	0.715	75	73	74	80	75.50	2.69
80	0.670	81	79	80	84	81.00	1.87
85	0.628	85	84	85	88	85.50	1.50
90	0.595	90	89	90	94	90.75	1.92
95	0.560	95	94	94	96	94.75	0.83

Table 4. Results at normal illumination

Actual Dist.		Estimated Distance (step)											
Step	meters	c1	c2	fa	ft	gl	gs	mk	mn	sb	tg	mean	std. dev.
10	3.750	10	10	11	11	11	10	10	10	10	10	10.3	0.46
15	2.850	17	11	15	17	15	14	14	13	14	14	14.4	1.68
20	2.500	17	17	20	20	20	19	19	20	19	19	19.0	1.09
25	1.930	23	23	25	25	25	25	24	23	25	24	24.2	0.87
30	1.720	28	27	30	30	29	30	30	28	31	29	29.2	1.17
35	1.465	35	32	36	35	35	37	32	33	36	36	34.7	1.68
40	1.320	51	38	42	42	40	40	38	39	40	39	40.9	3.62
45	1.170	46	44	46	45	45	45	43	45	45	44	44.8	0.87
50	1.080	51	50	50	46	50	44	49	51	50	50	49.1	2.16
55	0.965	55	54	55	54	55	55	55	56	56	54	54.9	0.70
60	0.900	63	59	60	60	62	60	60	56	58	59	59.7	1.85
65	0.822	64	68	65	68	63	56	63	62	64	62	63.5	3.23
70	0.770	64	67	70	72	71	64	70	72	67	70	68.7	2.86
75	0.715	74	73	75	75	75	72	73	74	73	75	73.9	1.04
80	0.670	78	78	80	82	95	80	78	80	80	81	81.2	4.78
85	0.628	83	82	85	85	84	84	84	84	84	84	83.9	0.83
90	0.595	89	88	90	92	92	92	88	89	89	90	89.9	1.51
95	0.560	95	95	88	95	95	95	95	95	95	95	94.3	2.10

Table 5. Results at 400 Lux

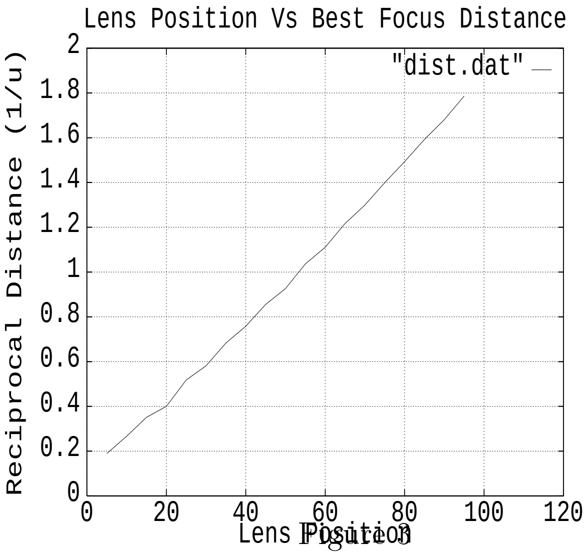


Figure 3

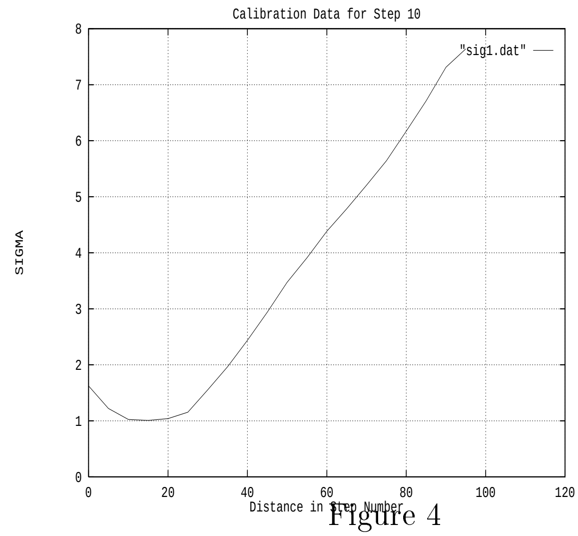


Figure 4

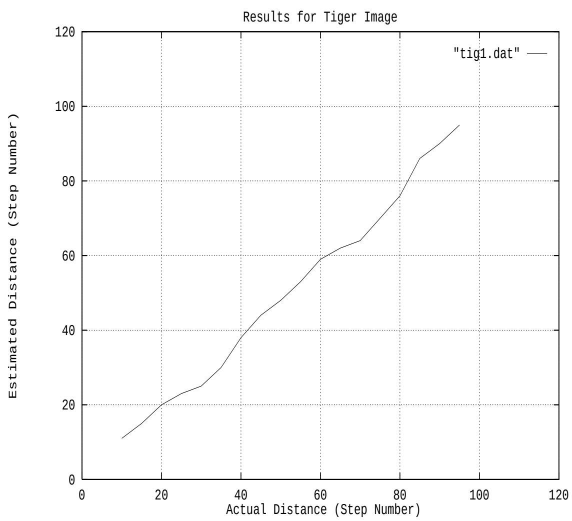


Figure 5

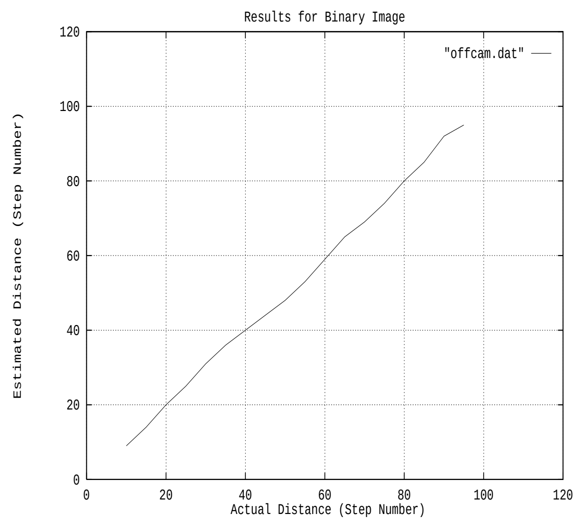


Figure 6

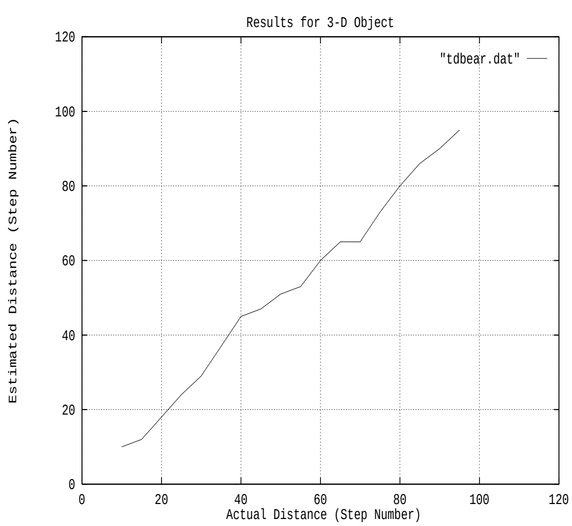


Figure 7

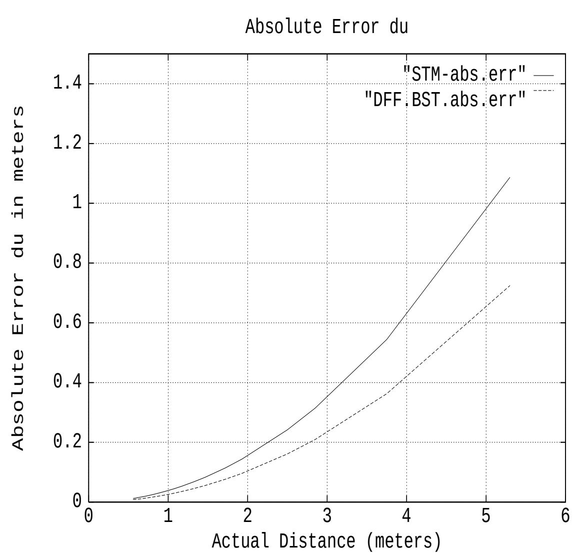
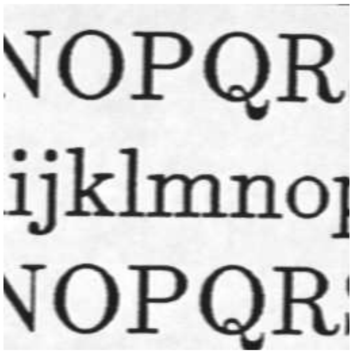


Figure 8



c1

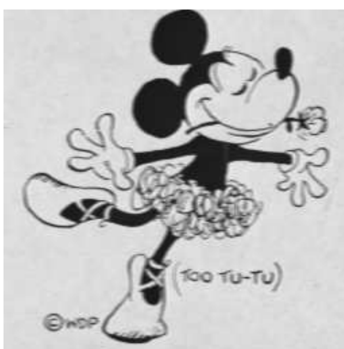

mk

sb


fa

gs


ft



mn


gl


tg

c2

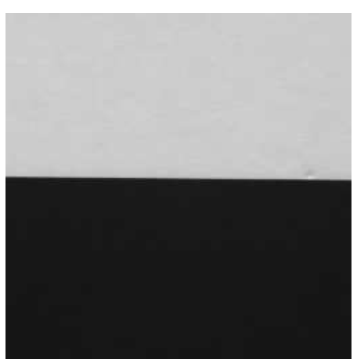

ev

Figure 9. Test Images in the Database

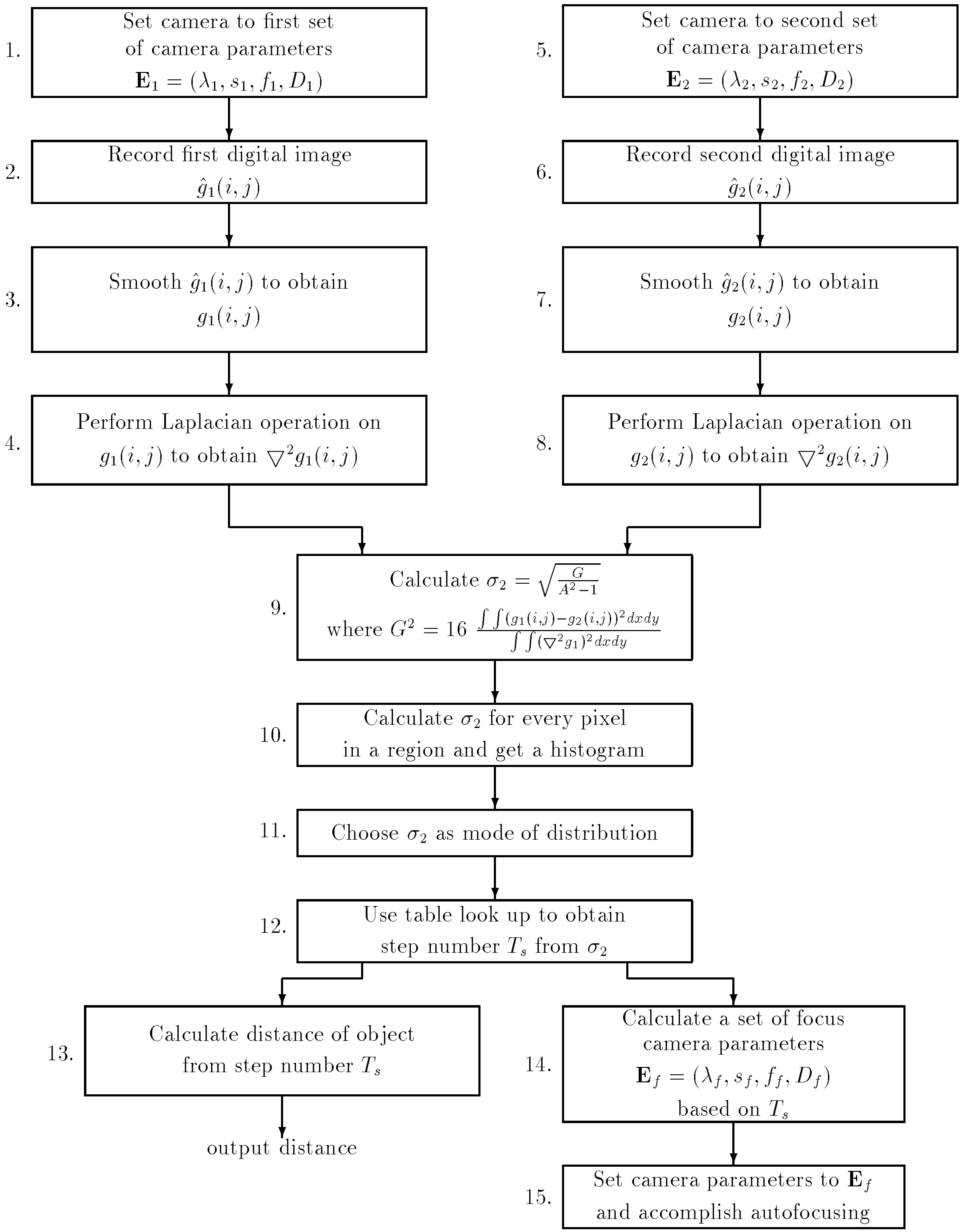

1. Set camera to first set of camera parameters $\mathbf{E}_1 = (\lambda_1, s_1, f_1, D_1)$
2. Record first digital image $\hat{g}_1(i, j)$
3. Smooth $\hat{g}_1(i, j)$ to obtain $g_1(i, j)$
4. Perform Laplacian operation on $g_1(i, j)$ to obtain $\nabla^2 g_1(i, j)$
5. Set camera to second set of camera parameters $\mathbf{E}_2 = (\lambda_2, s_2, f_2, D_2)$
6. Record second digital image $\hat{g}_2(i, j)$
7. Smooth $\hat{g}_2(i, j)$ to obtain $g_2(i, j)$
8. Perform Laplacian operation on $g_2(i, j)$ to obtain $\nabla^2 g_2(i, j)$
9. Calculate $\sigma_2 = \sqrt{\frac{G}{A^2 - 1}}$ where $G^2 = 16 \, \frac{\int\int (g_1(i,j) - g_2(i,j))^2 dx dy}{\int\int (\nabla^2 g_1)^2 dx dy}$
10. Calculate σ_2 for every pixel in a region and get a histogram
11. Choose σ_2 as mode of distribution
12. Use table look up to obtain step number T_s from σ_2
13. Calculate distance of object from step number T_s
output distance
14. Calculate a set of focus camera parameters $\mathbf{E}_f = (\lambda_f, s_f, f_f, D_f)$ based on T_s
15. Set camera parameters to $\mathbf{E}_f$ and accomplish autofocusing

Figure 10. Flow Chart of the Algorithm