International Journal of Computer Vision, 13, 3, 271–294 (1994)

Depth from Defocus: A Spatial Domain Approach

MURALI SUBBARAO AND GOPAL SURYA
Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350
e-mail: murali@sbee.sunysb.edu and gopal@sbee.sunysb.edu

Received December 15, 1992; revised May 6, 1993 and October 1993

Abstract. A new method named STM is described for determining distance of objects and rapid autofocusing of camera systems. STM uses image defocus information and is based on a new Spatial-Domain Convolution/Deconvolution Transform. The method requires only two images taken with different camera parameters such as lens position, focal length, and aperture diameter. Both images can be arbitrarily blurred and neither of them needs to be a focused image. Therefore STM is very fast in comparison with Depth-from-Focus methods which search for the lens position or focal length of best focus. The method involves simple local operations and can be easily implemented in parallel to obtain the depth-map of a scene. STM has been implemented on an actual camera system named SPARCS. Experiments on the performance of STM and their results on real-world planar objects are presented. The results indicate that the accuracy of STM compares well with Depth-from-Focus methods and is useful in practical applications. The utility of the method is demonstrated for rapid autofocusing of electronic cameras.

1 Introduction

Passive techniques of ranging or determining distance of objects from a camera is an important problem in computer vision. Stereo vision (Horn 1986) is perhaps the most popular technique. The major computational problems associated with stereo are the *correspondence problem* and detection of *occlusion*. Recently, Depth-from-Focus (DFF) methods (Horn 1968; Jarvis 1983; Krotkov 1987; Nayar 1992; Schlag et al. 1983; Subbarao et al. 1992; Tenenbaum 1970) have attracted the attention of researchers as they do not suffer from the problems associated with stereo.

DFF methods are based on the fact that in the image formed by an optical system such as a convex lens, objects at a particular distance (or depth) from the lens will be focused whereas objects at other distances will be blurred or defocused by varying degrees depending on their distance. The distance u of focused objects (Figure 1) depends on the focal length f of the lens and the distance v between the lens and the position of the focused image. The relation between f, v and u is expressed by the well-known lens formula

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

The problem of *focusing* is to find and adjust the value of focal length f or the distance s between the lens and the image detector (Figure 1) or both so that a specified object is focused. One way of focusing then is to vary f and/or s in steps until the observed image of the object is in sharpest focus. Once the values of f and s which correspond to focusing the object are found, the distance of the object can be calculated using the lens formula. Therefore DFF is essentially a search method which requires acquiring and processing many images (about 10–12 in practice). Recent work comparing different DFF methods can be found in Krotov (1987), Nayar (1992), and Subbarao et al. (1992). The disadvantage of taking a large number of images is the amount of time required to adjust the camera parameters (lens position and/or focal length) before taking each image. This involves mechanical motion of camera parts which is much slower than electronic computation. During the entire period of adjusting camera parameters, the scene must remain stationary.

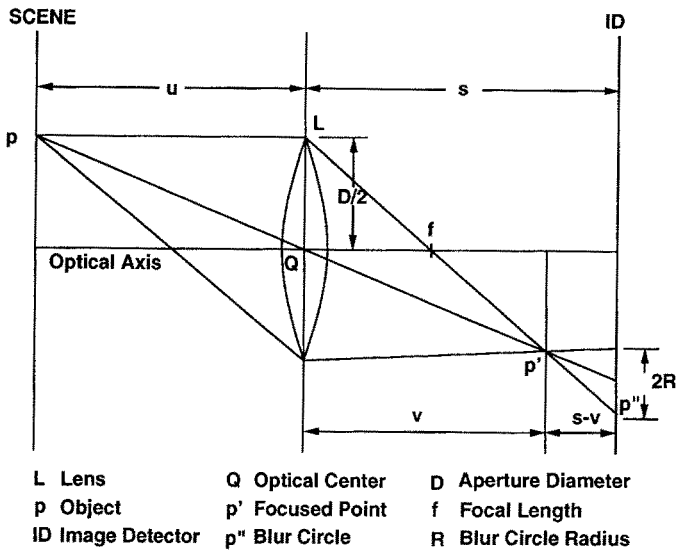


Fig. 1. Image formation in a convex lens.

Recently some researchers (Ens and Lawrence 1991; Pentland 1987 and 1989; Subbarao and Wei 1992; Subbarao 1988 and 1989b; Subbarao and Surya 1992; Lai and Fu 1992) have proposed methods for finding distance of an object which do not require focusing the object. They take the level of defocus of the object into account in determining distance. Therefore this approach is called *Depth-from-Defocus* (DFD). DFD methods do not involve searching for f and s values which correspond to focusing the object. Therefore these methods require processing only a few images (about 2–3) as compared to a large number (about 10–12) of images in the DFF methods. In addition, only a few images are sufficient to determine the distance of all objects in a scene using the DFD methods, irrespective of whether the objects are focused or not.

Several DFD methods have been proposed and demonstrated for objects with brightness edges (Grossman 1987; Pentland 1987; Subbarao and Natarajan 1988; Subbarao 1989b and 1989c; Lai and Fu 1992). In this case the underlying focused image is assumed to be a step edge or an edge of known form. A measure of edge blur is computed and this is used to estimate the distance of the object. In estimating the object distance from the edge blur, either a formula derived theoretically or a look-up table obtained experimentally by calibration is used. DFD methods for a set of simple objects such as lines, stripes, and blobs is discussed in Subbarao (1989c).

DFD methods for arbitrary objects have been proposed by some researchers (Ens and Lawrence 1991; Pentland 1989; Subbarao and Wei 1992; Subbarao 1988). Pentland (1989) proposed a method of distance estimation for general objects using only two images, one of them being a focused image and another one a blurred image. The focused image was obtained by setting the aperture diameter to be very small (pin-hole). A ratio of the Fourier power between the two images was computed in a narrow band of frequencies. This ratio was shown to be related to the amount of blur. The ratio was used as an index into a look up table to obtain the object distance. In the experiments, over a 1 cubic meter workspace, Pentland reports an accuracy of 2.5% standard error. A very small aperture has two main problems: (i) it increases diffraction effects thus distorting the acquired image, and (ii) it increases the exposure period of the camera for a given Signal-to-Noise ratio, thus slowing the method.

Subbarao (1988, 1989b) and Subbarao and Wei (1992) have proposed a general Fourier domain based method which does not restrict the camera parameter settings or the form of the point spread function of the camera. One version of their method uses only a few one-dimensional Fourier coefficients and therefore is computationally efficient. The method was applied for autofocusing and ranging in a large number of experiments over a working space of 0.6 meter to infinity. An accuracy of 3.7% RMS error was reported in autofocusing application. In ranging application, the accuracy was about 4% RMS error at 0.6 meter distance and changed linearly to about 30% RMS error at 5.0 meter distance. The performance of this method is comparable to the STM method presented here.

Ens and Lawrence (1991) 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. They reported an RMS error of 1.3% in terms of the distance from the camera in the range 0.80 meter to 0.95 meter. The main disadvantages of the method are that it is based on a smoothness assumption and it is computation intensive.

In this paper a new method of depth estimation using a new Spatial Domain Convolution/Deconvolution Transform (S-Transform) (Subbarao 1991) is described. This method, named S-Transform Method or STM, uses only two images taken with different camera parameters. All the computations are done in the spatial domain and are local in nature. Hence this method can yield denser depth maps (due to local computations) and can be implemented in parallel. STM has been implemented on a prototype camera system named Stonybrook Passive Autofocusing and Ranging Camera System or SPARCS. A large number of experiments (about 600) using natural objects have yielded an RMS error of about 2.3% in autofocusing application. The percentage error in distance estimation is about 2.3% at 0.6 meter and increases linearly to about 20% at 5.0 meters. Considering the fact that only 2–3 blurred images are used, these results compare well with the results obtained by a DFF method of about 1.6% error at 0.6 meter and increasing linearly to about 12.5% error at 5.0 meter distance. Two variations of STM are described and implemented, one where the lens position and focal length are changed and another where the diameter of camera aperture is changed.

Next section describes the camera model. Section 3 describes the theory of S Transform relevant to STM. Subsequent sections present the theory and implementation of STM.

2 Camera Model

A schematic diagram of a camera system with variable camera parameters is shown in Figure 2. It consists of an optical system with two lenses L1 and L2. The effective focal length f is varied by moving one lens with respect to the other. O.A. is the optical axis, P1 and P2 are the principal planes, Q1 and Q2 are the principal points, ID is the image detector, D is the aperture diameter, s is the distance between the second principal plane and the image detector, u is the distance of the object from the first principal plane, and v is the distance of the focused image from the second principal plane.

The distance s, focal length f and the aperture diameter D will be referred together as *camera parameters* and denoted by $\mathbf{e}$, i.e.,

$$\mathbf{e} = (s, f, D). \tag{2}$$

In order to illustrate the theoretical basis of STM we take the optical system to be circularly symmetric around the optical axis, and use a paraxial geometric optics model (Gaskill 1978) for image formation. This is a good approximation in practice to actual image formation process modeled by physical optics (Born and Wolf 1980; Subbarao and Lu 1992). However, STM itself is applicable to physical optics model also.

In Figure 2, if the object point p is not in focus, then it gives rise to a blurred image p'' on the image detector ID. According to geometric optics, the blurred image of p has the same shape as the lens aperture but scaled by a factor. This holds irrespective of the position of p on the object plane. Since we have taken the aperture

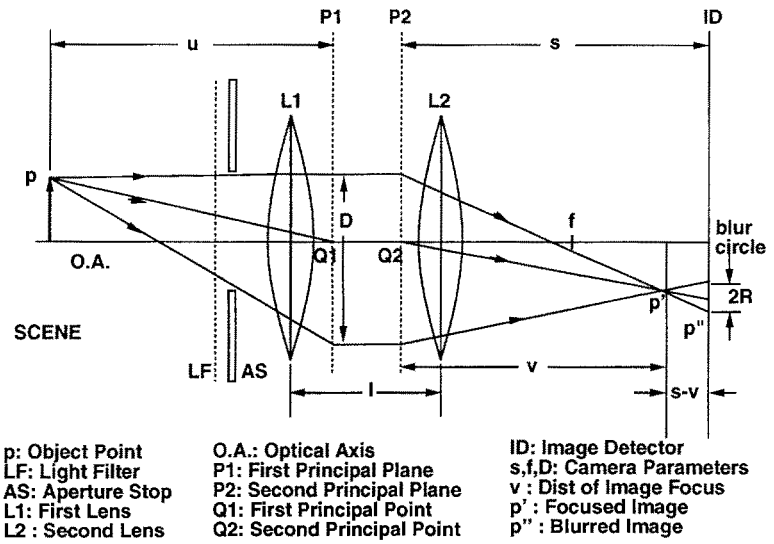


Fig. 2. Camera model and camera parameters.

to be circular, the blurred image of p is also a circle with uniform brightness inside the circle and zero outside. This is called a *blur circle.*

Let the light energy incident on the lens from the point p during one exposure period of the camera be one unit. Then, the blurred image of p is the response of the camera to a unit point source and hence it is the Point Spread Function (PSF) of the camera system. This PSF will be denoted by $h(x, y)$.

Let R be the radius of the blur circle and q be the scaling factor defined as $q = 2R/D$. In Figure 2, from similar triangles, we have

$$q = \frac{2R}{D} = \frac{s - v}{v} = s\left[\frac{1}{v} - \frac{1}{s}\right] \tag{3}$$

Substituting for $1/v$ from Eq. (1) in the above equation, we obtain

$$R = q\frac{D}{2} = s\frac{D}{2}\left[\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right] \tag{4}$$

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 a practical camera system, if two images $g_i(x, y)$ for $i = 1, 2$ are taken at camera parameter settings of $\mathbf{e}_i$, then image magnification and mean image brightness may change even though nothing has changed in the scene. For example, moving the lens away from the image detector will increase image magnification (because magnification is proportional to s) and changing the aperture diameter changes mean image brightness (which is proportional to $\pi(D/2)^2$). In order to compare the blur in images g_1 and g_2 in a correct and consistent manner, they must be first normalized with respect to these factors. Normalization with respect to image brightness is carried out by dividing the image brightness at every point by the mean brightness of the image.

Normalization with respect to image magnification is more complicated. It can be done by image interpolation and resampling such that all images correspond to the same field of view (Subbarao 1989a). The relation between an original image $g(x, y)$ taken with $s = s_1$ and the corresponding magnification normalized image $g_n(x, y)$ taken with $s = s_0$ is given by $g_n(x/s_1, y/s_1) = g(x/s_0, y/s_0)$. However, in most practical applications, the magnification change is less than 3% and can be ignored. It

is probably for this reason that most previous literature fails to mention the magnification correction. But this cannot be overlooked from a theoretical point of view.

In the following discussion we assume that the images have been normalized. Without loss of generality, we assume that the magnification has been normalized corresponding to $s = s_0$. After magnification normalization, the normalized radius $R' = s_0 R/s$ of the blur circle can be expressed as a function of the camera parameter setting $\mathbf{e}$ and object distance u as

$$R'(\mathbf{e}; u) = \frac{Ds_0}{2}\left(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}\right). \tag{5}$$

If we assume the camera to be a lossless system (i.e., no light energy is absorbed by the camera system) then

$$\iint h(x, y)dx\,dy = 1 \tag{6}$$

because the light energy incident on the lens was taken to be one unit. Using this and the fact that the blur circle has uniform brightness inside a circle of radius R' and zero outside, we obtain the PSF to be a cylindrical function:

$$h_1(x, y) = \begin{cases} \frac{1}{\pi R'^2} & \text{if } x^2 + y^2 \leq R'^2 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where h_1 is the PSF according to paraxial geometric optics.

In practice, the image of a point object is not a crisp circular patch of constant brightness as suggested by geometric optics. Instead, due to diffraction, polychromatic illumination, lens aberrations, etc., it will be a roughly circular blob with the brightness falling off gradually at the border rather than sharply. Therefore, as an alternative to the above cylindrical PSF model, often (Horn 1986; Pentland 1987; Schreiber 1986; Subbarao 1988) a two-dimensional Gaussian is suggested which is defined by

$$h_2(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{8}$$

where σ is a spread parameter corresponding to the *standard deviation* of the distribution of the PSF. In practice, it is found that (Subbarao 1989b and 1989c) σ is proportional to R', i.e.

$$\sigma = kR' \quad \text{for} \quad k > 0 \tag{9}$$

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

$$k = \frac{1}{\sqrt{2}} \tag{10}$$

is a good approximation (Subbarao and Natarajan 1988; Subbarao, 1989c and 1990). Since the blur circle radius R' is a function of $\mathbf{e}$ and u, σ can be written as $\sigma(\mathbf{e}, u)$. (However, the image of an actual point light source for our camera was quite close to a cylindrical function and was far from a Gaussian.)

If the radius R' is a constant over some region on the image plane, the camera acts as a linear shift invariant system. 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.,

$$g(x, y) = h(x, y) * f(x, y) \tag{11}$$

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. It can be defined as the square root of the second central moment of the function h. For a rotationally symmetric function it is given by

$$\sigma_h^2 = \iint (x^2 + y^2)h(x, y)\,dx\,dy \tag{12}$$

Using the polar co-ordinate system it can be shown (Subbarao and Natarajan 1988) that the spread parameter σ_{h_1} corresponding to h_1 is $R'/\sqrt{2}$. Therefore from equation (5) we have

$$\sigma_{h_1} = mu^{-1} + c \tag{13}$$

where

$$m = -\frac{Ds_0}{2\sqrt{2}} \quad \text{and} \quad c = \frac{Ds_0}{2\sqrt{2}}\left[\frac{1}{f} - \frac{1}{s}\right] \tag{14}$$

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 (5), (9) and (10) we again obtain σ_{h_2} in the same form as

$$\sigma_{h_2} = mu^{-1} + c. \tag{15}$$

In fact it has been shown (Subbarao 1989c) that even for an arbitrarily shaped aperture, σ_h is linearly related to inverse distance u^{-1}.

3 S Transform

A new Spatial-Domain Convolution/Deconvolution Transform (S Transform) has been developed for images and n-dimensional signals (Subbarao 1991) for the case of arbitrary order polynomials. The transform has been defined for both continuous signals and discrete signals. Here we summarize briefly only those results relevant to STM.

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

$$f(x, y) = \sum_{m=0}^{3} \sum_{n=0}^{3-m} a_{m,n} x^m y^n \tag{16}$$

where $a_{m,n}$ are the polynomial coefficients. (This restriction on the form of f will be relaxed later in Section 5.)

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

$$h_{m,n} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^m y^n h(x, y)\, dx\, dy \tag{17}$$

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

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x - \zeta, y - \eta) h(\zeta, \eta)\, d\zeta d\eta \tag{18}$$

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

$$\begin{aligned} f(x - \zeta, y - \eta) &= \sum_{0 \le m+n \le 3} \frac{(-\zeta)^m}{m!} \frac{(-\eta)^n}{n!} f^{m,n}(x, y) \end{aligned} \tag{19}$$

where

$$f^{m,n}(x, y) \equiv \frac{\partial^m}{\partial x^m} \frac{\partial^n}{\partial y^n} f(x, y) \tag{20}$$

$$\begin{aligned} \Rightarrow g(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{0 \le m+n \le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x, y) \\ &\quad \times \zeta^m \eta^n h(\zeta, \eta)\, d\zeta d\eta \\ &= \sum_{0 \le m+n \le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x, y) \\ &\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \zeta^m \eta^n h(\zeta, \eta)\, d\zeta d\eta \\ &= \sum_{0 \le m+n \le 3} \frac{(-1)^{m+n}}{m!n!} f^{m,n}(x, y) h_{m,n} \end{aligned} \tag{21}$$

Equation (21) 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

$$\begin{aligned} h_{0,1} &= h_{1,0} = h_{1,1} = h_{0,3} = h_{3,0} \\ &= h_{2,1} = h_{1,2} = 0 \text{ and } h_{2,0} = h_{0,2} \end{aligned} \tag{22}$$

Also from equation (6)

$$h_{0,0} = 1 \tag{23}$$

Therefore we obtain

$$f(x, y) = g(x, y) - \frac{h_{2,0}}{2} \left(f^{2,0}(x, y) + f^{0,2}(x, y) \right) \tag{24}$$

Applying $\frac{\partial^2}{\partial x^2}$ to the above expression on either side and noting that derivatives of order higher than 3 are zero (because f is cubic), we obtain

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

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

$$f^{0,2}(x, y) = g^{0,2}(x, y) \tag{26}$$

Therefore

$$f(x,y) = g(x,y) - \frac{h_{2,0}}{2}\nabla^2 g(x,y) \qquad (27)$$

where ∇^2 is the Laplacian operator. Equation (27) 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 *Inverse S-Transform* (Subbarao 1991).

Using the definitions of the moments of h and the definition of the spread parameter σ_h of h, we have $h_{2,0} = h_{0,2} = \sigma_h^2/2$, and the above deconvolution formula can be written as

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

In the following section we describe the application of this formula to the problem of distance estimation from blurred images.

4 Determining Distance

In this section we develop a theoretical basis for determining distance. Let $f(x,y)$ be the focused image of a planar object at distance u. The *focused image* $f(x,y)$ at a point (x,y) of a scene is defined as the total light energy incident on the camera aperture (entrance pupil) during one exposure period from the object point along the direction corresponding to (x,y) (Subbarao and Nikzad 1990).

Let $g_1(x,y)$ and $g_2(x,y)$ be two images of the object recorded for two different camera parameter settings $\mathbf{e}_1$ and $\mathbf{e}_2$ where

$$\mathbf{e}_1 = (s_1, f_1, D_1) \quad \text{and} \quad \mathbf{e}_2 = (s_2, f_2, D_2). \qquad (29)$$

The images g_1 and g_2 are normalized with respect to magnification, brightness, and other factors such as sensor response and vignetting as necessary (Subbarao 1989a).

For a planar object perpendicular to the optical axis, the blur circle radius R' is a constant over the image of the object (this may not be obvious at first sight, but it can be proved easily). In this case the camera acts as a linear shift invariant system. Therefore g_i will be equal to the convolution of the focused image $f(x,y)$ with the corresponding point spread function $h_i(x,y)$. In brief this can be expressed by $g_1 = h_1 * f$ and $g_2 = h_2 * f$. Let the spread parameter σ_h for h_1 be σ_1 and for h_2 be σ_2.

Now from equation (13) we can write

$$\sigma_1 = m_1 u^{-1} + c_1 \qquad (30)$$

where

$$m_1 = -\frac{D_1 s_0}{2\sqrt{2}} \quad \text{and} \quad c_1 = \frac{D_1 s_0}{2\sqrt{2}}\left[\frac{1}{f_1} - \frac{1}{s_1}\right]. \qquad (31)$$

Similarly we obtain

$$\sigma_2 = m_2 u^{-1} + c_2 \qquad (32)$$

where

$$m_2 = -\frac{D_2 s_0}{2\sqrt{2}} \quad \text{and} \quad c_2 = \frac{D_2 s_0}{2\sqrt{2}}\left[\frac{1}{f_2} - \frac{1}{s_2}\right]. \qquad (33)$$

Therefore,

$$u^{-1} = \frac{\sigma_1 - c_1}{m_1} = \frac{\sigma_2 - c_2}{m_2}. \qquad (34)$$

σ_1 can then be expressed in terms of σ_2 as

$$\sigma_1 = \alpha\sigma_2 + \beta \qquad (35)$$

where

$$\alpha = \frac{m_1}{m_2} \quad \text{and} \quad \beta = c_1 - c_2\frac{m_1}{m_2}. \qquad (36)$$

We assume that in a small image neighborhood the focused image $f(x,y)$ can be adequately approximated by a cubic polynomial in (x,y) as in equation (16). This assumption will be relaxed in the next section. In our application, the image neighborhood is of size 9×9 pixels. Now we can apply the results from the previous section, particularly the deconvolution expression (28) and obtain the following relations:

$$f = g_1 - \tfrac{1}{4}\sigma_1^2\nabla^2 g_1 \qquad (37)$$

$$f = g_2 - \tfrac{1}{4}\sigma_2^2\nabla^2 g_2 \qquad (38)$$

In the above two relations, the dependence of all functions on (x,y) is understood but has been

dropped from notation only for convenience. These two relations express the focused image f in terms of the blurred (observed) images g_1, g_2, and the spread parameters σ_1 and σ_2. Equating the right hand sides of equations (37) and (38) we obtain

$$g_1 - \tfrac{1}{4}\sigma_1^2\nabla^2 g_1 = g_2 - \tfrac{1}{4}\sigma_2^2\nabla^2 g_2. \tag{39}$$

It can be easily verified for a third degree polynomial f that $\nabla^2 g_1 = \nabla^2 g_2$. Therefore, in the above equation, $\nabla^2 g_1$ and $\nabla^2 g_2$ can be replaced by

$$\nabla^2 g = \frac{(\nabla^2 g_1 + \nabla^2 g_2)}{2} \tag{40}$$

Further, using relations (35) and (39) we obtain

$$a\sigma_2^2 + b\sigma_2 + c = 0 \tag{41}$$

where

$$a = \tfrac{1}{4}(\alpha^2 - 1)\nabla^2 g \tag{42}$$

$$b = \tfrac{1}{2}\alpha\beta\nabla^2 g \tag{43}$$

$$c = (g_2 - g_1) + \tfrac{1}{4}\beta^2\nabla^2 g \tag{44}$$

The values of α and β are determined using the camera parameter values in relations (31), (33) and (36). The Laplacians $\nabla^2 g_1$ and $\nabla^2 g_2$ are computed from the two observed images g_1 and g_2. Therefore, the coefficients a, b and c can be computed from a knowledge of the camera parameters and the observed images using relations (42), (43) and (44). Having computed the coefficients a, b and c, we can solve for σ_2 by solving the quadratic equation (41). The distance u of the object is then obtained from equation (32). The fact that the quadratic equation results in two solutions for σ_2 does pose a problem. The easiest way to overcome this two-fold ambiguity is to reduce the equation to a linear equation by forcing the coefficient a to be zero. This approach is followed in one version of our implementation. As an alternative, one may record a third image g_3 and solve for σ_2 again using g_3 in place of g_1. The common root for σ_2 in the two cases gives the correct root. However, the ambiguity persists if the two roots for the second pair of images are the same as that for the first pair of images. (This condition arises very rarely.) This is the basic principle of STM. However, the method described above needs to be modified to make it applicable in practice.

Due to noise, the two focused images derived from the two blurred images may not be the same and equation (39) may not be valid. In order to make the method robust in the presence of noise, the following modification was made. Equating the right hand sides of equations (37) and (38) we get

$$g_1 - g_2 = \tfrac{1}{4}(\sigma_1^2 - \sigma_2^2)\nabla^2 g \tag{45}$$

Squaring first and then integrating over a small region around the point (x, y) we get

$$\iint (g_1 - g_2)^2 dx\, dy = \tfrac{1}{16}(\sigma_1^2 - \sigma_2^2)^2 \iint (\nabla^2 g)^2\, dx\, dy \tag{46}$$

which can be expressed as

$$(\sigma_1^2 - \sigma_2^2)^2 = G^2 \tag{47}$$

where

$$G^2 = 16\,\frac{\iint (g_1 - g_2)^2\, dx\, dy}{\iint (\nabla^2 g)^2\, dx\, dy} \tag{48}$$

$$\Rightarrow (\sigma_1^2 - \sigma_2^2) = G' \tag{49}$$

where $G' = \pm G$. The sign of G' is ambiguous, but this ambiguity is not inherent. It was introduced by the squaring of equation (45). The ambiguity can be resolved from the given images g_1 and g_2 in one of several ways. As one example, if g_1 is more blurred than g_2 then $\sigma_1^2 > \sigma_2^2$ and therefore the sign is positive, otherwise the sign is negative. It is easy to determine which of g_1 and g_2 is more blurred. From the theory on Depth-from-Focus methods (Subbarao et al. 1992) it is well-known that the gray-level variance of an image is a good measure of the degree of focus of the image. Therefore, if v_1, v_2 are the gray-level variances of g_1, g_2 respectively, then the sign is positive if $v_1 < v_2$ and negative otherwise. Therefore

$$G' = \begin{cases} +G & \text{if } v_1 < v_2 \\ -G & \text{otherwise} \end{cases}$$

Now substituting for σ_1 in terms of σ_2 using equation (35) into equation (49) yields

$$\sigma_2^2(\alpha^2 - 1) + 2\alpha\beta\sigma_2 + \beta^2 = G' \qquad (50)$$

The above equation can be solved as a quadratic in σ_2.

In our experiments, two variations of STM, named STM1 and STM2, were implemented. In STM1 the lens position was changed in acquiring the two images g_1 and g_2. This resulted in changing the parameters s and f of the camera but the aperture diameter remained the same (i.e. $f_1 \neq f_2$ and $s_1 \neq s_2$ but $D_1 = D_2$). In this case we get $\alpha = 1.0$ and therefore the above quadratic equation in σ_2 reduces to a linear equation. Therefore we get the unique solution:

$$\sigma_2 = \frac{G' - \beta^2}{2\beta} \qquad (51)$$

In STM2, only the diameter of camera aperture was changed in acquiring the two images g_1 and g_2. All other camera parameters remained constant (i.e. $s_1 = s_2$ and $f_1 = f_2$ but $D_1 \neq D_2$). In this case we get $\beta = 0.0$ and $\alpha = D_1/D_2$. Therefore the quadratic equation in σ_2 reduces to

$$\sigma_2 = \pm\sqrt{\frac{G'}{\alpha^2 - 1}} \qquad (52)$$

In this case we get two solutions for σ_2. One way to obtain a unique solution is to set $s_2 = f_2$. In this case the sign of the right hand side above is negative. This is the approach used in our implementation.

Ideally it should be possible to compute the value of σ_2 at one pixel (x, y) in the image and obtain an estimate of the distance. But because of noise and digitization, it is necessary to combine information from many pixels in an image region. In our implementation several alternatives were tried. Finally the following scheme worked well: σ_2 was computed at each pixel in a neighborhood of size 48×48 and a histogram of the values was obtained. The histogram was smoothed by a Parzen window and the mode of the resulting distribution was taken to be the best estimate of σ_2. Once σ_2 is determined the object distance u can be obtained using a look-up table or calculated from equation (32).

5 Smoothed Differentiation Filters

In the previous section we assumed a local cubic polynomial model for the focused image $f(x, y)$ in deriving STM. This assumption can be removed by using a set of smoothing filters proposed by Meer and Weiss (1992) so that STM can be applied to arbitrary focused images. Meer and Weiss (1992) have proposed a set of discrete image smoothing filters for estimating images and their derivatives. These filters essentially provide an efficient way for fitting polynomials to image brightness in small neighborhoods through simple separable convolution operation. The polynomial fitting is implicit and it is done subject to least-square error minimization. In our implementation we used one version of the filters where all data points have equal weight. The filters are based on Chebyshev polynomials as described in Appendix A in Meer and Weiss (1992).

The filter for image smoothing by fitting a quadratic or cubic polynomial is

$$L_0(n) = -\frac{3[5n^2 - (3N^2 + 3N - 1)]}{(2N - 1)(2N + 1)(2N + 3)} \qquad (53)$$

where the support of the filter is $n = -N, -(N-1), \ldots, -1, 0, 1, \ldots, N-1, N$. This filter is separable and therefore can be first applied along rows and then along columns. The effective smoothing convolution kernel in this case is $L = L_0 * L_0^T$ where L_0^T is the transpose of L_0. The filter for estimating the second order image derivatives is

$$L_2(n) = -\frac{30[3n^2 - N(N + 1)]}{N(N + 1)(2N - 1)(2N + 1)(2N + 3)} \qquad (54)$$

In the implementation of STM, the result of applying the above filter along rows and columns were summed to get an estimate of the Laplacian of the image. In our implementation N was chosen to be 4 so that the window size becomes 9×9 pixels.

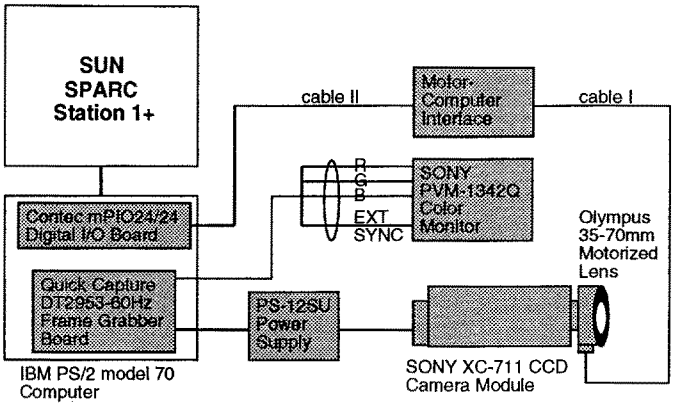


Fig. 3. Block diagram of SPARCS.

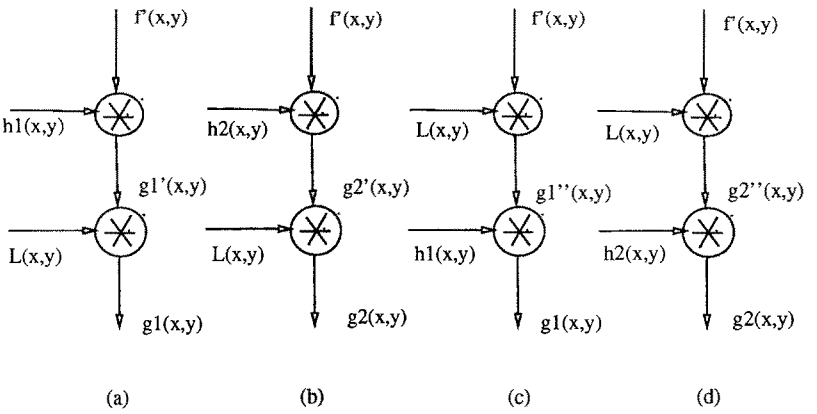


Fig. 4. Smoothing using least square filters.

Let the underlying focused image of an object be $f'(x, y)$ which is not a cubic polynomial, and let the two blurred images of f' corresponding to two point spread functions h_1 and h_2 be g_1' and g_2' respectively (Figure 4). Then we have $g_1' = h_1 * f'$ and $g_2' = h_2 * f'$. Now consider the effect of smoothing the blurred images g_1' and g_2' using a filter such as $L = L_0 * L_0^T$ which fits a cubic polynomial. If g_1 and g_2 are the smoothed images corresponding to g_1' and g_2' respectively, then we have $g_1 = L * (h_1 * f')$ and $g_2 = L * (h_2 * f')$ (Figure 4, (a) and (b)). Using commutative and associative properties of the convolution operation, we can write $g_1 = h_1 * (L * f')$ and $g_2 = h_2 * (L * f')$. Therefore, the smoothed images g_1 and g_2 can be thought of as the blurred images of the focused image f where $f = L * f'$ (Figure 4, (c) and (d)). However, f is the result of fitting a cubic polynomial to the underlying focused image f'. Therefore g_1

Fig. 5. Actual object at step 10 and images with lens steps 10,40,70.

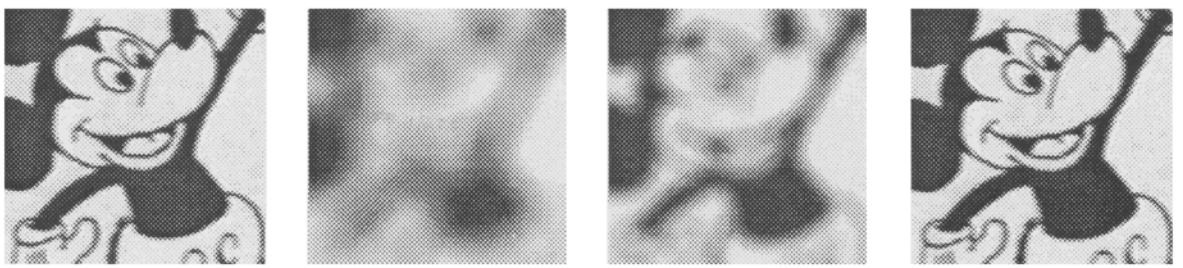


Fig. 6. Actual object at step 60 and images with lens steps 10,40,70.

and g_2 are the blurred images of a hypothetical focused image f which can be modeled by a cubic polynomial. For this reason, STM can be applied to g_1 and g_2 to estimate σ_2.

6 Implementation

6.1 SPARCS

STM described above was implemented on a camera system named Stonybrook Passive Autofocusing and Ranging Camera System (SPARCS). SPARCS was built over the last three years in our laboratory. A block diagram of the system is shown in Figure 3. SPARCS consists of a SONY XC-711 CCD camera and an Olympus 35–70 mm 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 captured images 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. 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 55 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. In effect, the system is set up such that, a C program running on the PS/2 can move the lens to any desired step number and take pictures and process them.

Table 1 shows some important data of the lens used in SPARCS. In this table, the first column specifies lens position in terms of step number of the stepper motor. The second column is the focal length f, third column is the param-

Table. 1. Lens data.

Step	FL	SI	D0
0	36.235	36.180	9034.0
5	36.086	36.132	5300.0
10	35.938	36.083	3750.0
15	35.72	36.035	2850.0
20	35.646	35.988	2500.0
25	35.502	35.941	1930.0
30	35.359	35.894	1720.0
35	35.217	35.848	1465.0
40	35.076	35.802	1320.0
45	34.937	35.757	1170.0
50	34.798	35.712	1080.0
55	34.661	35.667	965.0
60	34.524	35.622	900.0
65	34.389	35.578	822.0
70	34.255	35.534	770.0
75	34.121	35.491	715.0
80	33.989	35.448	670.0
85	33.859	35.406	628.0
90	33.728	35.363	595.0
95	33.380	35.250	560.0

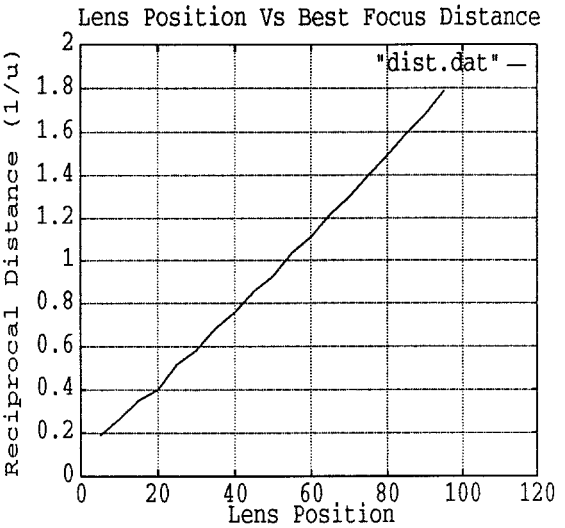


Fig. 7. Distance vs step number.

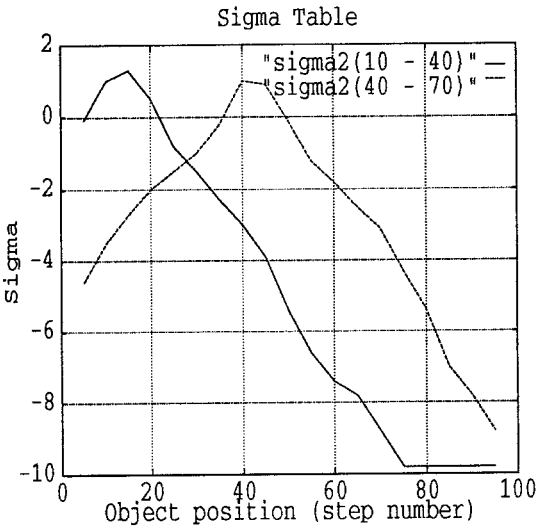


Fig. 8. Sigma vs distance.

eter s which specifies the distance between the image detector and the second principal plane of the lens, and the last column specifies the distance D_0 of an object which will be in best focus when the lens position is as specified in the first column. This data was obtained by the manufacturer of the lens by computer simulation and provided to us. This data is for the case when the zoom setting on the lens is 35 mm focal length. It is clear from this data that when the lens step number is changed, not only the parameter s but also the focal length changes by a small amount. Figure 7 shows a plot of the lens step number (the first column) along the x-axis and the reciprocal of best focused distance $1/D_0$ along the y-axis. This plot indicates that the lens step number and the reciprocal of best focused distance have an almost linear relationship. This is in fact predicted by the lens formula (1). Based on this relationship, we often find it convenient to specify distances of objects in terms of lens step number rather than in units of length such as meter. For example, when the "distance" of an object is specified as step number n, it means that the object is at such a distance D_0 that it would be in best focus when the lens is moved to step number n. The precise relationship between n and D_0 is given by Figure 7. In SPARCS, some experiments based on Depth-from-Focus methods indicated

that the data of Table 1 should be shifted by 12 lens steps, i.e. a value of 12 should be added to each of the entries in the first column. We believe that this is due to mechanical assembly error between the lens and the CCD camera. We have taken this fact into account in reporting the results of our experiments in the following discussion.

A missing piece of information in Table 1 is the dependence of the diameter of camera aperture on lens position. We believe that the diameter also changes by a small percentage when the lens moves from one end to the other, but this data is not available to us. Therefore we have taken the diameter to be constant. The diameter was calculated from the F-number and the zoom focal length.

As mentioned before, two versions of STM were implemented. In the first version named STM1, only the lens position was changed in obtaining two images g_1 and g_2, but the diameter of the lens aperture was not changed. Changing the lens position changes the parameters s and f as shown in Table 1. In the second version of STM named STM2, only the diameter of the lens aperture was changed but the other camera parameters s, f were unchanged in obtaining the two images. First we present the results for STM1 and then for STM2.

6.2 *STM1*

The overall operation of SPARCS for finding distance and autofocusing of an object is summarized as a flow chart in Figure 17. The stepwise operation is also explained briefly with comments below. In the experiments, initially, the zoom setting of the lens was set to be 35 mm focal length and the F-number was set to be 4. The camera gain was set to +6 db.

The lens is first moved to step 10 and a first image $g_1(x, y)$ is obtained. Optionally we can specify the number of image frames (typically 4) to be recorded which are then averaged to reduce noise. Such frame averaging is particularly needed under low illuminations, and in the presence of flickering illumination such as fluorescent lamps. This was clearly evident from a number of tests on SPARCS.

The lens is then moved to step 40 and a second image $g_2(x, y)$ is recorded. Again several frames may be recorded and averaged. The object to be ranged/focused can be selected by specifying a region in the image. The default region is the center of the image. The size of the region is also an option and the default size is 72×72. The two images are then normalized with respect to brightness. This is done by dividing the grey level of each pixel by the mean grey level of the entire image. Our implementation does not normalize the images with respect to other types of distortions such as vignetting and sensor response characteristics, as their effects are not significant for our camera. As mentioned earlier we have also ignored the magnification normalization, as the change in magnification due to change in lens position was found to be negligible (about 2%).

The images are then smoothed using the least-squares polynomial fit filters proposed by Meer and Weiss (1992). The filter coefficients are derived from equation (53) and the filter size is 9×9. The Laplacian of the two smoothed images are then obtained using the differentiation filters of Meer and Weiss (1992) given by equation (54).

The sign of G' is found by computing the gray-level variances of the original (unsmoothed) images g_1 and g_2. G^2 is calculated at every pixel by integrating over a 9×9 window centered at the pixel. G' is then calculated at every pixel. The value of the camera constants α and β are calculated from a knowledge of the camera parameters (see Table 1). An estimate of σ_2 is then obtained at every pixel using equation (51). Due to border effects of smoothing filter and integration, the estimates of σ_2 is limited to the interior 48×48 region of the original 72×72 images. A histogram of the estimated σ_2 is computed. The bin size of the histogram was 0.1 (the expected range of σ_2 was from about -10.0 to $+10.0$). The histogram was smoothed using a Parzen Window of size 5. The mode of the histogram was taken to be the best estimate of σ_2. This value is used to estimate the distance of the object. In autofocusing application, from σ_2, the lens step number which will bring the object to focus is determined. The lens is then moved

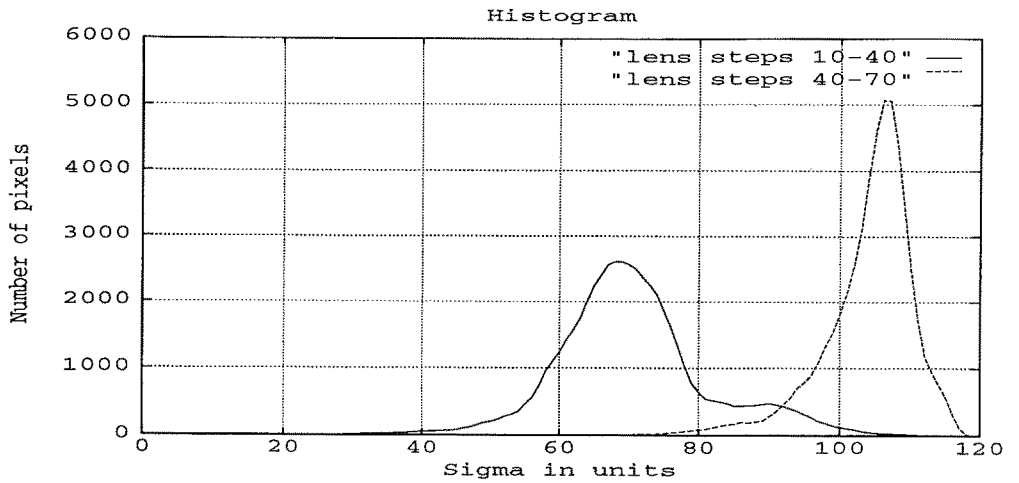


Fig. 9. Histograms obtained when an object is placed at step 10.

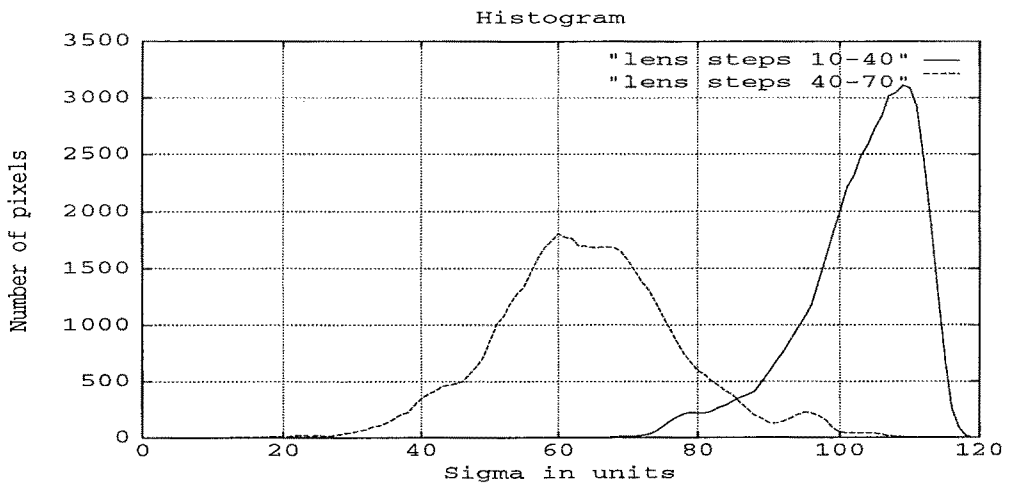


Fig. 10. Histograms obtained when an object is placed at step 60.

to this step number to accomplish autofocusing.

Once the value of σ_2 is estimated, equation (32) can be used to determine the distance of the object. However, in our implementation, for obtaining the object distance or lens step number for focusing from the computed value of σ_2, a look-up table is used. The look-up table itself is obtained through calibration and this method was found to be more accurate than the direct method of using equation (32). The calibration procedure used by us is as follows. First an object is placed at a known distance and then σ_2 is obtained exactly as described above. This procedure is repeated for several different objects at the same distance and the average σ_2 and the distance are recorded. This gives one entry of the look-up table. All other entries are obtained by repeating the above procedure for all possible distances of the objects. As for autofocusing, the relation between object distance and the best focused lens position was obtained by using a Depth-from-Focus algorithm which is based on the maximization of the energy of smoothed image gradient magni-

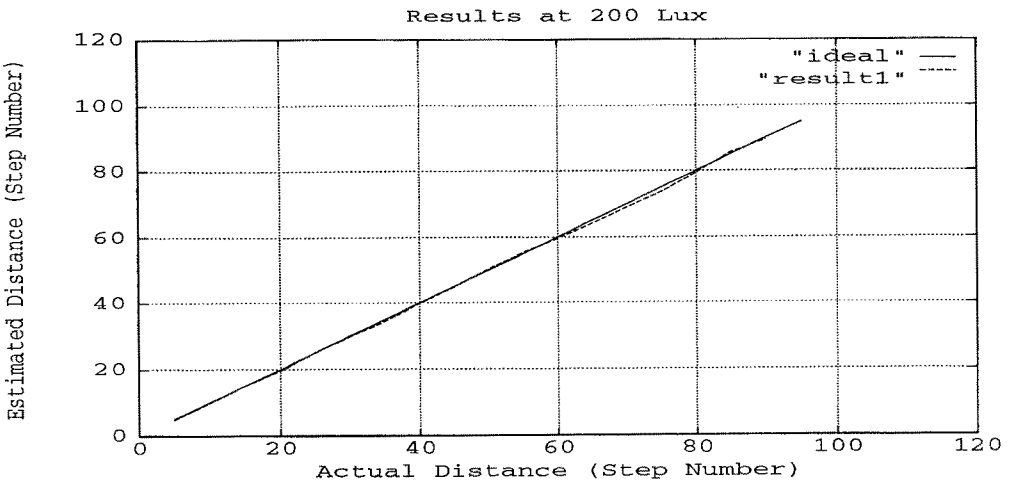


Fig. 11. Results obtained at 200 lux illumination.

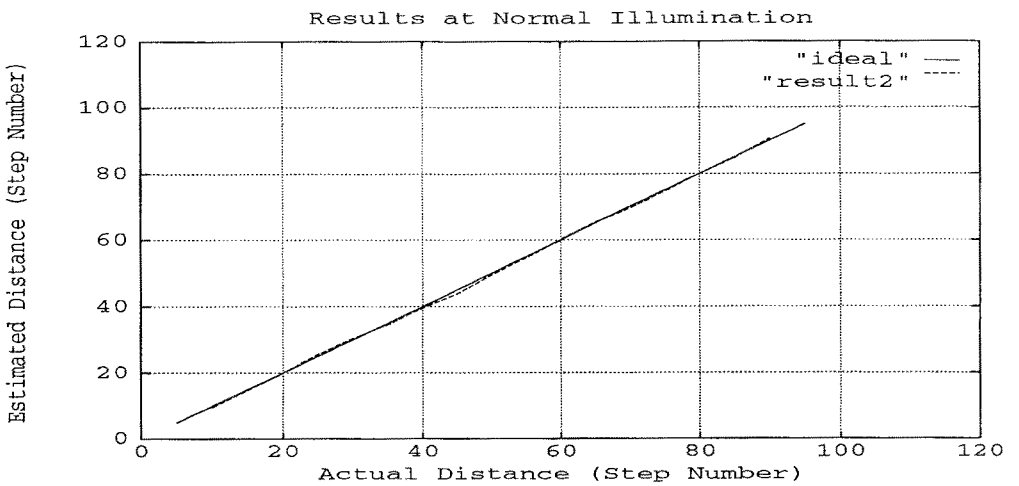


Fig. 12. Results obtained at room illumination.

tude (Subbarao et al. 1992). For every possible distance, the best focused lens step number was determined for several different objects and the mean was recorded in the look-up table.

In our experiments, it was found that if the object is very close to the camera, then the image g_1 is very highly blurred. In this case the results of ranging were unreliable. In such a case, a third image g_3 was taken after moving the lens to step number 70 and the image pair g_2 and g_3 were used in the estimation of distance. The use of this third image is due to practical reasons. Theoretically, only two images are necessary and sufficient. In our implementation, whenever the use of the first two images g_1, g_2 (taken at steps 10, 40 respectively) resulted in an object distance greater than step 45, the third image g_3 was taken (at step 70).

Two typical histograms of σ_2 are shown in Figures 9 and 10. In these figures, the histogram obtained with first pair of images (g_1, g_2) is indicated by the plot "lens steps 10–40" and the histogram obtained with the second pair of images (g_2, g_3) is indicated by the plot "lens steps

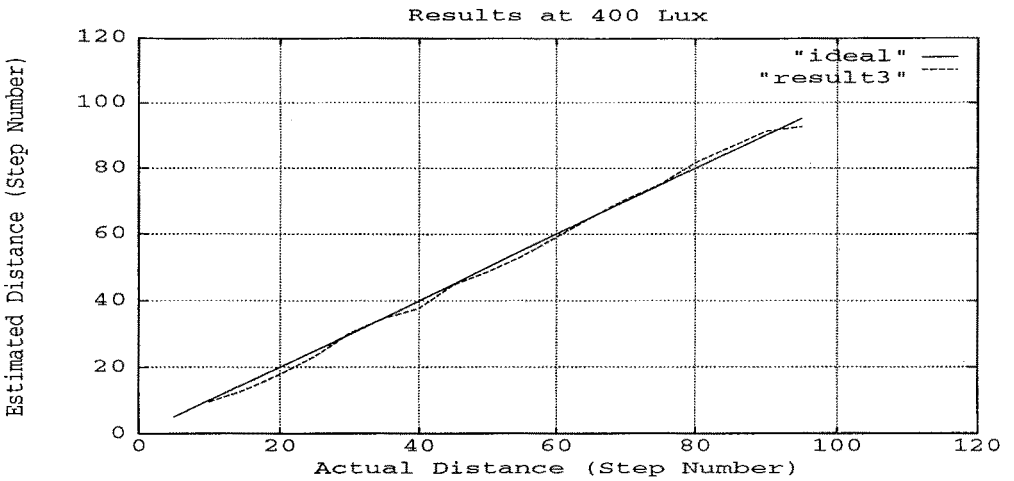


Fig. 13. Results obtained at 400 lux illumination.

40–70". In Figure 9, the object is closer to step 10 and hence the plot has a sharper peak for the first pair. In Figure 10, the object is closer to step 70 and hence the second pair of images yields a sharper histogram. In general as the object moves farther away from either of the two positions where the images are taken, the histogram becomes more and more flat. The Q-factor (ratio of peak value to width at half the peak value) of the histogram may be used as a goodness measure of the results.

Experiments were performed on five objects at normal room illumination (about 200 to 300 lux), five objects at 200 lux illumination, and 10 objects at 400 lux illumination. For each object the distance was varied from step 5 (step 10 in some cases) to step 95. All these images have been saved in an image database named SPARCS.DB1. Some of the objects in the database are shown in Figure 20. The total number of experiments is 355. The mean results are plotted in Figures 11–13. The actual distance of the object measured in step numbers is along the x axis, and the estimated distance (in step number) is along the y axis. Under ideal conditions, the plots would have been diagonals running from bottom-left to top-right, which is indicated on the plots by the "ideal" plot. Some of the objects were very difficult ones such as thin lines and edges. The Root Mean Square (RMS) error was calculated for each of the three cases. Out of 97 steps, the RMS errors were 1.48 steps at room illumination, 2.26 steps error at 200 lux illumination, and 2.28 steps at 400 lux illumination. Since there are 97 steps, this error corresponds to about 2.5 percent error.

6.2.1 Error Analysis. We shall use the accuracy achieved by Depth-from-Focus methods as a benchmark against which to compare the accuracy of STM. The DFF methods usually take a large 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 the performances of many of them are nearly the same (Subbarao et al. 1992). Since DFF methods involve exhaustive search for the focused position, we believe that the accuracy that can be obtained by any 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. A number of experiments were performed with a DFF method using the same objects used for the STM experiments. These experiments yielded an RMS error of 1.52 steps out of 97 steps. The RMS error of about 2.25 steps for STM compares well with this, considering the fact that only 2–3 blurred images are used.

The relationship between the reciprocal of

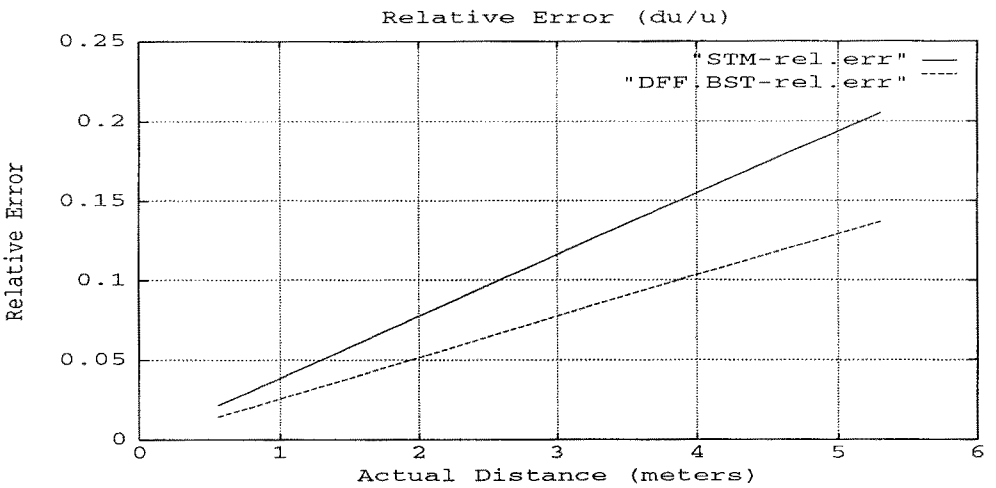


Fig. 14. Relative error.

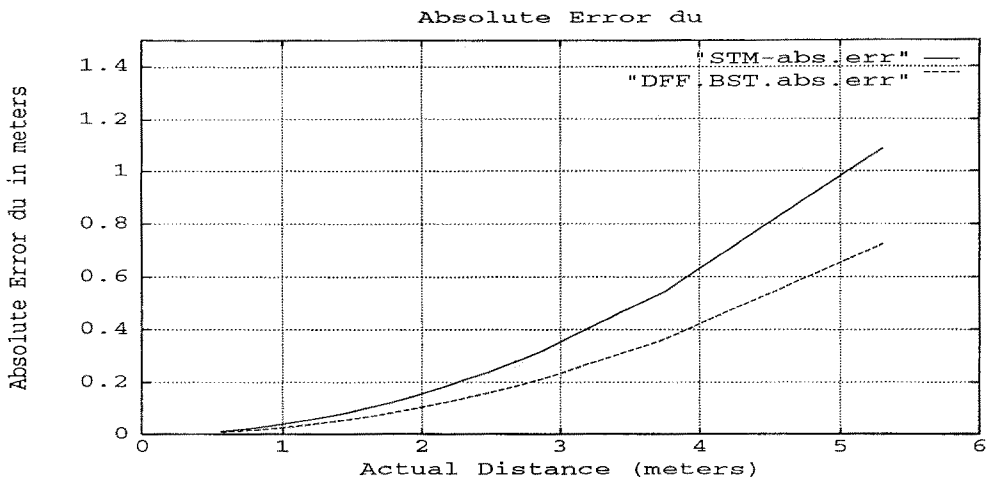


Fig. 15. Absolute error.

the object distance $1/u$ versus the lens step number is almost linear (see Figure 7) and can be expressed as

$$1/u = ax + b \tag{55}$$

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 0.03 mm. 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 (55):

$$|\delta(1/u)| = a|\delta x| \tag{56}$$

$$\Rightarrow \left|\frac{\delta u}{u}\right| = a|\delta x|u \tag{57}$$

$$\Rightarrow |\delta u| = a|\delta x|u^2 \tag{58}$$

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

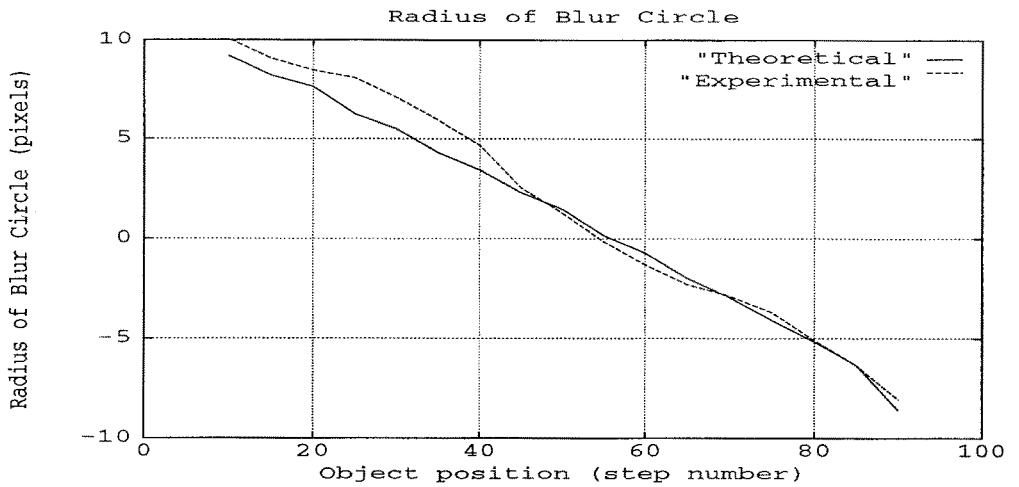


Fig. 16. Radius of blur circle with lens step 70.

error $|\delta u|$ in actual distance increases quadratically with distance. For our camera, using a Depth-from-Focus method (1992) the constants were found to be $a = 0.0172$ and $b = -0.1143$.

Setting $|\delta x|$ to be the RMS error of 1.52 steps for DFF and 2.25 steps for STM1 respectively, a plot of relative error $|\frac{\delta u}{u}|$ is shown in Figure 14 and a plot of the absolute error is shown in Figure 15. In Figure 14 we see that for STM 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 error obtained by the DFF approach of about 1.6% at 0.6 meter and increasing linearly to about 12.5% at 5 meter distance. Figure 15 shows that for STM, 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.

A comparison between the actual radius of blur circle R' obtained from an experiment (using equation (9)) and that predicted by the geometric optics model (equation (5)) is shown in Figure 16. It can be seen that the error in terms of the radius of blur circle is less than 1 pixel for most distances.

6.3 STM2

The procedure for calibration and experiments for STM2 is similar to STM1. The average sigma values are plotted in Figure 18. The experiment is first tried with two pictures g_1 and g_2 taken at lens position fixed at 0 but F-numbers 4 and 8 respectively. Fixing the lens position at step 0 assures that the focused image is always behind the image detector (because objects at infinity are focused at step 0 and all other objects come to focus at higher step numbers). Therefore a unique solution is obtained for σ_2. If the estimated distance is greater than step 60, then the object is assumed to be too close to the camera and two more pictures are taken at step 60. These two are then used in estimating the distance.

Experiments on STM2 were conducted on four different objects at room illumination (about 200–300 Lux) and ten different objects at 400 Lux illumination. These objects were the same ones used in STM1 experiments and all the images are available as a database. 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 mean values of the results are plotted in Figure 19. The overall RMS error is about 2.25 steps out of 97

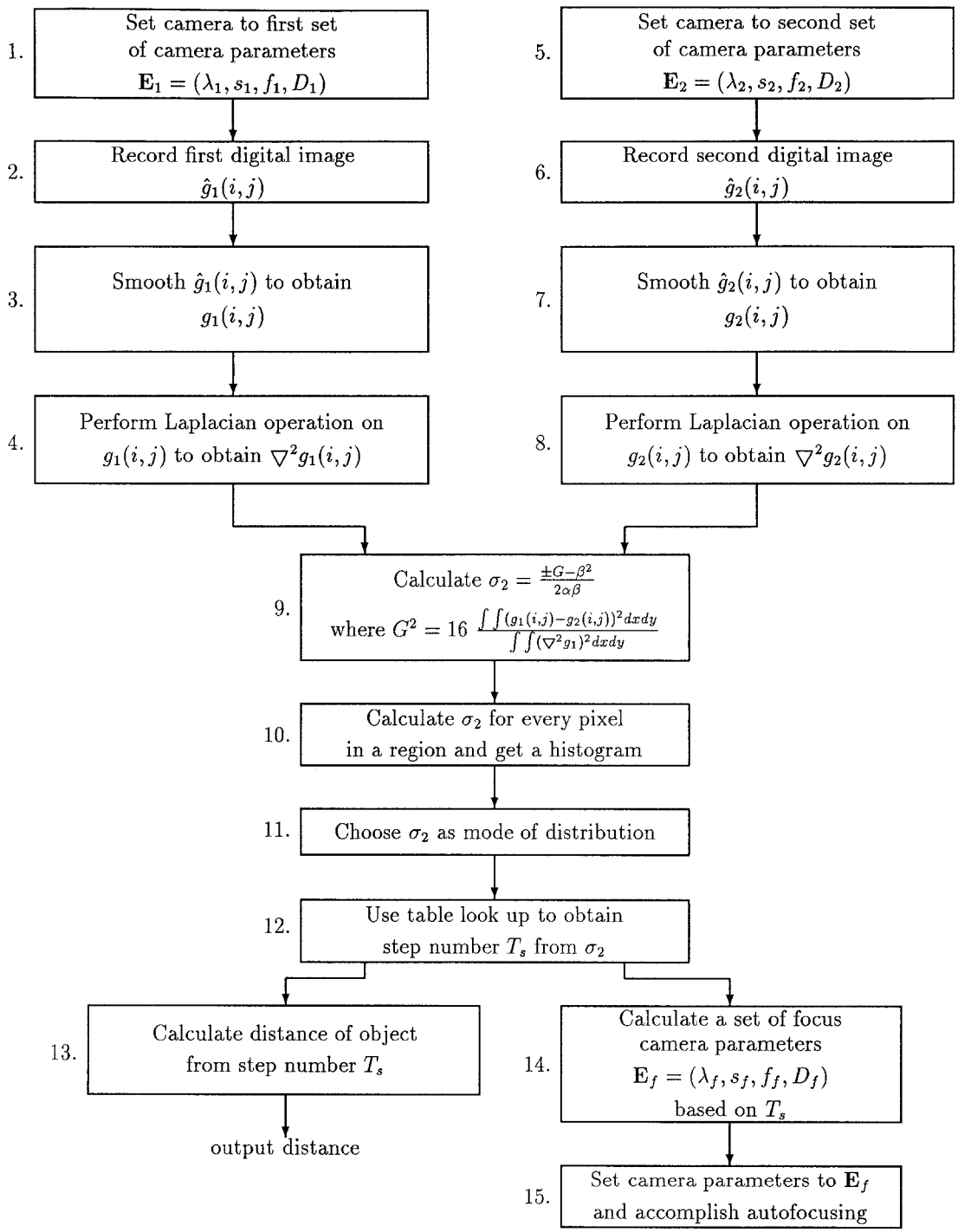


Fig. 17. Flow chart of STM.

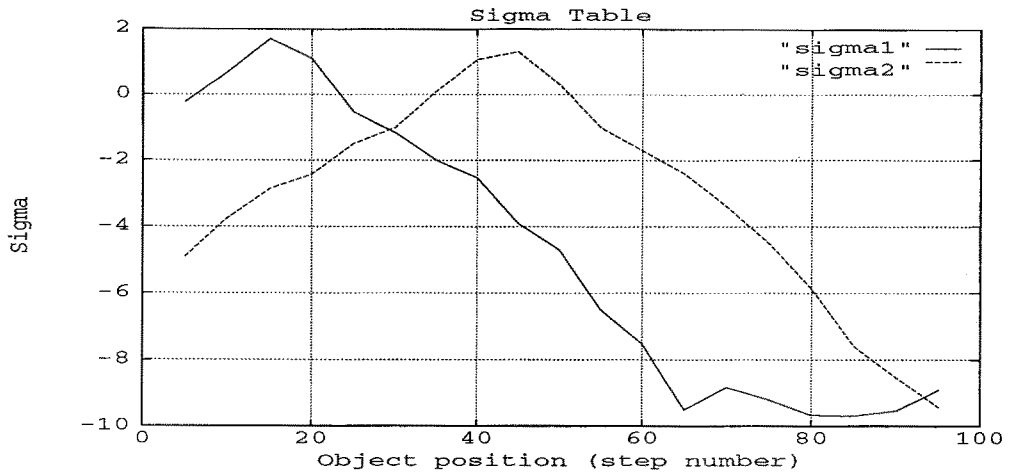


Fig. 18. Calibration plots for STM2.

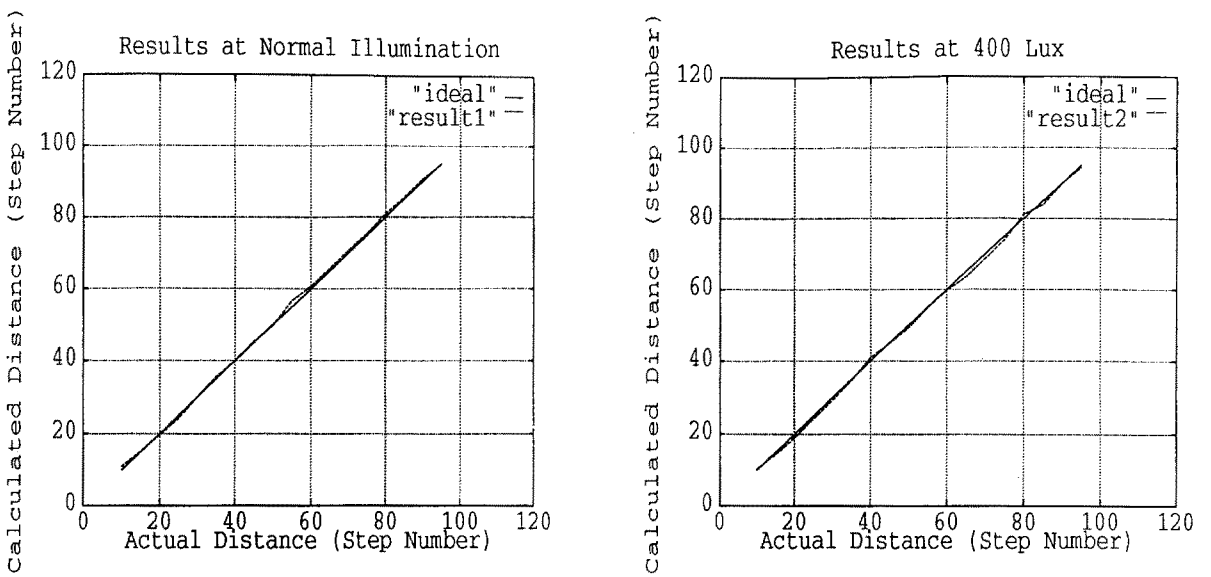
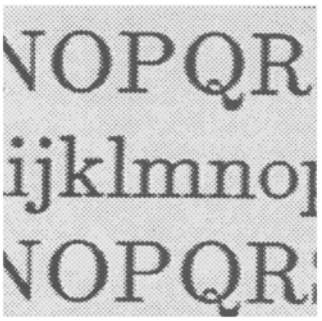


Fig. 19. Experimental results for STM2.

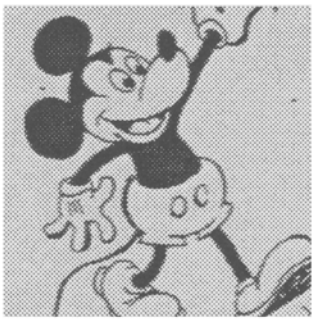
steps. This accuracy is very similar to that of STM1.

6.4 *3-Dimensional Objects*

In the previous section, planar objects were used so that a rigorous performance and error analysis could be done. Here we give the results of determining distance of some 3D objects. In Figure 21, (a) and (b) are two images of a cone taken with lens positions 40 and 70. The cone is about 1.5 meters long with black and white stripes on it. The axis of the cone is placed roughly along the optical axis of the camera and the tip is about 0.7 meter from the camera. The images were divided into overlapping regions of 32×32 pixels and STM1 algorithm was



c1

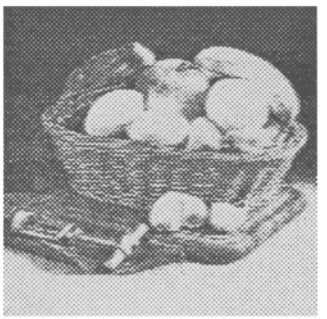


mk

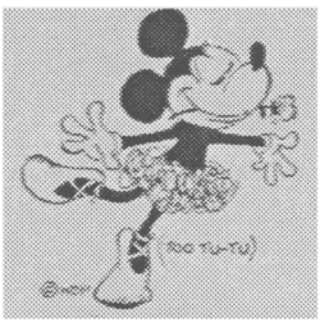
sb

fa

gs

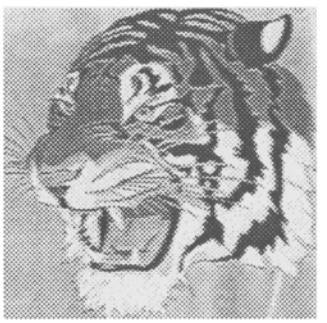


ft



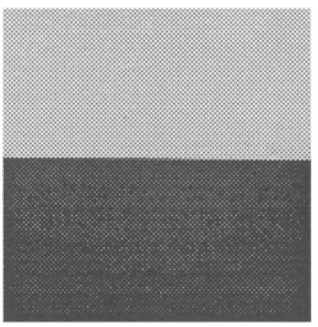
mn

gl



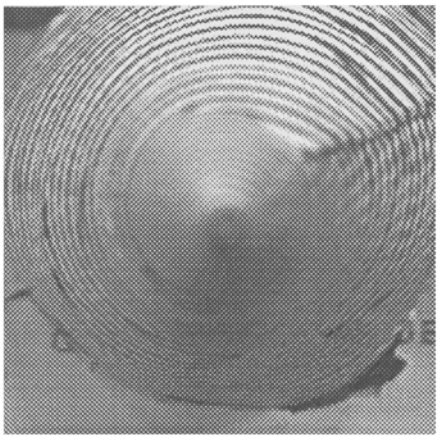
tg

c2

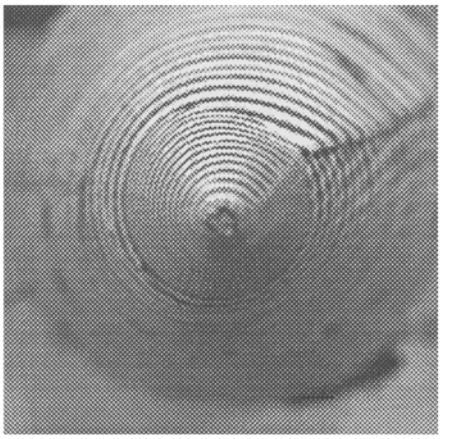


ev

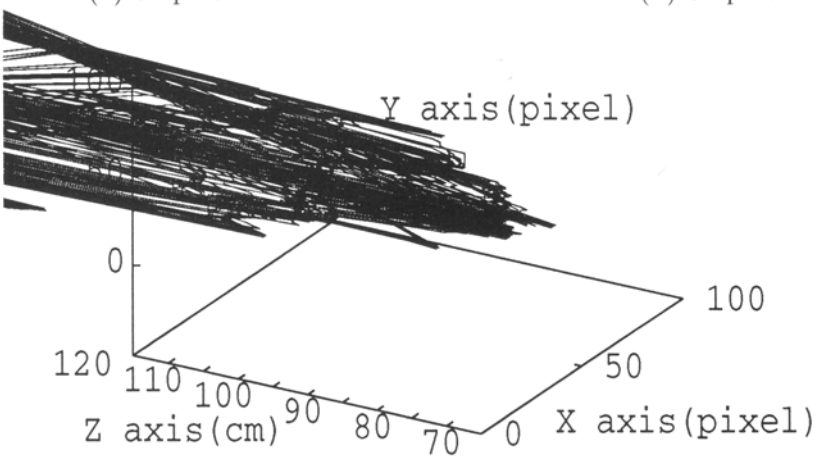
Fig. 20. Test images in the database.



(a) Step 40

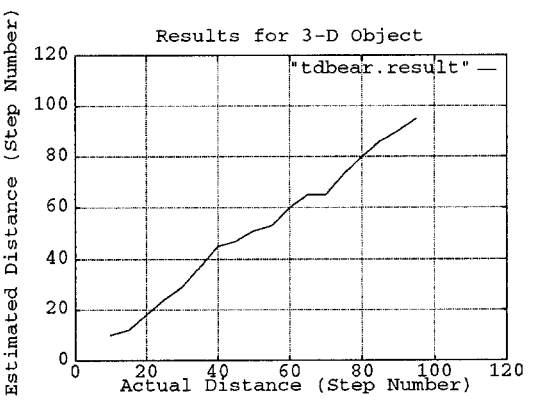


(b) Step 70



(c) Depth Map for Cone.

(d) Teddy Bear



(e) Results on Teddy Bear

Fig. 21.

used to get one depth estimate at every 4 pixel intervals. The resulting depth map is shown in Figure 21 (c). The depth-map is roughly in agreement with the ground truth.

Figure 21 (d) shows a 3-D object (Teddy Bear), which has a depth variation of about 10 cm. The result of running STM1 on it is shown in Figure 21 (e). The face of the Teddy Bear was the region chosen for processing. We get an approximate depth estimate in such cases where the depth discontinuities are not too large.

In the case of 3-D objects, blurred images cannot be modeled as the result of convolving the focused image with the PSF of the camera. Therefore the distance estimated by STM will be in error. The error depends on the shape and appearance of objects. For objects with small depth variations, STM gives an estimate of "average" distance of the objects in the image window being processed.

7 Conclusions

The theory and implementation of a new DFD method named STM has been presented. It has been successfully demonstrated on an actual camera system built by us. Experimental results indicate that STM is useful for passive ranging and rapid autofocusing. The ranging accuracy of STM is high for nearby objects and decreases with increasing distance. The ranging accuracy of this method can be improved somewhat by using a DFF method which searches for the best focused position in a small interval near the distance estimated by STM. This combination of DFD and DFF methods together result in a powerful new technique for ranging which is both fast and accurate.

If we want to find the distance of only one object, then one could use binary or Fibonacci search in a DFF method. The order of complexity is about the same for both binary and Fibonacci search. However if one wants to obtain a coarse depth-map of a scene (e.g. the cone object in Section 6.4), DFF requires a large uniformly spaced sequence of images whereas STM needs only two to three images. In the case of search methods there is no closed form solution for the distance of objects and that is what distinguishes Depth-from-Focus and Depth-from-Defocus.

In comparison with the stereo method of ranging, the DFD methods do not suffer from the *correspondence* problem, but they are in general less accurate than stereo vision (assuming focal length to be less than base line, as for example in human vision system). Therefore the DFD methods such as STM 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 STM. However a random illumination pattern can be projected onto such objects to make them "textured". STM can then be used.

The wavelength of light λ can be considered as another camera parameter because focal length changes with wavelength (Born and Wolf 1980). STM can be implemented by taking two pictures using two different colors. Color filters may be placed in front of the lens for the purpose.

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 can be made much more accurate by designing cameras with small depth of field for the purpose of ranging.

We are currently investigating the use of S-Transform for deblurring the blurred images.

Acknowledgments

The support of this research by the National Science Foundation and Olympus Optical Corporation is gratefully acknowledged. Mr. Tse-Chung Wei helped in the creation of the image database SPARCS.DB1. Mr. Tae Choi provided the images of the cone.

References

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

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

J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, Inc., 1968.

P. Grossman, "Depth from Focus," *Pattern Recognition Letters* 5, pp. 63–69, Jan. 1987.

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

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

J. Ens 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.

R.A. Jarvis, "A Perspective on Range Finding Techniques for Computer Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5, No. 2, pp. 122–139, March 1983.

E. Krotkov, "Focusing," *International Journal of Computer Vision*, 1, 223–237, 1987.

S. Lai and 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.

P. Meer and I. Weiss, "Smoothed Differentiation Filters for Images," *Journal of Visual Communication and Image Representation*, 3, 1, 1992.

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.

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, 1987.

A.P. Pentland, "A Simple Real-time Range Camera," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, June 1989.

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.

W.F. Schreiber, *Fundamentals of Electronic Imaging Systems*, Springer-Verlag, Section 2.5.2., 1986.

M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters," *Second International Conference on Computer Vision*, Florida, USA, pp. 149–155, December 1988.

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), 1989a.

M. Subbarao, "Efficient Depth Recovery through Inverse Optics," Editor: H. Freeman, *Machine Vision for Inspection and Measurement*, Academic Press, Boston, pp. 101–126, 1989b.

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, 1989c.

M. Subbarao, "On the Depth Information in the Point Spread Function of a Defocused Optical System," Tech. Report No. 90.02.07, Feb. 1990, Computer Vision Laboratory, Dept. of Electrical Engineering, State University of New York, Stony Brook, NY 11794-2350.

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, 1991.

M. Subbarao, T.S. Choi, and A. Nikzad, "Focusing Techniques," Vol. 1823, Proceedings of SPIE conference, OE/TECHNOLOGY'92, Boston, Nov. 1992, pp. 163–174.

M. Subbarao and M. Lu, "Computer Modeling and Simulation of Camera Defocus," Vol. 1822, *Proceedings of SPIE conference*, OE/TECHNOLOGY '92, Boston, Nov. 1992, pp. 110–120.

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.

M. Subbarao and A. Nikzad, "A Model for Image Sensing and Digitization in Machine Vision," OE/BOSTON90, SPIE conference, Boston, Nov. 1990, Vol. 1385, pp. 70–84.

M. Subbarao and G. Surya, "Application of Spatial-Domain Convolution/Deconvolution Transform for Determining Distance from Image Defocus," Vol. 1822, Proceedings of SPIE conference, OE/TECHNOLOGY'92, Boston, Nov. 1992, pp. 159–167.

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.

J.M. Tenenbaum, *Accommodation in Computer Vision*, Ph.D. Dissertation, Stanford University, Nov. 1970.