Shape from Focus

Shree K. Nayar

CMU-RI-TR-89-27

Carnegie Mellon University
The Robotics Institute
Technical Report

(4)

Shape from Focus

Shree K. Nayar

CMU-RI-TR-89-27

Department of Electrical and Computer Engineering
The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

November 1989

This research work was supported by the Production Engineering Research Laboratory, Hitachi Ltd., Yokohama, Japan.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): CMU-RI-TR-89-27	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: The Robotics Institute, Carnegie Mellon University; 6b. OFFICE SYMBOL (*If applicable*)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (*City, State, and ZIP Code*): Pittsburgh, PA 15213	7b. ADDRESS (*City, State, and ZIP Code*)
8a. NAME OF FUNDING / SPONSORING ORGANIZATION; 8b. OFFICE SYMBOL (*If applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (*City, State, and ZIP Code*)	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (*Include Security Classification*): Shape from Focus

12. PERSONAL AUTHOR(S): Shree K. Nayar

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*Year, Month, Day*)	15. PAGE COUNT
Technical	FROM _____ TO _____	November 1989	32

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (*Continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (*Continue on reverse if necessary and identify by block number*)

Rough surfaces pose a challenging shape extraction problem. Images of rough surfaces are often characterized by high frequency intensity variations, and it is difficult to perceive the shapes of these surfaces from their images. The shape-from-focus method described in this paper uses different focus levels to obtain a sequence of object images. The sum-modified Laplacian (SML) operator is developed to compute local measures of the quality of image focus. The SML operator is applied to the image sequence, and the set of focus measures obtained at each image point are used to compute local depth estimates. We present two algorithms for depth estimation. The first algorithm simply looks for the focus level that maximizes the focus measure at each point. The other algorithm models the SML focus measure variations at each point as a Gaussian distribution and uses this model to interpolate the computed focus measures to obtain more accurate depth estimates. The algorithms were implemented and tested using surfaces of different roughness and reflectance properties. We conclude with a brief discussion on how the proposed method can be applied to smooth textured and smooth non-textured surfaces.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (*Include Area Code*)	22c. OFFICE SYMBOL

Contents

".....the effort to produce accommodation (focus) usually results in some degree of convergence (depth perception), even when one eye is covered - the so-called accommodation-convergence-synkinesis. It exists even in people who never had binocular vision....."

Gerald Westheimer, The Eye, [17]

Abstract

Rough surfaces pose a challenging shape extraction problem. Images of rough surfaces are often characterized by high frequency intensity variations, and it is difficult to perceive the shapes of these surfaces from their images. The *shape-from-focus* method described in this paper uses different focus levels to obtain a sequence of object images. The sum-modified-Laplacian (SML) operator is developed to compute local measures of the quality of image focus. The SML operator is applied to the image sequence, and the set of focus measures obtained at each image point are used to compute local depth estimates. We present two algorithms for depth estimation. The first algorithm simply looks for the focus level that maximizes the focus measure at each point. The other algorithm models the SML focus measure variations at each point as a Gaussian distribution and use this model to interpolate the computed focus measures to obtain more accurate depth estimates. The algorithms were implemented and tested using surfaces of different roughness and reflectance properties. We conclude with a brief discussion on how the proposed method can be applied to smooth textured and smooth non-textured surfaces.

1 Introduction

1.1 Motivation

The advancement of three-dimensional machine vision is largely dependent on the development of efficient and reliable shape extraction methods. Shape extraction, in turn, requires a sound understanding of various surface reflection mechanisms and the image formation process. Many extraction methods, for diffuse and specular surfaces, have been developed in the past. However, the extraction problem associated with *rough* surfaces has not received sufficient attention. All surfaces encountered in practice are rough at some level detail. At that level, they exhibit high frequency spatial surface variations that are often random in nature. In many vision applications, the spatial surface variations are comparable in dimensions to the viewing area of individual picture elements of the imaging sensor. Hence, image intensities produced by such surfaces vary in an unpredictable manner from one sensor element to the next, and it is difficult to obtain dense and accurate surface shape information by using existing techniques, such as, structured light, shape-from-shading, stereopsis, etc. Therefore, a practical and reliable solution to this rather difficult extraction problem is desirable.

1.2 Background

We propose to use focus analysis to recover the shape of surfaces. Previously, focus analysis has been used to automatically focus imaging systems and obtain *sparse* depth information from the observed scene. Horn [1] proposed focusing imaging systems by using the Fourier transform and analyzing the frequency content in the image. Tenebaum [2] developed the gradient magnitude maximization method that uses the sharpness of edges to optimize focus quality. Jarvis [3] proposed the sum-modulus-difference that is computed by summing the first intensity differences between neighboring pixels along a scan-line and is used as a measure of focus quality. Schlag et.al. [4] implemented and tested various automatically focusing algorithms.

More recently, Krotkov [5][6] evaluated and compared the performance of different focus criterion functions. Krotkov also proposed a method to estimate the depth of an image area. Pentland [7] suggested the evaluation of image blur to determine the depth of image points. Grossmann [8] has proposed the estimation of depth of edge points by analyzing the blur of the edges due to defocusing. Darrell and Wohn [9] have developed a depth from focus method that obtains an image sequence by varying the focus level and uses Laplacian and Gaussian pyramids to calculate depth. Subbarao [10] suggests the change of intrinsic camera parameters to recover the depth map of a scene. Ohta et.al. [11] and Kaneda et.al. [12] have used images corresponding to different focus levels to obtain a single level of high focus quality.

1.3 Proposed Approach

In this paper, we develop a shape-from-focus method. We start by defining visibly rough surfaces that produce textured images with high frequency intensity variations. We review the image formation process and show that a defocused imaging system plays the role of a low-pass filter. The shape-from-focus method moves the unknown object with respect to the imaging system and obtains a sequence of images that correspond to different levels of object focus. The sum-modified-Laplacian (SML) focus operator is developed to measure the relative degree of focus between images. The operator is applied to the image sequence to obtain a set of focus measures at each image point. The focus measure values at each point are modeled and interpolated to obtain accurate depth estimates. Experimental results indicate that the method is capable of extracting *dense* and accurate shape information with appreciable invariance to texture strength and type.

Though the shape-from-focus method is developed with the intention of extracting the shape of rough surfaces, it can also be applied to smooth surfaces that are textured as a result of variations in surface reflectance properties. Smooth surfaces that do not have textures can also be handled by illuminating these surfaces with high resolution light patterns to produce textured images.

2 Visibly Rough Surfaces

In the study of reflection, a rough surface is defined as one whose smallest spatial variations have dimensions that are much larger than the wavelength of the incident electromagnetic wave. This is the concept of optical roughness. In this paper, we introduce the notion of *visible roughness*; a surface is considered to be rough if the dimensions of its spatial variations are comparable to the viewing area of individual elements (e.g. pixels) of the sensor (e.g. camera) used to observe the surface. The surface shown in Fig.1 is composed of a large number of facets[1]. While the surface appears to have a smoothly varying global shape, $z(x, y)$, the orientation α of individual facets may deviate substantially from the mean surface orientation in the facet vicinity. Although the facet orientations are dependent on the global shape of the surface and on the orientations of neighboring facets, they may exhibit some degree of randomness.

Now let us consider the images of rough surfaces generated by using a finite resolution sensor. The number of facets that contribute to the image irradiance at a pixel location depends on the magnification of the optics used to project the surface onto the image plane of the sensor. We define two levels of magnification; *multi-facet level* and *facet level*. At the

[1]No assumptions are made regarding the size of the facets. Hence, these facets may or may not represent the *micro-facets* defined in [13], [14].

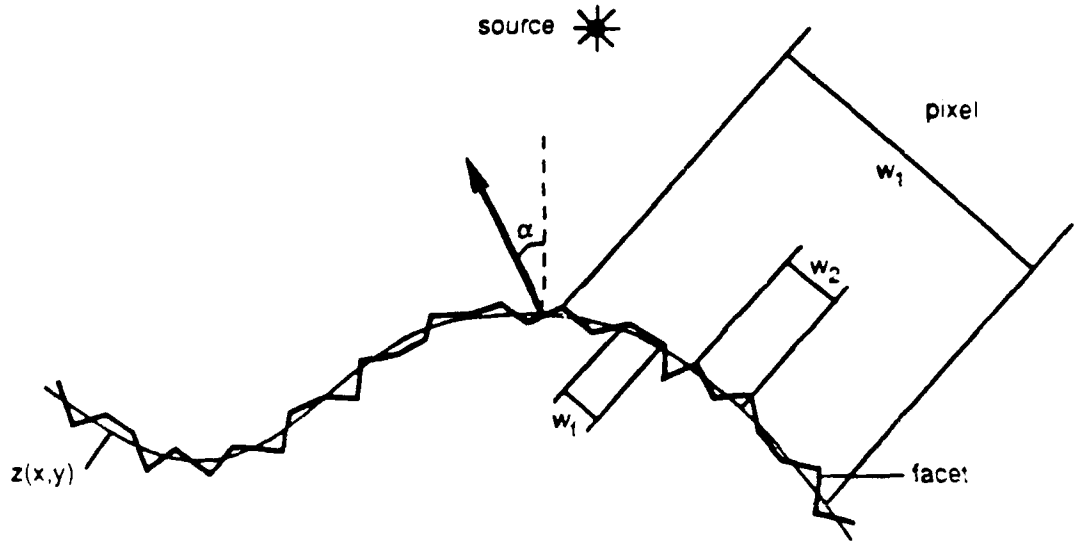


Figure 1: Rough Surface.

multi-facet level, the pixel width w is very large compared to the facet size w_f (i.e. $w = w_1$). In this case, the surface patch projected on a pixel may be modeled by assigning it a mean orientation value and a roughness which is determined by the probability function of its facet orientations [13] [14]. The pixel intensities are continuous functions of the angle of incident light and can be expressed as a linear combination of the diffuse lobe and specular lobe components [14], the relative strengths of the two components depending on the reflectance properties of the facets.

At facet level magnification, on the other hand, the pixel width w is comparable to the facet width w_f (i.e. $w = w_2$), and only one or few facets are viewed by each pixel. As a result of the random nature of facet orientations, the image intensity values are expected to vary substantially and unpredictably from one pixel to the next. This is true for both specular as well as diffuse facets as the radiance of both are dependent on the angle of incident light. Therefore, at the facet level, the surface produces images that are rich in *texture*[2] and we say that the surface is *visibly rough*. But, why do we use facet level measurements when multi-facet level measurements will provide us with image intensities that can perhaps be used to extract shape information? In many practical instances, the desirable resolution of shape information is unobtainable at the multi-facet level.

[2]There are many notions of what is meant by the term texture. Here, we define texture as a noticeable fluctuation in the intensities of neighboring image pixels [16] The textures produced by rough surfaces may be periodic, nearly periodic, or random. No assumptions are made regarding the type of texture.

3 Focused and Defocused Images

In this section, we briefly review the image formation process and describe defocused images as processed versions of focused images. There are two approaches to the study of optics, and hence, also to the analysis of the image formation process. The *physical optics* approach is based on electromagnetic wave theory [6], and it analyzes diffraction effects to derive an exact image formation model. *Geometrical optics*, on the other hand, uses the short wavelength of light to simplify the analysis, and the resulting image formation model may be viewed as an approximation to the corresponding physical optics model. Since our sensor (i.e. CCD camera) lacks sufficient spatial resolution to make diffraction effects significant, we confine ourselves to the geometrical analysis.

Fig.2 shows the basic image formation geometry. All light rays that are radiated by the object P and intercepted by the lens are refracted by the lens to converge at the point Q on the image plane. The relationship between the object distance o, focal distance of the lens f, and the image distance i, is given by the Gaussian lens law:

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}. \tag{1}$$

Each point on the object plane is projected onto a single point on the image plane, thus causing a clear or *focused* image $I_f(x,y)$ to be formed on the image plane. If, however, the sensor plane does not coincide with the image plane and is displaced from the image plane by a distance δ, the energy received from the object by the lens is uniformly distributed over a circular[3] patch on the sensor plane. Fig.2 may be used to establish the relationship between the radius r of the circular patch and the sensor displacement δ. By using similar triangles, we find that:

$$r = \frac{\delta R}{i} \tag{2}$$

where R is the radius of the lens. It is also possible to convince oneself that the radius r of the circular patch is independent of P's location on the object plane. Therefore, the blurred or *defocused* image $I_d(x, y)$ formed on the sensor plane can also be obtained by convolving the focused image $I_f(x, y)$ with a circular symmetric function $p(x, y)$ of unit volume, called the "pillbox" function:

$$I_d(x, y) = p(x, y) * I_f(x, y) \tag{3}$$

where:

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

[3] The shape of the patch also depends on the shape of the aperture of the imaging system. We are assuming the aperture to be circular.

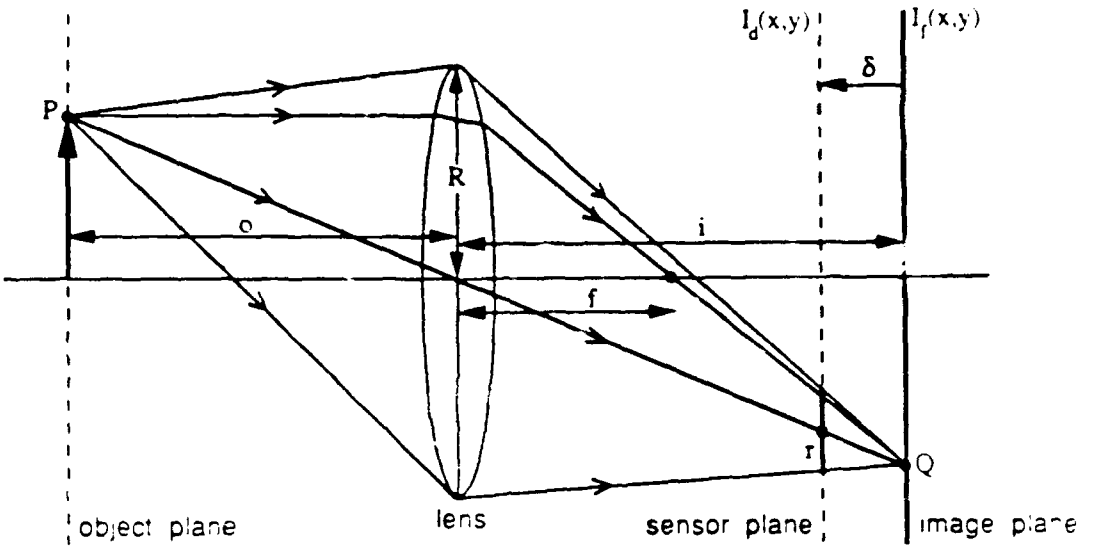


Figure 2: Formation of focused and defocused images.

Note that the defocusing effect is observed for both positive and negative sensor displacements.

Now let us analyze the defocusing process in the frequency domain (u, v). If $I_F(u, v)$, $P(u, v)$, and $I_D(u, v)$ are the Fourier transforms of $I_f(x, y)$, $p(x, y)$, and $I_d(x, y)$, respectively, we can express eq.3 as:

$$I_D(u, v) = P(u, v) . I_F(u, v) \tag{5}$$

The pillbox function $p(x, y)$ is a rotationally symmetric function, and the Fourier transform of rotationally symmetric functions can be computed as a single integral instead of a double integral by using the Hankel transform. If we define $\rho = \sqrt{u^2 + v^2}$, we can express $P(u, v)$ in polar coordinates as:

$$\overline{P}(\rho) = \frac{2J_1(r\rho)}{(r\rho)} \tag{6}$$

where $J_1(r\rho)$ is the first-order Bessel function. The quantity $J_1(r\rho)/(r\rho)$ plays the same role in two dimensions as the "sinc" function does in one dimension. $\overline{P}(\rho)$ is a rotationally symmetric function of which a one-dimensional section is shown in Fig.3. $\overline{P}(\rho)$ allows low frequencies of $I_F(u, v)$ to be passed unattenuated while it attenuates higher frequencies. Hence, we see that defocusing is a *low-pass* filtering process.

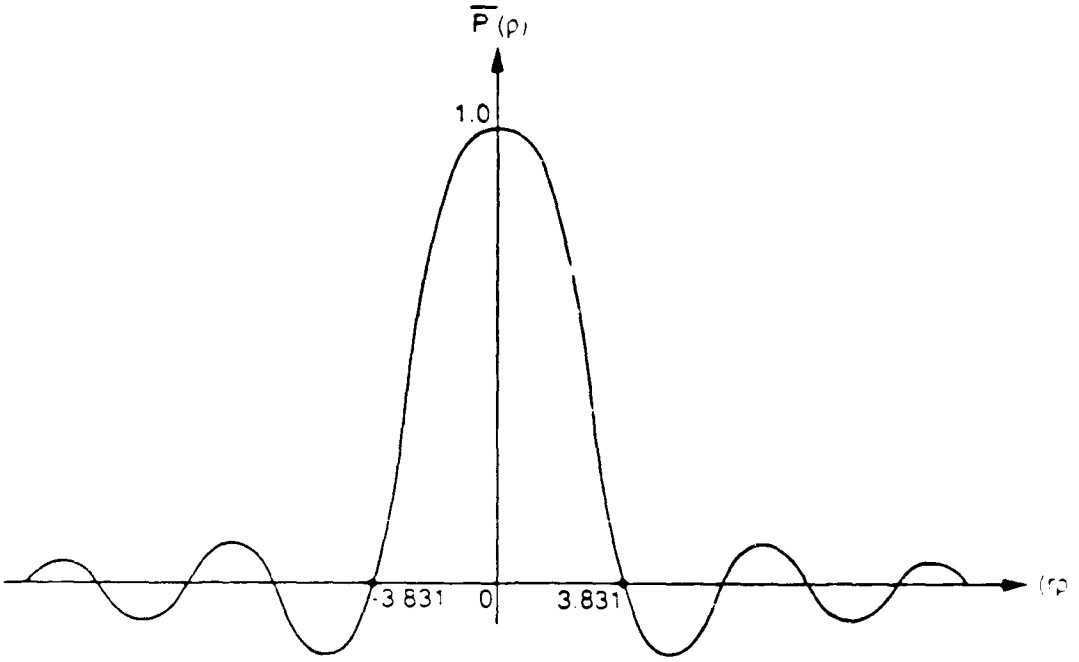


Figure 3: One-dimensional section of $\overline{P}(\rho)$.

Due to the sharp cut-off of $p(x, y)$, $\overline{P}(\rho)$ oscillates about zero and the amplitude of oscillation decreases asymptotically as $(r\rho)^{-3/2}$. The first zero of $\overline{P}(\rho)$ occurs at $r\rho = 3.83171$. Therefore, as the sensor displacement δ increases, the defocusing radius r increases, and a smaller range of low frequencies is contained in the defocused image.

From Fig.2, it can be seen that a defocused image of the object can be obtained in three ways: by displacing the sensor with respect to the image plane, by moving the lens, or by moving the object with respect to the object plane. Moving the lens or sensor plane with respect to one another causes the following problems:

- The magnification of the system varies, thereby causing the image coordinates of the object points to change.
- The area on the sensor plane over which light energy is distributed varies, thereby causing a variation in image brightness.

In order to overcome these problems, we propose to vary the degree of focus by moving the object[4] with respect to a fixed configuration of the optical system and sensor.

[4]Object movement is easily realized in industrial inspection applications.

This approach ensures that the focused areas of the image are always subjected to the same magnification.

4 Shape from Focus: An Overview

The shape-from-focus method is based on the observations made in the previous sections.

- At the facet level magnification, rough surfaces produce images that are rich in texture.
- A defocused optical system plays the role of a low-pass filter

Fig.4 shows a rough surface of unknown shape placed on a translational stage. The reference plane shown corresponds to the initial position of the stage. The configuration of the optics and sensor define a single plane, the "focused[5] plane," that is perfectly focused onto the sensor plane. The distance d_f between the focused and reference planes, and the displacement d of the stage with respect to the reference plane, are always known by measurement. Let us focus our attention on the surface element, s, that lies on the unknown surface, S. If the stage is moved towards the focused plane, the image will gradually increase in its degree of focus (high frequency content) and will be perfectly focused when s lies on the focused plane. Further movement of the element s will again increase the defocusing of its image. If we observe the image area corresponding to the s and record the stage displacement $d = \bar{d}$ at the instant of maximum focus, we can compute the height d_s of s with respect to the stage as $d_s = d_f - \bar{d}$. In fact, we can use the value of $\bar{d}$ to determine the distance of s with respect to the focused plane, sensor plane, or any other coordinate system defined with respect to the imaging system. This approach may be applied independently to all surface elements to obtain the shape of entire surface S.

To automatically detect the instant of "best" focus, we will develop an image focus measure. In the above discussion, the stage motion and image acquisition were assumed to be continuous processes. In practice, however, it is not feasible to acquire and process such a large number of images in a reasonable amount of time. Therefore, we obtain only a finite number of images; the stage is moved in increments of Δd, and an image is obtained at each stage position ($d = n.\Delta d$). By studying the behavior of the focus measure, we develop an interpolation method that uses a small number of focus measures to compute accurate depth estimates. An important feature of the method is its local nature; the depth estimate at an image point is computed only from focus measures recorded at that point. Consequently, the method can adapt well to variations in texture type and content over the object surface.

[5]The focused plane is the same as the object plane defined in the previous section. A different name is used here as the object does not necessarily lie on the focused plane.

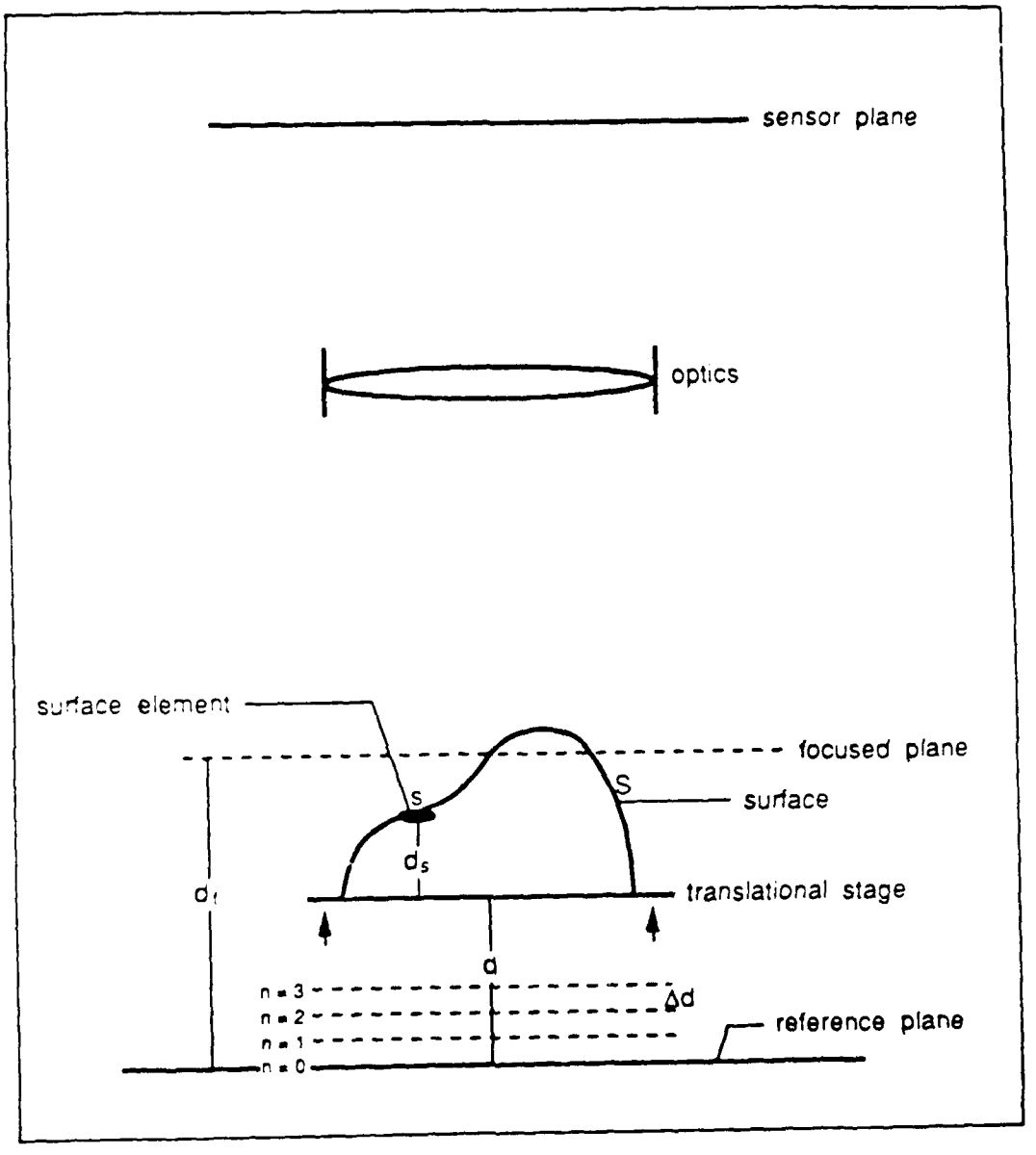


Figure 4: Shape from focus.

5 A Focus Measure Operator

In order to measure the quality of focus in a "small" image area, we would like to develop a focus measure operator. The operator must respond to high frequency variations in image intensity, and ideally, must produce maximum response when the image area is perfectly focused. The high frequency content of an image area can be determined by using the Fourier transform and analyzing the frequency distribution. However, since Fourier transforms are expensive to compute without special purpose hardware, we seek an alternative method.

A few focus measure operators have been proposed and used in the past. Generally, the objective has been to find an operator that behaves in a stable and robust manner over a variety of images such as images of outdoor scenes, text, etc. Such an approach is essential while developing automatically focusing imaging systems that have to deal with "general" scenes. Bearing in mind that we are dealing with textured images, we develop an operator that is particularly well-suited to such images. In the next section, we will evaluate our operator and compare it with other ones by using textured image samples.

One way to high-pass filter an image is to determine its second derivative. For two-dimensional images, the *Laplacian* is very often used:

$$\Delta^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \tag{7}$$

where $I(x, y)$ is the image intensity at the point (x, y). We note that in the case of the Laplacian the second derivatives in the x and y directions can have opposite signs and tend to cancel each other. An example of such an instance is illustrated in Fig.5; the partial derivatives are equal in magnitude but opposing in sign, i.e. $\Delta^2 I = 0$. In the case of textured images, this and similar instances may occur frequently, and the Laplacian is prone to behave in an unstable manner. We overcome this problem by defining the *modified Laplacian* as:

$$\Delta^2_M I = \left|\frac{\partial^2 I}{\partial x^2}\right| + \left|\frac{\partial^2 I}{\partial y^2}\right| \tag{8}$$

Note that the modified Laplacian is always greater or equal in magnitude to the Laplacian. The discrete approximation to the Laplacian is usually a 3×3 operator. In order to accommodate for possible variations in the size of texture elements, we compute the partial derivatives by using a variable spacing (*step*) between the pixels used to compute the derivatives. Hence, the discrete approximation to the modified Laplacian is computed as:

$$ML(x, y) = | 2I(x, y) - I(x - step, y) - I(x + step, y) | + \tag{9}$$

$$| 2I(x, y) - I(x, y - step) - I(x, y + step) | \tag{10}$$

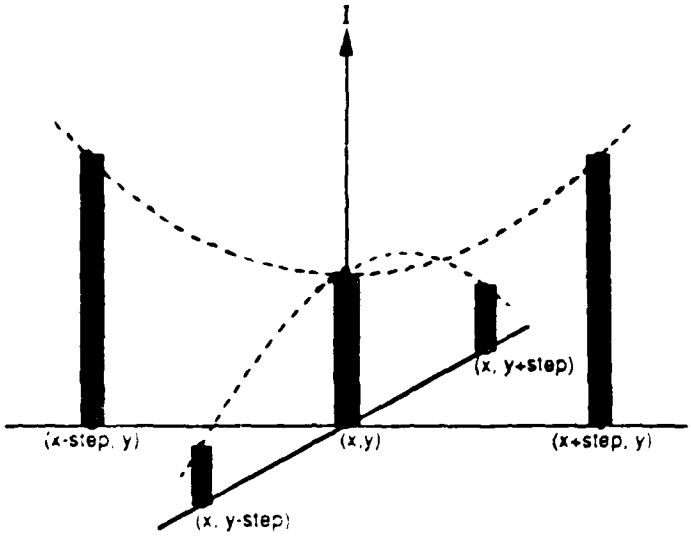


Figure 5: A texture instance with zero Laplacian value.

Finally, the focus measure at a point (i,j) is computed as the sum of the modified Laplacian, in a "small" window around (i,j), that are greater than a threshold value:

$$F(i,j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} ML(x,y) \quad for\ ML(x,y) \geq T_1 \tag{11}$$

where, the parameter N determines the window size used to compute the focus measure. In contrast to automatically focusing methods, we typically use a small window of size 3×3 or 5×5, i.e. $N = 1$ or $N = 2$. We shall refer to the above focus measure as the *sum-modified-Laplacian* (SML).

6 Evaluating the Focus Measure

We evaluate the SML focus measure in three stages. First, we analyze its behavior as a function of the distance between the observed surface and the focused plane. Next, we see how the SML measure is affected by the selection if its parameters *step* and T_1. Finally, we compare the performance of the SML operator with other operators used in the past.

A detailed description of the experimental set-up is given in a later section. In the following experiments, texture samples are attached to a translational stage (Fig.4) and the distance, d_s, from each sample to the stage is known by measurement. Images of the samples

are obtained using a microscope and a 256×256 pixel CCD camera. The complete imaging system has a physical resolution of approximately $1\mu m$ per pixel width.

In Fig.6, the focus measure functions of two samples are shown. Sample X has high texture content while sample Y has relatively weaker texture. Both samples are made of a paste containing resin and tungsten particles. The variable size of the tungsten particles gives the surfaces a randomly textured appearance. For each sample, the stage is moved in increments (Δd) of $1\mu m$, an image of the sample is obtained, and the focus measure is computed using an evaluation window size of 10×10 pixels. The vertical lines in Fig.6 indicate the known initial distances (d_f - d_s) of the samples from the focused plane. The focus measures were computed using parameter values of *step* = 1 and $T_1 = 7$. No form of temporal filtering was used to reduce the effects of image noise, as we intend to use unfiltered focus measures to estimate the depth of surface points. Though the measure values are slightly noisy, they peak very close to the expected peak positions (vertical lines in Fig.6). We see that, the focus measure function peaks sharply for the stronger texture and it peaks relatively slowly and with a lower peak value for the weaker texture. However, the sharpness of the focus measure function depends not only on the texture strength but also the "depth of focus" of the imaging system. The depth of focus, in turn, depends on the magnification and aperture size of the imaging system. We will assume that the depth of focus is constant for all our experiments.

Fig.7 shows the focus measure computed as a function of the parameter *step*, for the sample X shown in Fig.6. Once again, an evaluation window size of 10×10 pixels and a threshold value of $T_1 = 7$ were used. We see that, for sample X, a maximum measure value is computed at *step* = 4. However, it may be noted that the effective size of the focus measure evaluation window increases with the *step* size. As we are interested in local depth estimates, *step* values of 1 or 2 are generally used. Fig.8 shows the effect of varying the threshold T_1, for both focused and defocused images of sample X. A good value of T_1 is one that produces a high measure value for the focused image and low measure values for defocused images. From Fig.8 we see that, for sample X, $T_1 = 7$ appears to be a good choice. However, from a number of unreported experiments, we find that though the peak value of the focus measure function $F(d)$ varies with the parameter values, the same parameter values may be used to obtain sharp, unimodal, focus measure peaks for surface elements of varying degrees of texture content.

In Fig.10, we compare the SML focus measure with three other popular measures: the *Tenengrad*, *variance*, and *sum-Laplacian* (SL). All of these measures are defined and evaluated by Krotkov [5]. Here, we use four 10×10 sample windows to compare the measures. The differences between the four samples are illustrated in Fig.9. Sample A lies on the edge of a textured surface; only a part of the window has strong texture. Sample B lies just inside the edge of the texture. Sample C lies well within edge and sample D lies on a surface patch that is oriented away from the viewing direction of the camera. Real samples

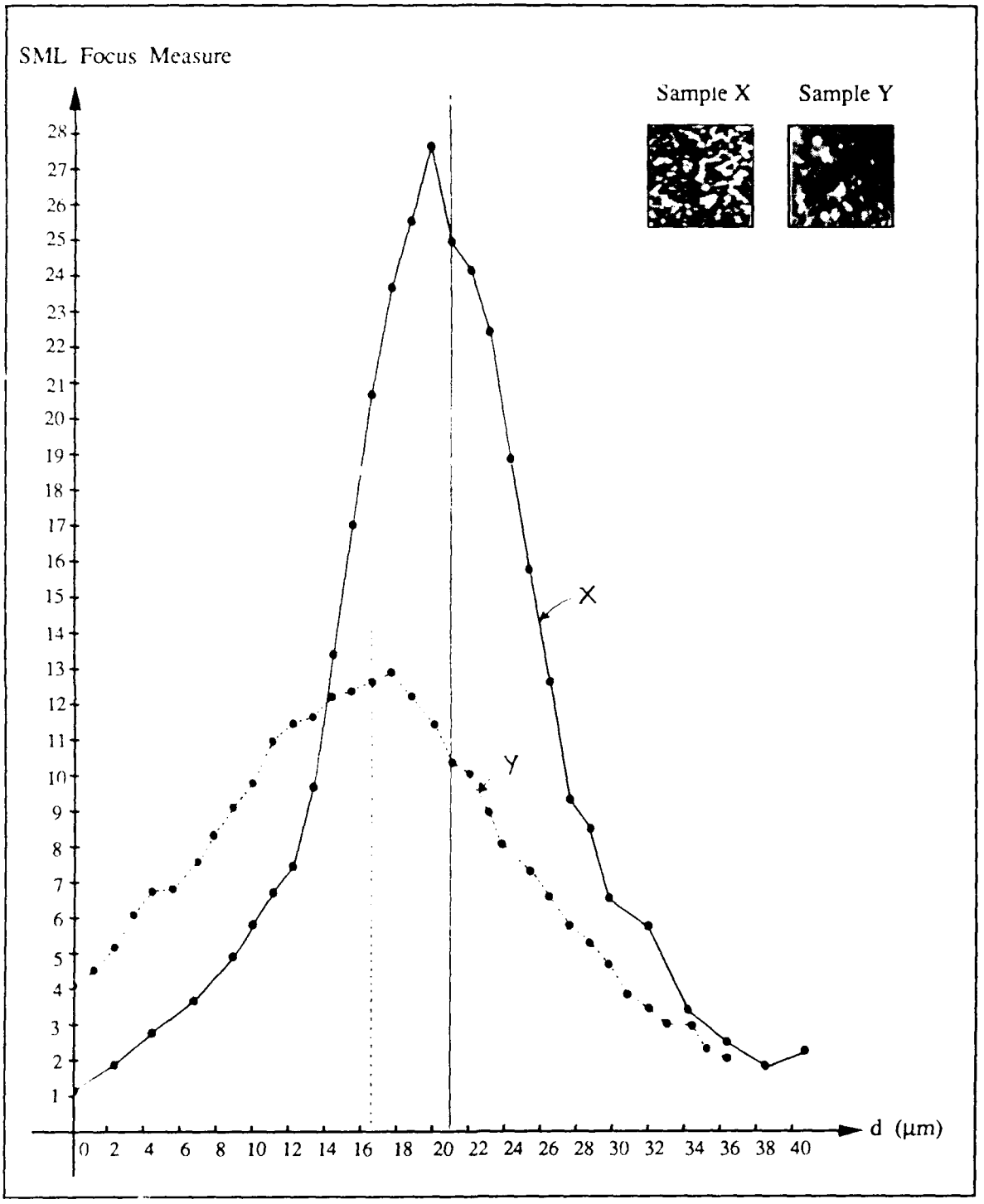


Figure 6: SML focus measure function computed for two texture samples.

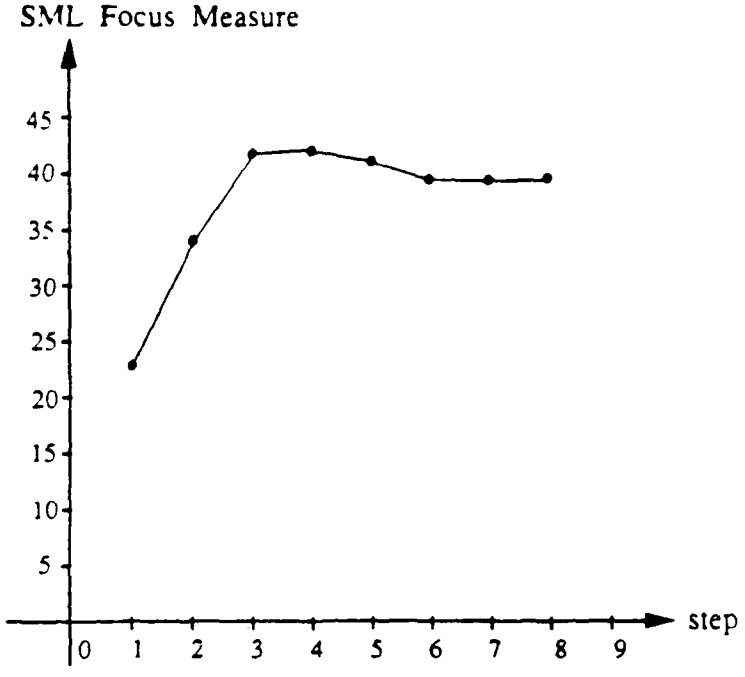


Figure 7: SML focus measure as a function of the operator parameter *step*.

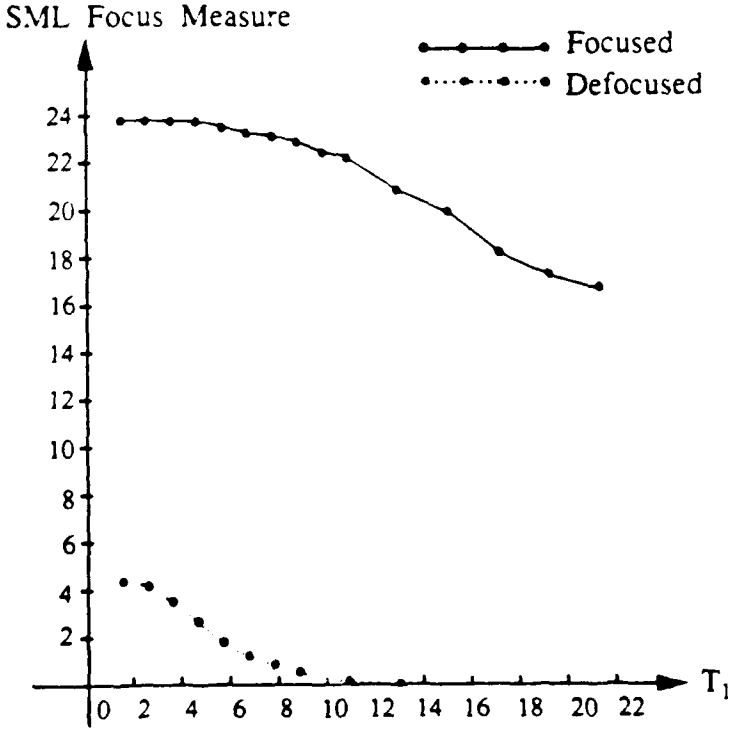


Figure 8: SML focus measure as a function of the operator parameter T_1.

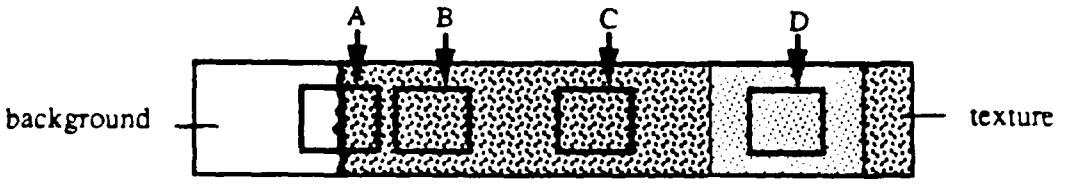
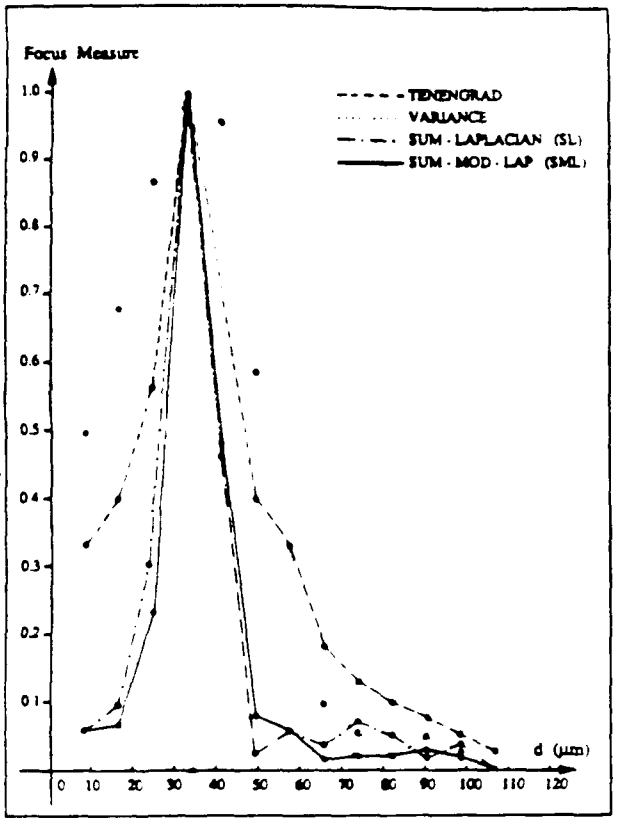


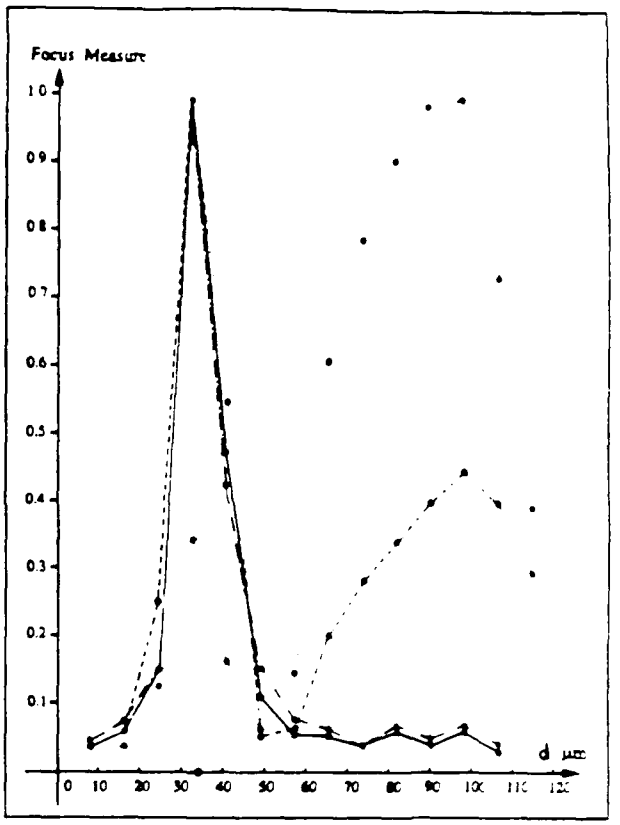
Figure 9: Texture samples used to compare focus measure operators.

and images were used to conduct the experiments. Since the comparison graphs are meant to project the global behaviors of the operators, they are plotted using large incremental displacements of the stage ($\Delta d = 10\mu m$). Therefore, the actual peaks in the focus measure function can lie between two computed measure values. The measure values produced by each operator are normalized by the maximum measure value. Once again, the vertical line in each graph indicate the known depth of each sample.

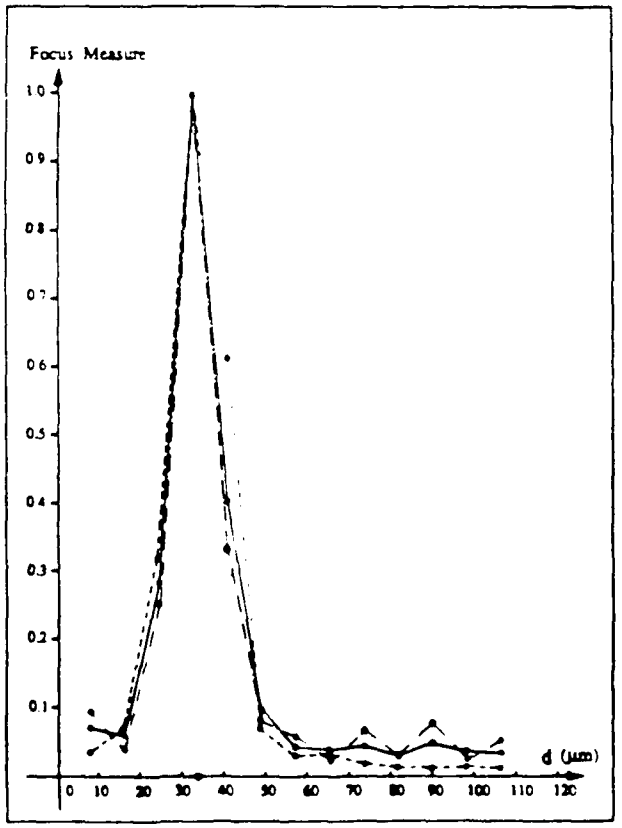
The Tenengrad was computed using a threshold value[5] of $T = 1$, and the SL and SML were computed using $T_1 = 7$ and $step = 1$. We find that the SL and SML measures peak sharply and close to the expected peak positions (vertical lines). However, for lower measure values, the SL is more noisy than the SML. For weaker textures than the ones used above, this causes multimodal measure functions. The Tenengrad and variance operators produce relatively weaker modes for sample A as both operators respond to first-order variations in image intensity. When the texture edge in sample A is defocused, the first-order variations are strong although the second-order variations are not. As a result, both these operators produce high measure values while the SL and SML operators do not. Sample B represents a particularly interesting case. When the edge of the textured surface is highly defocused, it is also subjected to a different magnification than when it is in focus. Hence, the edge coordinates change and the image of sample A shifts to the image area of sample B. Since even the highly defocused image of the edge has substantial first-order intensity variations, the Tenengrad and variance operators produce high measure values. This effect is rather significant when the mean intensities of the textured and background areas are very different. In such cases, the Tenengrad and variance operators produce bimodal focus measure functions that result in erroneous depth estimates. From these experiments, we conclude that the SML operator is best suited for measuring the focus quality of textured images.



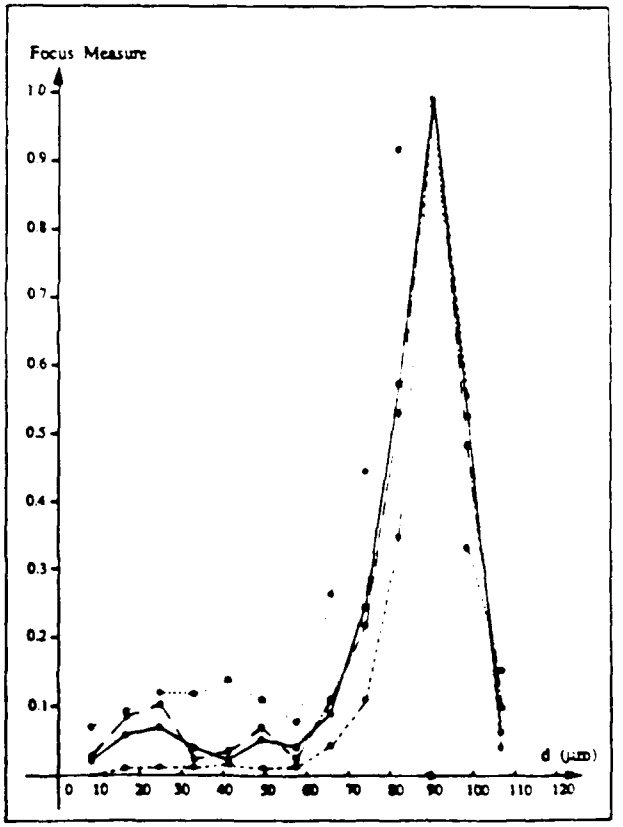
(a) Sample A



(b) Sample B



(c) Sample C



(d) Sample D

Figure 10: Comparison between Tenegrad, variance, SL, and SML operators.

7 Sampling the Focus Measure Function

We can represent the focus measure function of an image point (x, y) as $F(x,y,d)$. Since depth estimation is a local operation, we will focus our attention on a single image point, bearing in mind that the same estimation method can be applied to all other image points. Therefore, we will denote the focus measure function as simply $F(d)$. By studying the variation of $F(d)$ in Fig.6, we find that $F(d)$ may be assumed to have a Gaussian distribution with mean value $\bar{d}$ and standard deviation σ_F (Fig.11). The mean $\bar{d}$ corresponds to the stage displacement at which $F(d)$ is maximum, i.e. $F(\bar{d}) = F_{peak}$. As the texture content of the surface element increases, F_{peak} increases and σ_F decreases. Each surface element, therefore, is expected to have its own F_{peak} and σ_F values.

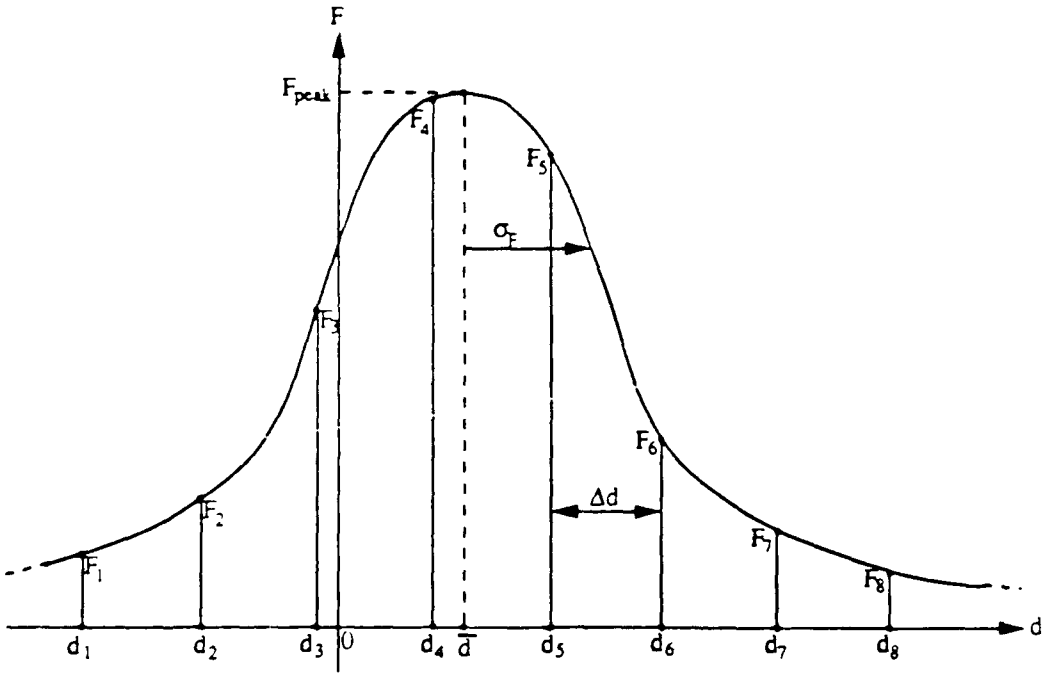


Figure 11: Sampling the SML focus measure function.

If we use very small stage displacement ($\Delta d \approx 0$), the number of images to be obtained and processed is too large from the perspective of practical implementation. Hence, we use large displacements to obtain a few images of different focus levels and use the Gaussian model to interpolate the small number of focus measures to obtain depth estimates at each image point. Computing the focus measure at a finite number of displacements is equivalent to *sampling* the function $F(d)$ (Fig.11); at each displacement d_i we compute the

focus measure $F(d_i)$ to obtain the set $\{F(d_i) \mid i = 1,2,....M\}$. We show in the following section that a minimum of *three* focus measures are needed perform the Gaussian interpolation. In theory, therefore, depth estimates may be obtained from only three images of the surface. However, since the Gaussian model only approximates the focus measure function, we use the condition $\sigma_F \leq \Delta d \leq 2\sigma_F$ to ensure that evaluation of at least one focus measure in the $\pm \sigma_F$ range of $F(d)$. Note that displacements are applied to all object points. Therefore, by applying the above condition to the image point that has maximum texture content, we can ensure that a few or many focus measures will be computed in the $\pm \sigma_F$ range at all image points.

We note that the value of σ_F also increases with the depth of focus of the imaging system. Therefore, for objects of larger dimensions also, only a small number of images may used by increasing the depth of focus.

8 Depth Estimates from Focus Measures

In this section, we describe the estimation of depth of a surface point (x, y) from the focus measure set $\{F(d_i) \mid i = 1,2,....M\}$. We use the parameter $\bar{d}$ to represent the depth of the surface point. For convenience, the notation F_i is used to represent the focus measure value $F(d_i)$. We present algorithms for two different depth estimation methods. Each algorithm may be applied to all points in the image to obtain depth maps.

8.1 Coarse Resolution Depth Estimation

The first algorithm simply looks for the displacement value d_i that maximizes the focus measure and assigns that value to $\bar{d}$.

Algorithm 1

Step 1: Let $k = 1$, $F_{max} = 0$.

Step 2: If $F_k > F_{max}$, $F_{max} = F_k$ and $\bar{d} = d_k$.

Step 3: If $k < M$, $k = k + 1$, go to step 2.

Step 4: If $F_{max} < T_2$, the point (x, y) belongs to *background*. Stop.

This simple algorithm may be used to compute rough depth estimates. The performance of the algorithm is directly dependent on the selection of Δd.

8.2 Depth Estimation by Gaussian Interpolation

The second algorithm uses the Gaussian distribution to model the focus measure function $F(d)$ and interpolates the computed measure values to obtain more accurate depth estimates. One approach is to fit all computed F_i values to the Gaussian model. However, we feel that more accurate depth estimates can be obtained, while saving computations, by using the Gaussian distribution to model only the peak of $F(d)$. The following algorithm uses only three focus measures, namely, F_{m-1}, F_m, and F_{m+1}, that lie on the largest mode of $F(d)$, such that, $F_m > F_{m-1}$ and $F_m > F_{m+1}$ (Fig.12).

Using the Gaussian model, the focus measure function may be expressed as:

$$F = F_{peak} \exp\left\{-\frac{1}{2}\left(\frac{d-\bar{d}}{\sigma_F}\right)^2\right\} \tag{12}$$

where $\bar{d}$ and σ_F are the mean and standard deviation of the Gaussian distribution (Fig. 15). Using natural logarithm, we can rewrite Eq. 12 as:

$$\ln F = \ln F_{peak} - \frac{1}{2}\left(\frac{d-\bar{d}}{\sigma_F}\right)^2 \tag{13}$$

By substituting each of the three measures F_{m-1}, F_m, and F_{m+1}, and its corresponding displacement value in Eq. 13, we obtain three equations that can be solved for $\bar{d}$ and σ_F:

$$\bar{d} = \frac{(\ln F_m - \ln F_{m+1})(d_m^2 - d_{m-1}^2) - (\ln F_m - \ln F_{m-1})(d_m^2 - d_{m+1}^2)}{2\,\Delta d\{(\ln F_m - \ln F_{m-1}) + (\ln F_m - \ln F_{m+1})\}} \tag{14}$$

$$\sigma_F^2 = -\frac{(d_m^2 - d_{m-1}^2) + (d_m^2 - d_{m+1}^2)}{2\{(\ln F_m - \ln F_{m-1}) + (\ln F_m - \ln F_{m+1})\}} \tag{15}$$

Using Eq. 12, we can find F_{peak} from σ_F and $\bar{d}$ as:

$$F_{peak} = F_m / \exp\left\{-\frac{1}{2}\left(\frac{d_m-\bar{d}}{\sigma_F}\right)^2\right\} \tag{16}$$

If F_{peak} is large and σ_F is small, the focus measure function has a "strong" peak, indicating high surface texture content in the vicinity of the image point (x, y). Thus, the values of F_{peak} and σ_F can be used to segment the observed scene into regions of different texture content.

The following algorithm first finds the measures F_{m-1}, F_m, and F_{m+1} that correspond to the strongest[6] peak of $F(d)$ and then uses these measures to estimate the depth $\bar{d}$ by Gaussian interpolation.

[6]Due to image noise and variations in magnification, the focus measure function may be multi-modal with one strong peak and one or more weak ones.

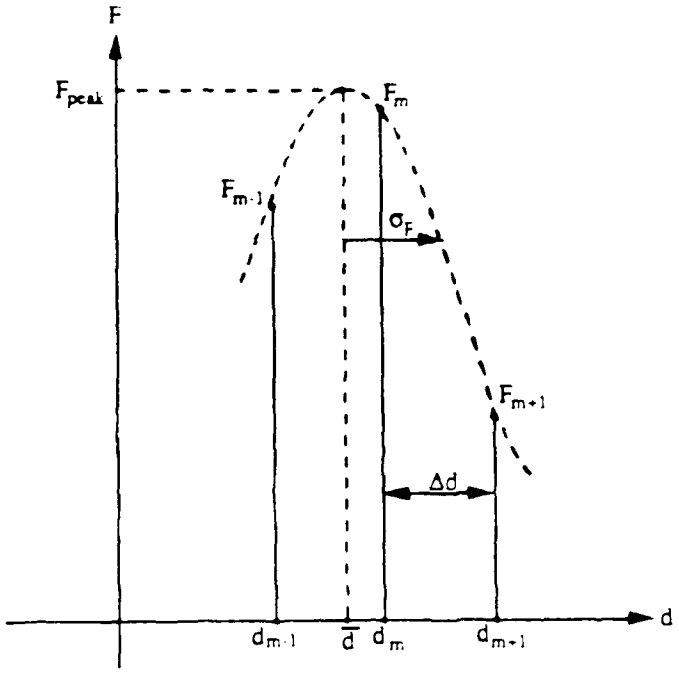


Figure 12: Gaussian interpolation of focus measures.

Algorithm 2

Step 1: Let $k = 3$, $F_{m-1} = 0$, $F_m = 0$, $F_{m+1} = 0$, $d_m = 0$.

Step 2: If $F_{k-1} > F_m$, $F_{k-1} > F_k$, and $F_{k-1} > F_{k-2}$, then:

$F_m = F_{k-1}$,
$F_{m-1} = F_{k-2}$,
$F_{m+1} = F_k$,
$d_m = d_{k-1}$.

Step 3: If $k < M$, $k = k + 1$, go to step 2.

Step 4: $d_{m-1} = d_m - \Delta d$ and $d_{m+1} = d_m + \Delta d$. Determine $\bar{d}$, σ_F, and F_{peak} using Eqs. 14, 15, and 16.

Step 5: If $F_{peak} < T_3$ or $\sigma_F > T_4$, the image point (x, y) belongs to *background*. Stop.

Since the values of F_{peak} and σ_F are only useful for texture classification, their evaluation may be avoided to save computations.

9 Experiments

9.1 Experimental Set-up

Fig.13 shows a photograph of the experimental set-up used to demonstrate the shape-from-focus method. A schematic diagram of the set-up is shown in Fig.14. A microscope is used to magnify the object surface, and images are obtained using a CCD camera with 256×256 pixels. Camera images are digitized and processes using a computer. The magnification of the imaging system can be varied from ×5 to ×160. The object is placed on a translational stage that is used to move the object through the focused plane of the imaging system. Stage displacements are monitored using an electronic displacement sensor that has an accuracy of within 0.1 μm. In most of our experiments, the bright field illumination (Fig.14) of the microscope was used to illuminate the object surface.

9.2 Results

The accuracy of the shape-from-focus algorithms was analyzed using a steel ball sample that was 1590μm in diameter. The ball was sand-papered to give it a rough surface. A camera image of the ball, under bright field illumination, is shown in Fig.15a. Due to the small depth of focus of the microscope, some areas of the ball are defocused. The bright field illumination decreases the texture intensity from the flat top section of the ball to the steep boundary area. As it is difficult to perceive the shape of most of the samples we have used from their camera images, we have also included scanning electron microscope (SEM) images of the samples. We hope that these images will provide sufficient shape cues to the reader. An SEM image of the ball is shown in Fig.15b. Incremental displacement of Δd = 100 μm were used to obtain 13 images of the bal, and a 5×5 SML operator was applied to the image sequence to obtain focus measures. Depth maps of the ball, generated by the coarse resolution and Gaussian interpolation algorithms, are shown in Fig.15c and 15d, respectively. The known size and location of the ball were used to obtain error maps by subtracting a smooth ball from the two depth maps. It is difficult to define the accuracy of the method as it depends on many factors: the surface texture, depth of focus of the imaging system, and the incremental displacement Δd. The table shown in Fig.15e shows the error statistics computed from the error maps corresponding to the two algorithms. A total of 23235 image pixels lie within the boundary of the ball. The number of depth values computes by each algorithm depends on the selected values for the thresholds T_2, T_3, and T_4. The error map corresponding to the Gaussian interpolation algorithm is shown in Fig.15f. We see that there is no obvious correlation between the errors and the surface orientation.

Fig.16 to 19 show samples with different surface reflectance and roughness properties, and their depth maps obtained using the Gaussian interpolation algorithm. All sample are

approximately 100 μm in width and an incremental displacement of Δd = 10 μm was used in all cases to obtain sequences of about 10 images each. Fig.16a and 16b show the camera and SEM images of a tungsten paste filling in a via-hole on a substrate [15] that is used to establish electrical connections between different components. Conditions such as excess filling and lack of filling cause electrical defects such as short and open circuits. The sample shown in Fig.16 has a bump on its surface, indicating excess filling. The specular reflectance and variable size of the tungsten particles gives the surface a random texture. The white background (Fig.16a) is the substrate area that has very weak texture. For this sample, we selected the threshold values to classify the substrate area as *background*. An arbitrary depth value is assigned to the background region.

Fig.17 shows another via-hole sample. In this case, the substrate and filling are hardened by baking. The baking process changes the reflectance and texture of the filling and also increases the texture content of the substrate. From the SEM image we see that the filling has some contamination on its surface and a concavity in the center. For this sample, the algorithm threshold values were selected to obtain the depth of the substrate area too. To accommodate for the large size of substrate texture elements, a *step* size of 2 was used. Two different views of the sample's depth map are shown in Fig.17c and d. Fig.18 shows another baked via-hole sample. In this case, the the via-hole is not sufficiently filled with tungsten paste. Again, two views of the computed depth maps are shown.

Fig.19 shows a via-hole filling that is plated with Nickel. The plating process increases the particale size and each particle is highly specular in reflectance. Due to the spherical shape of the particles (Fig.19b), the bright field illumination produces a single highlight at the center of each particles and some of the steep surface areas do not reflect any light. To demonstrate the use of special illumination, we used a multi-spot light source to illuminate the sample and obtained images of higher texture content, such as the one shown in Fig.19c. The computed depth map of the sample is shown in Fig.19d.

9.3 Discussion

The above experiments validate the effectiveness of the shape-from-focus method. The Gaussian interpolation algorithm performs stably over a wide range of textures. Errors in computed depth estimates result from factors, such as image noise, Gaussian approximation of the SML focus measure function, and weak textures in some image areas. Some detail of the surface roughness is lost due to the use of a finite size window to compute the SML focus measures.

Most of the above experiments were conducted using bright field illumination. In the last experiment, we demonstrated the use of special illumination to enhance the quality of textured images. Furthermore, light patterns can be projected on smooth surfaces, both

diffuse and specular,to generate textured images. Hence, by using special illumination, the shape-from-focus method can also be applied to smooth surfaces.

In the above experiments, object displacement and image aquisition were manually initiated. In a high-speed implementation, the object can be moved continuously, while images are obtained at fixed intervals of time. By using customized hardware, the SML focus operator can be applied to each image in a frame-time and the Gaussian interpolation can be implemented by using look-up tables. We estimate that a high-speed implementation of the method can generate surface depth maps in less than 1 second.

10 Conclusion

In this paper, we have presented shape-from-focus as a new method of extracting the shape of rough surfaces.

- To measure the quality of image focus, we developed the SML operator. By evaluating the SML operator and comparing it to other existing operators, we found that the SML operator is particularly well-suited for measuring the focus of textured images.
- We developed and tested two depth estimation algorithms and found, through numerous experiments, that the Gaussian interpolation algorithm produces accurate results for a variety of textures.
- The local nature of the depth estimation technique enables it to adapt to substantial variations in image texture.
- Though we have concentrated on rough surfaces in this paper, the shape-from-focus method can be directly applied to smooth textured surfaces. Smooth non-textured surfaces can also be handled by using special illumination techniques.

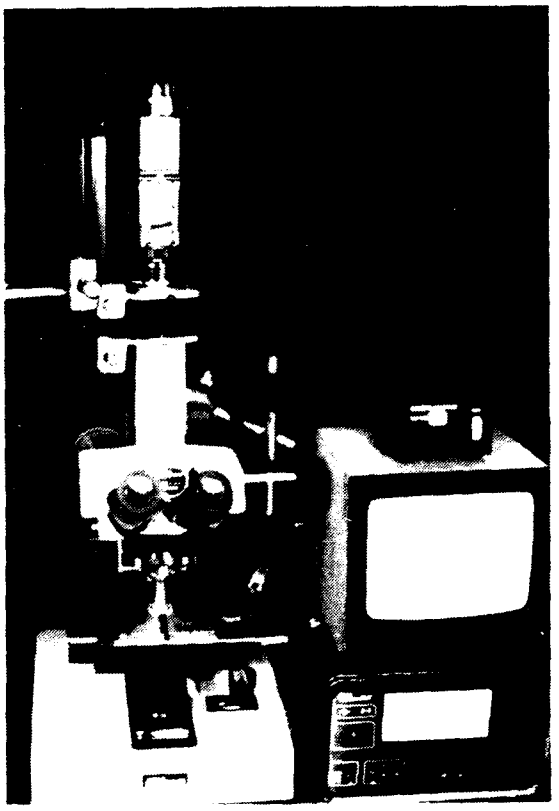


Figure 13: Photograph of the experimental set-up.

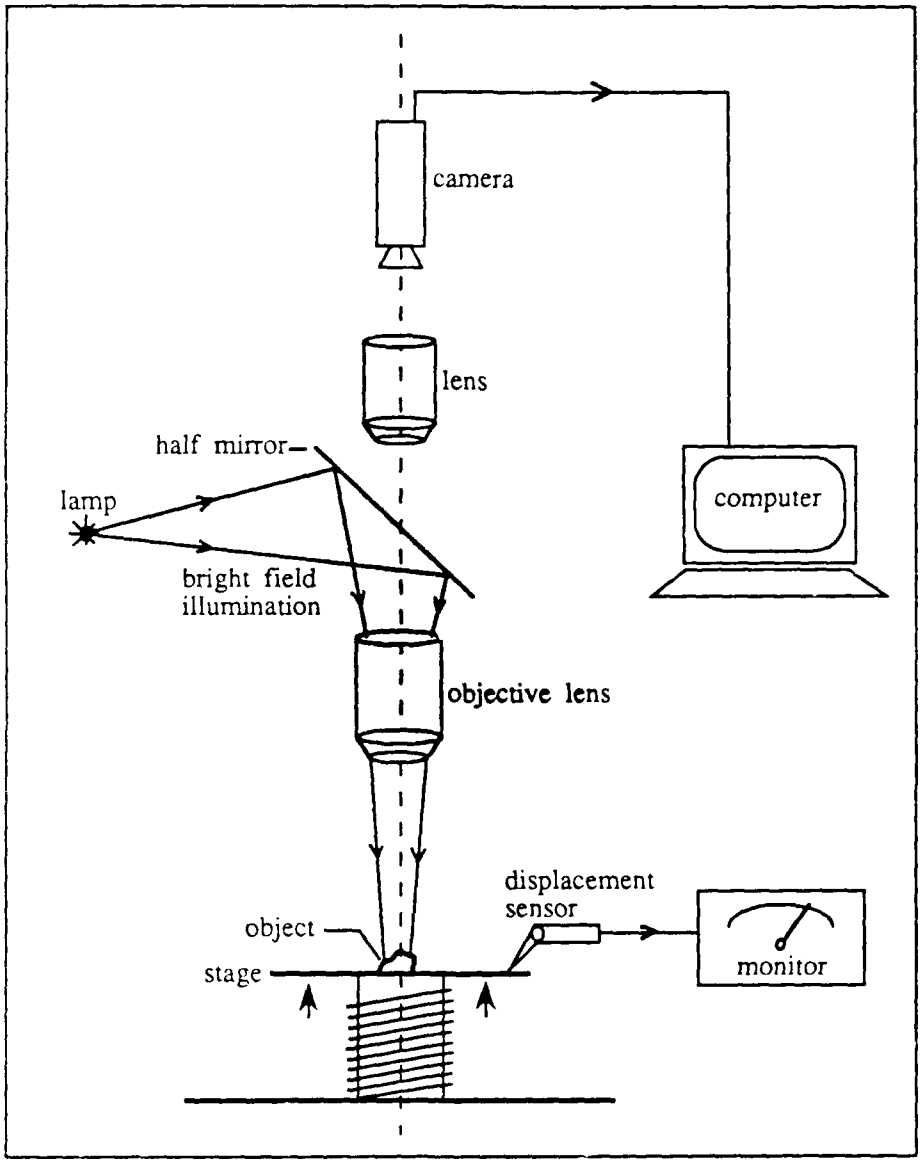
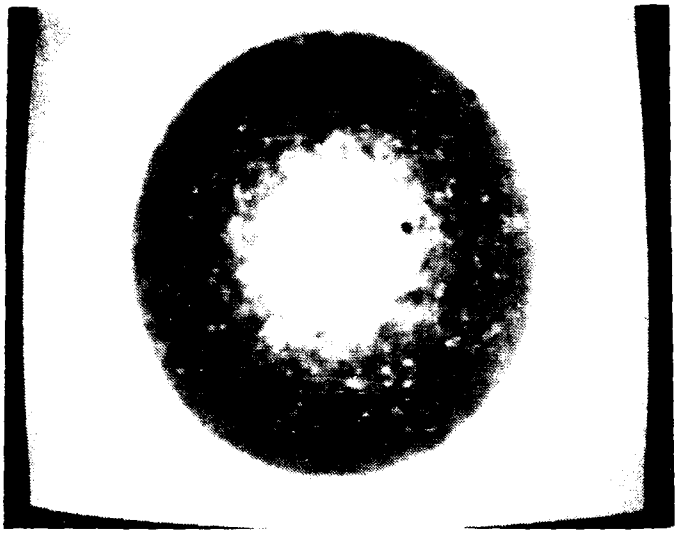
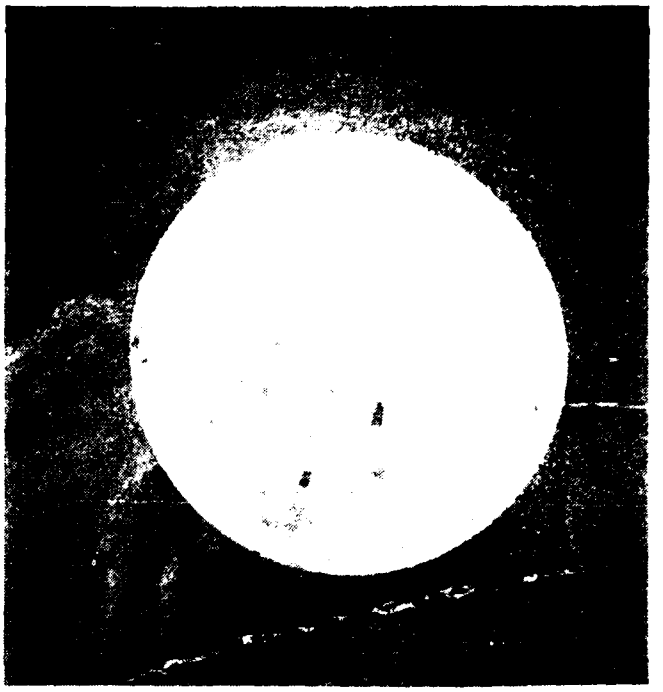


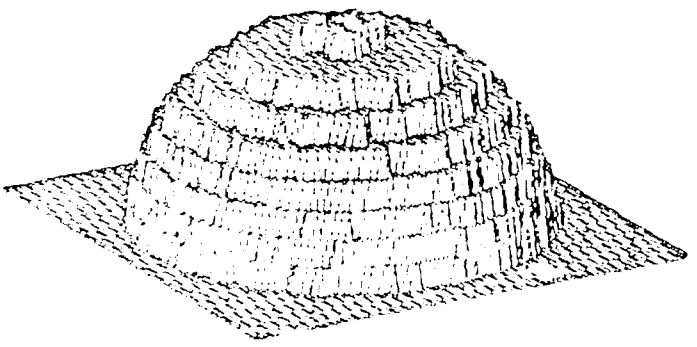
Figure 14: Schematic diagram of experimental set-up.



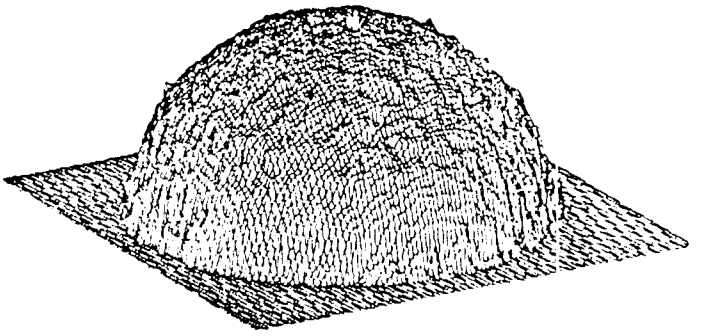
(a) Camera image.



(b) SEM image.



(c) Depth Map: coarse resolution.

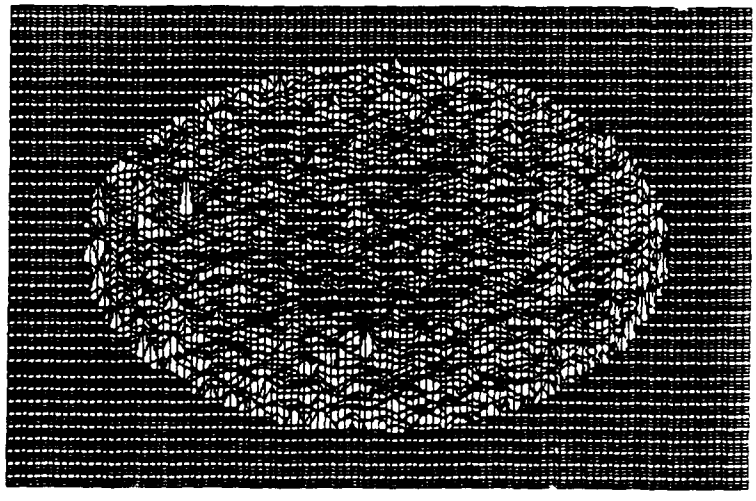


(d) Depth Map: Gaussian interpolation.

Figure 15: Steel ball.

Diameter of Test Sphere : 1590 μm		
	Coarse Interpolation	Gaussian Interpolation
Number of Points	22682	23257
Mean Error (μm)	7.861	3.857
Mean Absolute Error (μm)	30.32	13.815
Maximum Absolute Error (μm)	187.80	175.82

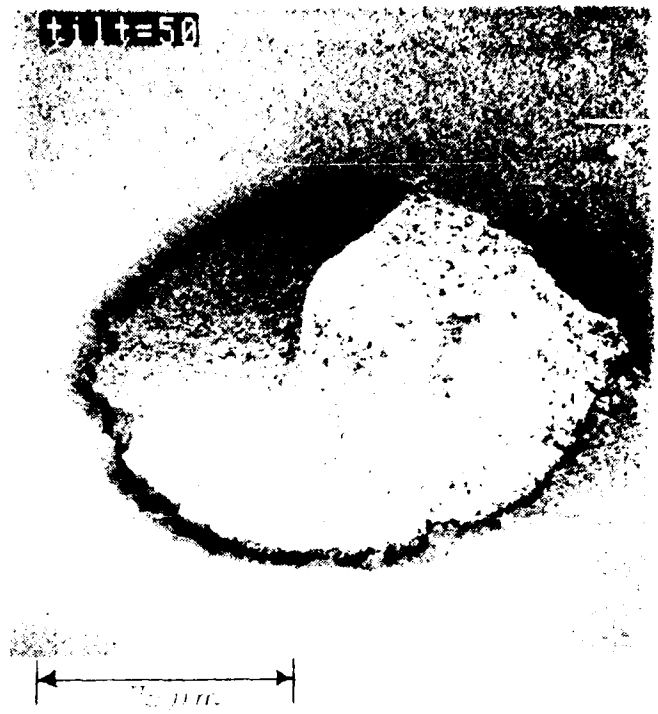
(e) Error statistics.



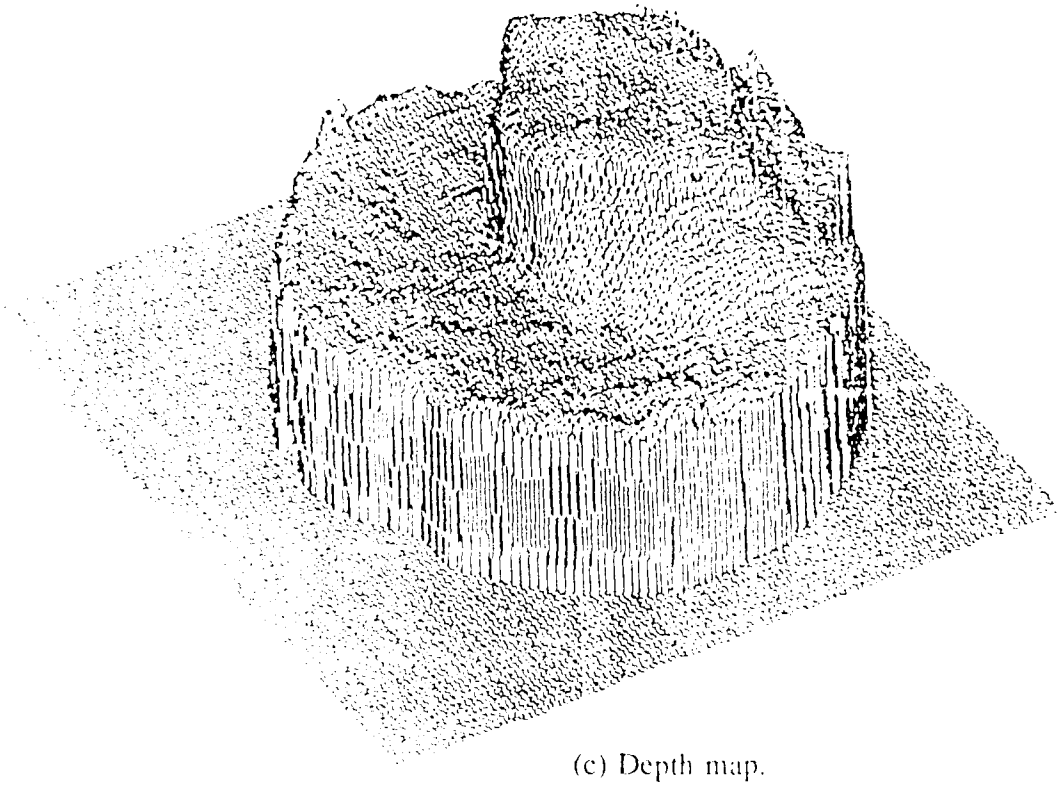
(f) Error map: Gaussian interpolation.

Figure 15: Steel ball. (continued)

(a) Camera image.



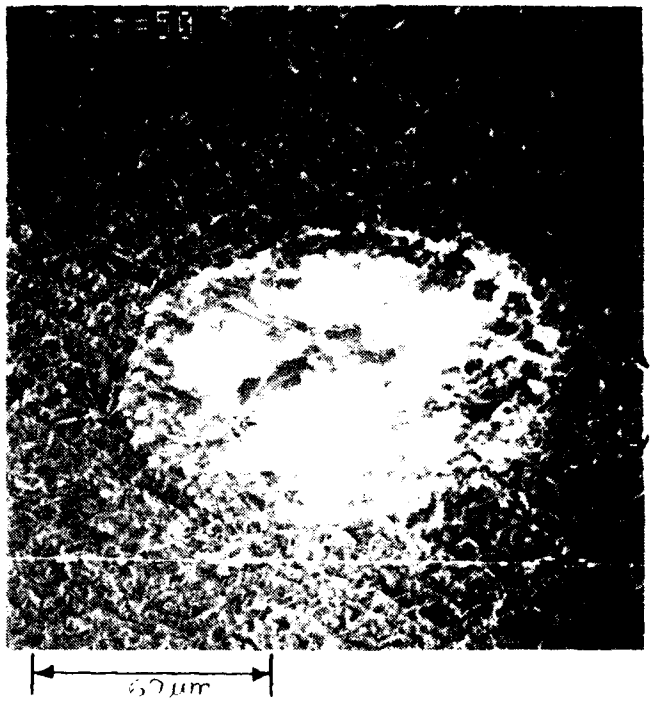
(b) SEM image.



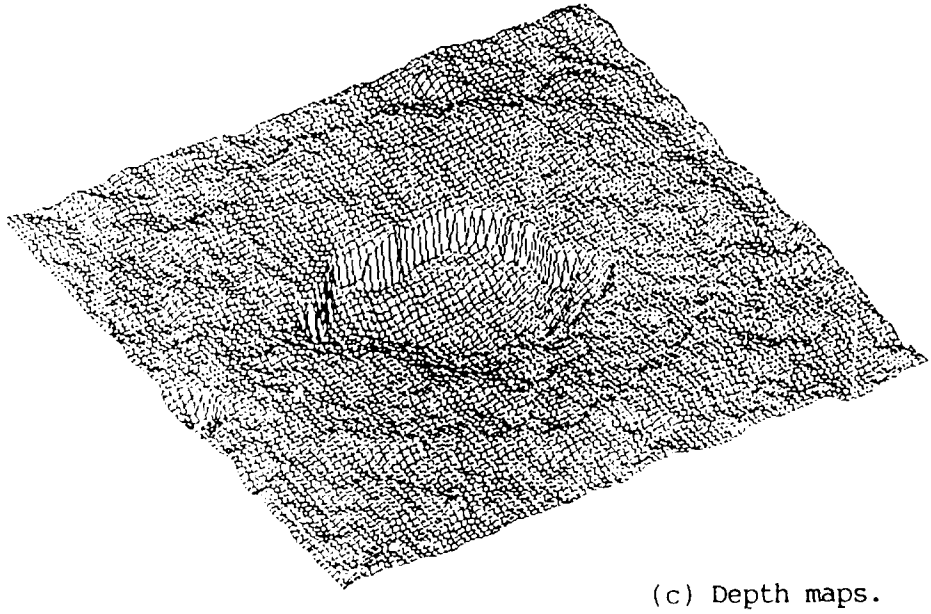
(c) Depth map.

Figure 16: Via-hole filling.

(a) Camera image.



(b) SEM image.



(c) Depth maps.

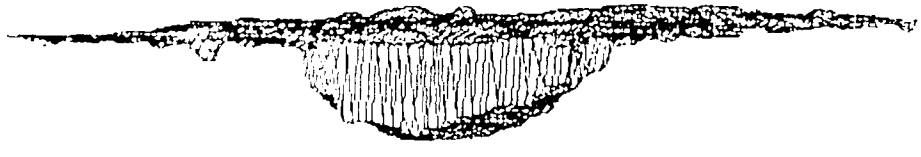
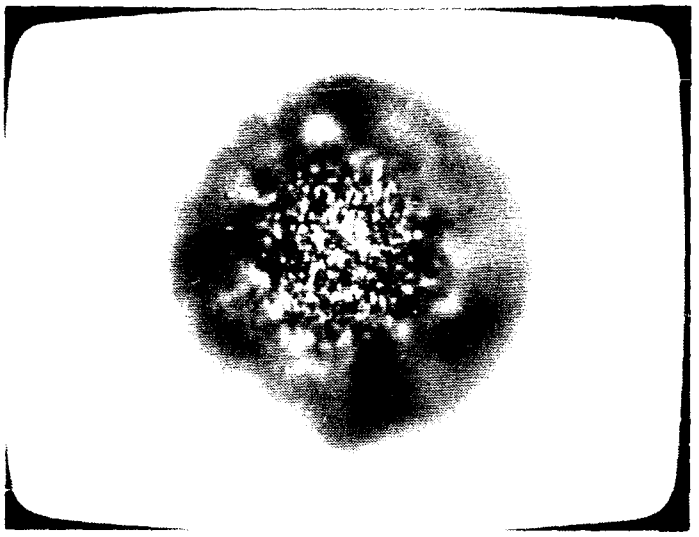
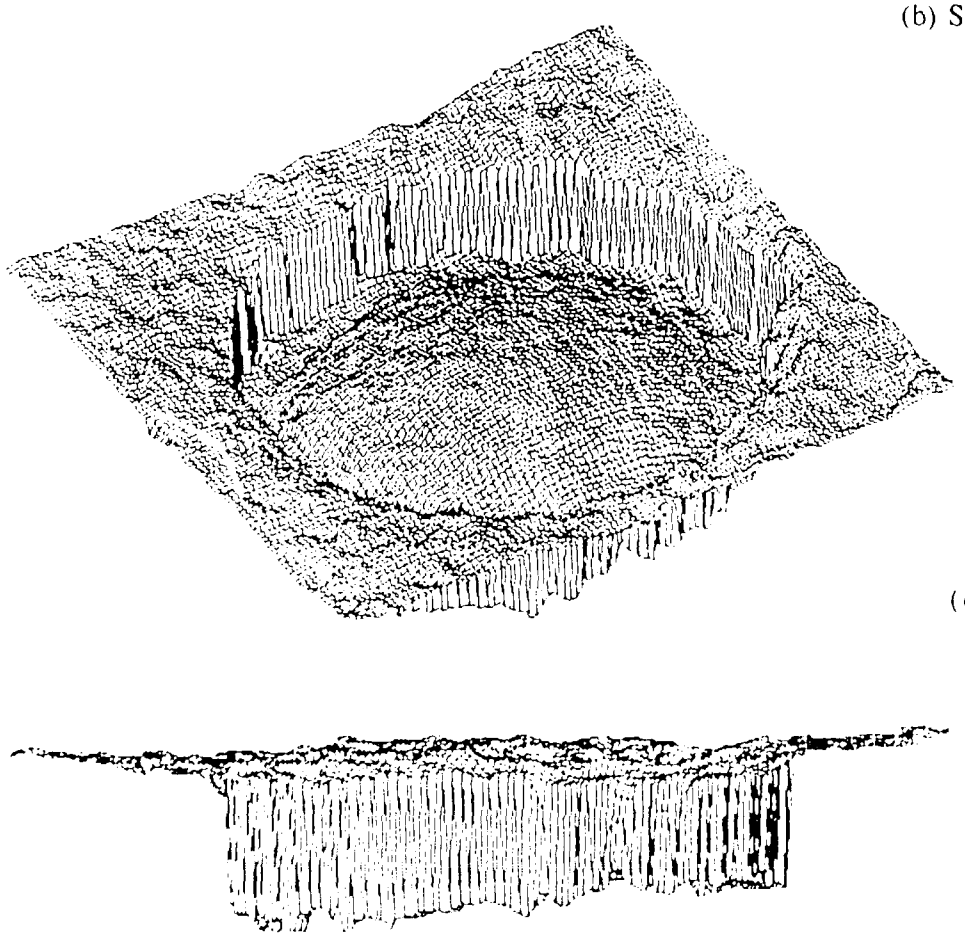


Figure 17: Baked via-hole filling.



(a) Camera image.

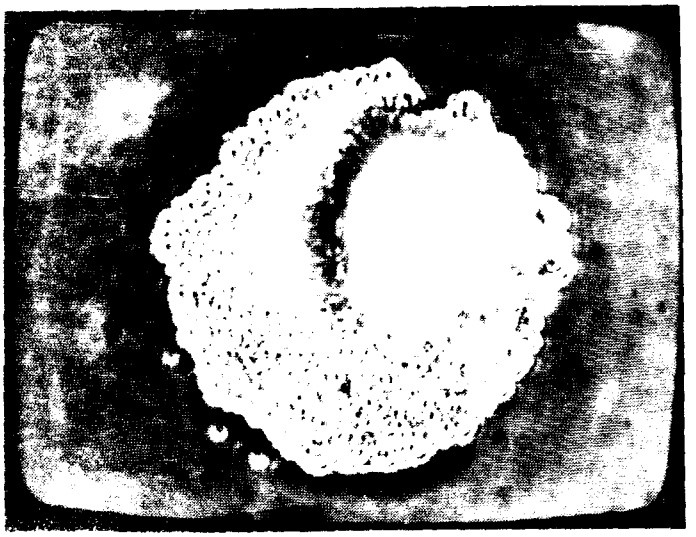
(b) SEM image.



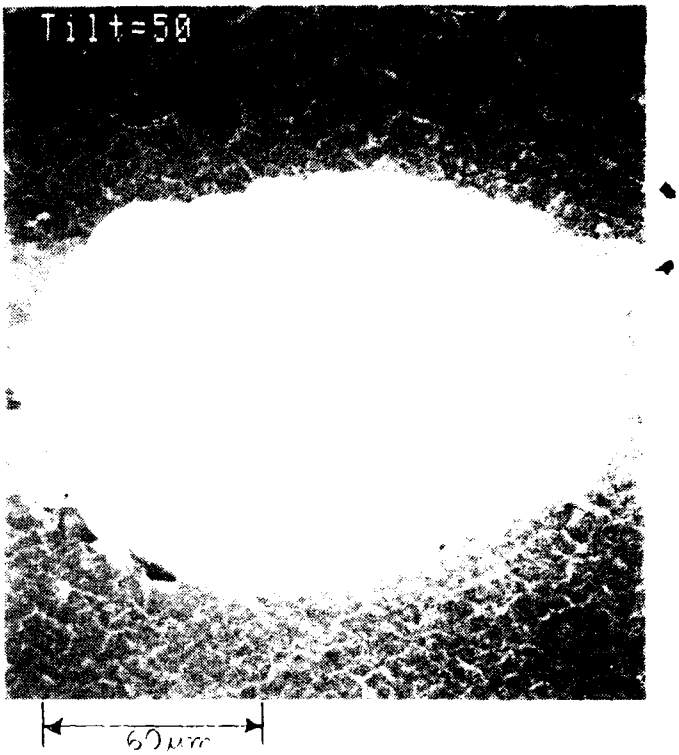
(c) Depth maps.

Figure 18: Baked via-hole filling.

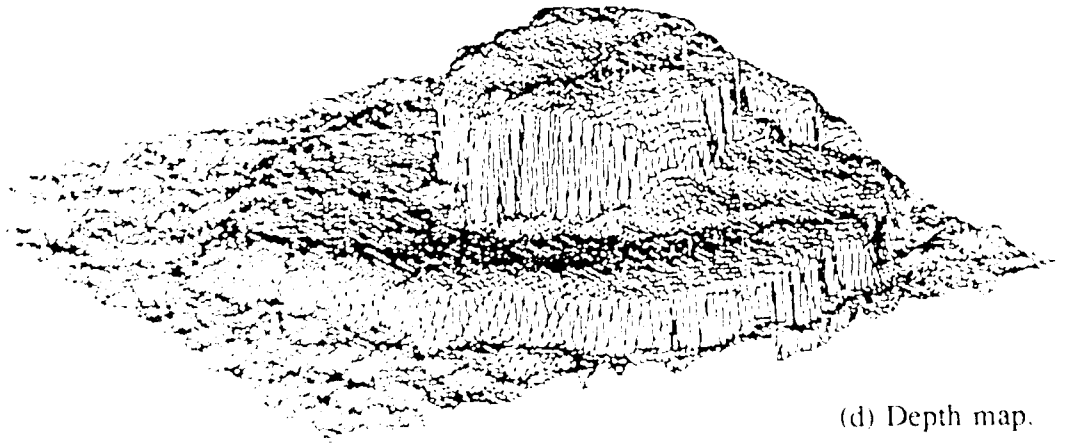
(a) Camera image: bright field illumination.



(b) Camera image: multi-spot illumination.



(c) SEM image.



(d) Depth map.

Fig. 20 Nickel-plated via-hole filling.

Acknowledgements

The author is extremely grateful to Yasuo Nakagawa of the Production Engineering Research Laboratory, Hitachi, for making this research effort possible and for his constant support and encouragement. The author wishes to thank Eric Krotkov, Katsushi Ikeuchi, and Takeo Kanade for their valuable comments and Kathy Porshe for reading earlier versions of this paper.

References

[1] B.K.P. Horn, *Focusing*, MIT Artificial Intelligence Laboratory, Memo No. 160, May, 1968.

[2] J.M. Tenebaum, *Accomodation in Computer Vision*, Ph.D. Thesis, Stanford University, 1970

[3] R.A. Jarvis, *Focus optimisation criteria for computer image processing*, Microscope, Vol. 24, No. 2, pp. 163-180, 1976.

[4] J.F. Schlag, A.C. Sanderson, C.P. Neumann, F.C. Wimberly, *Implementation of automatic focusing algorithms for a computer vision system with camera control*, Carnegie Mellon University, CMU-RI-TR-83-14, August, 1983.

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

[6] E. Krotkov, *Exploratory visual sensing with an agile camera*, Ph.D. Dissertation, TR-87-29, University of Pennsylvania, 1987.

[7] A. Pentland, *A new sense for depth of field*, Proc. IJCAI, Los Angeles, pp. 988-994, August, 1985.

[8] P. Grossmann, *Depth from Focus*, Pattern Recognition Letters, Vol. 5, pp. 63-69, 1987.

[9] T. Darrell and K. Wohn, *Pyramid Based Depth from Focus*, Proc. CVPR, pp. 504-509, 1988.

[10] M. Subbarao, *Direct Recovery of Depth Map 2: A New Robust Approach*, Technical Report, 87-03, State University of New York, Stony Brook, 1987.

[11] T. Ohta, K. Sugihara, and N. Sugie, *A Method for Image Composition using Image Variance*, Transactions of IECE, Vol. J66-D, No. 10, pp. 1245-1246.

[12] K. Kaneda, Y. Wakasu, E. Nakamae, E. Tazawa, *A Method of Pan-Focused and Stereoscopic Display Using a Series of Optical Microscopic Images,* Proc. of Fourth Symposium on Image Sensing Technologies in Industry, pp. 189-194, June, 1988.

[13] K. Torrance and E. Sparrow, *Theory for Off-Specular Reflection from Roughened Surfaces,* Journal of the Optical Society of America, No. 57, pp. 1105-1114, 1967.

[14] S. K. Nayar, K. Ikeuchi, T. Kanade, *Surface Reflection; Physical and Geometrical Perspectives,* Carnegie Mellon University, CMU-RI-TR-89-7, March, 1989.

[15] T. Ninomiya, M. Nomoto, Y. Nakagawa, *Automatic 2 1/2D Shape Inspection System for Via-Hole Fillings of Green Sheets by Shadow Image Analysis,* Proc. IEEE Intl. Conf. on Robotics and Automation, pp. 515-520, 1989.

[16] B. K. P. Horn, *Robot Vision,* MIT Press, 1986.

[17] G. Westheimer, *The Eye,* Medical Physiology, Thirteenth Edition, Mosby Co., 1974.