[4] T. S. Huang and S. D. Blostein, "Robust algorithms for motion estimation based on two sequential stereo image pairs," in *IEEE Conf. Comput. Vision Pattern Recognit.*, 1985, pp. 518–523.
[5] P. Huber, *Robust Statistics*. New York: Wiley, 1981.
[6] R.Y. Rubinstein, *Monto Carlo Optimization, Simulation and Sensitivity of Queuing Networks*. New York: Wiley, 1986.
[7] A. M. Waxman and J. H. Duncan, "Binocular image flows: Steps toward stereo-motion fusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 715–729, Nov. 1986.
[8] J. Weng, N. Ahuja, and T.S. Huang, "Two-view matching," in *Proc. 2nd Inter. Conf. Comput. Vision*, Tampa, FL, 1989, pp. 64–73.
[9] J. Weng and P. Cohen, "Robust motion estimation using stereo vision," in *Proc. IEEE Int. Workshop on Robust Comput. Vision*, Seattle, WA, Oct. 1990, pp. 367–388.
[10] G. S. Young and R. Chellappa, "3-D motion estimation using noisy stereo images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, No. 8, pp. 735–759, Aug. 1990.
[11] X. Zhuang and R. M Haralick, "A highly robust estimator for computer vision," in *Proc. 10th Int. Conf. on Pattern Recognit.*, vol. 1, pp. 545–550, Atlantic City, NJ, June 1990, pp. 17–22.
[12] X. Zhuang, T. Wang, and P. Zhang, "A highly robust estimator through partially likelihood function modeling and its application in computer vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no.1, pp. 19–35, 1992.

Shape from Focus

Shree K. Nayar and Yasuo Nakagawa

Abstract—**The shape from focus method presented here uses different focus levels to obtain a sequence of object images. The sum-modified-Laplacian (SML) operator is developed to provide local measures of the quality of image focus. The operator is applied to the image sequence to determine a set of focus measures at each image point. A depth estimation algorithm interpolates a small number of focus measure values to obtain accurate depth estimates. A fully automated shape from focus system has been implemented using an optical microscope and tested on a variety of industrial samples. Experimental results are presented that demonstrate the accuracy and robustness of the proposed method. These results suggest shape from focus to be an effective approach for a variety of challenging visual inspection tasks.**

Index Terms—**Defocus, focus operator, focus measure function, depth estimation, roughness and texture, automatic shape from focus.**

I. Introduction

Several recovery methods for diffuse and specular surfaces have been developed in the past. The recovery problem associated with rough surfaces, however, has not received sufficient attention. All surfaces encountered in practice are rough at some level of 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 resolution of the imaging system. This is most often the case with microscopic objects where a few microns of surface area can occupy an entire digital image. Image intensities produced by such surfaces vary in an unpredictable manner from one sensor element (pixel) to the next. Hence, it is difficult to obtain dense and accurate surface shape information by using existing passive and active sensing techniques, such as, binocular stereo, shape from shading, photometric stereo, and structured light. A practical and reliable solution to this rather difficult problem is therefore desirable.

In this correspondence, we develop a shape recovery technique that uses focus analysis to compute dense depth maps of rough textured surfaces. Focusing mechanisms play a vital role in the human vision system [8]. However, focus analysis alone does not serve as a reliable means of shape recovery. This limitation applies to machine vision as well. Previously, focus analysis has been used to automatically focus imaging systems [5], [17], [7], [10], [14], [9], or to obtain sparse depth information from the observed scene [9], [13], [4], [2], [15], [16], [11], [3]. In the context of general scenes, focus analysis may or may not provide the richness of information often required for scene interpretation. Here, we show that focus analysis can be put to great use by restricting ourselves to a particular class of surfaces. These are surfaces that produce textured images either due to their roughness or due to reflectance variations.

We begin by reviewing 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) 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. We show that focus measure variations due to defocusing can be approximated using a Gaussian model. This model is used to interpolate a small number of focus measures to obtain accurate depth estimates. Based on these ideas a fully automated shape-from-focus system has been developed. The system is applicable to objects that are up to a few hundred microns in size. An optical microscope is used to image the object. The translation stage of the microscope is motorized to enable controlled movement of the unknown object through the plane of focus of the imaging system. We conclude the correspondence with experiments conducted on industrial samples and a discussion on the proposed method.

II. Focused And Defocused Images

Fundamental to the concept of recovering shape by focus analysis is the relationship between focused and defocused images of a scene. In this section, we briefly review the image formation process and describe defocused images as processed versions of focused ones. A detailed discussion on this topic can be found in [1]. Figure 1 shows the basic image formation geometry. All light rays that are radiated by the object point P and intercepted by the lens are refracted by the lens to converge at the point Q on the image plane. For a thin lens, 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 it by a distance

Manuscript received July 2, 1992; revised June 24, 1993. Recommended for acceptance by Associate Editor R. Bolle.

S. K. Nayar is with the Department of Computer Science, Columbia University, New York, NY 10027 USA; e-mail: nayar@cs.columbia.edu.
Y. Nakagawa is with the Production Engineering Research Laboratory, Hitachi Ltd., Totsuka-Ku, Yokohama 244, Japan.

IEEE Log Number 9400036.

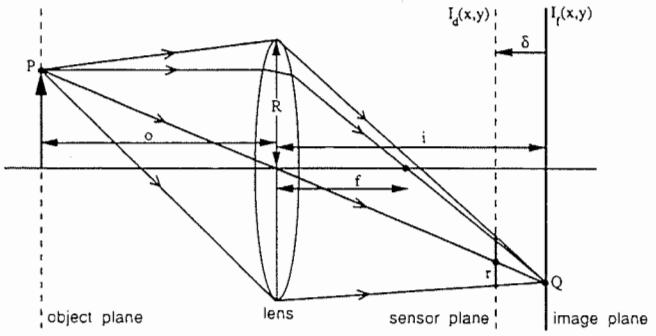


Fig. 1. Formation of focused and defocused images.

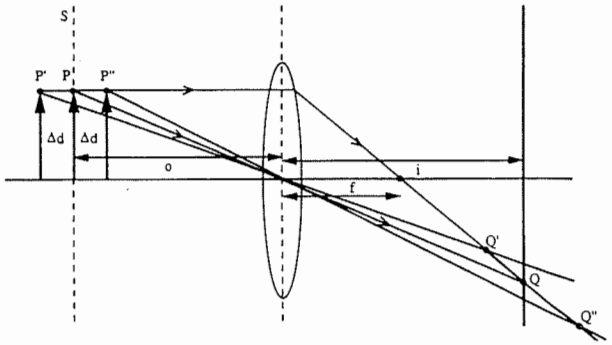


Fig. 2. Effect of object displacement on magnification.

δ, the energy received from the object by the lens is distributed over a circular[1] patch on the sensor plane. Fig. 1 may be used to establish the following relationship between the radius r of the circular patch and the sensor displacement δ:

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

where R is the radius of the lens. It is also possible to convince oneself that the radius r of the patch is independent of P 's location on the object plane. The distribution of light energy over the patch, or the *blurring function*, can be accurately modeled using physical optics [1]. Very often, a two-dimensional Gaussian function is used to approximate the physical model [13]. Then, the blurred or *defocused* image $I_d(x, y)$ formed on the sensor plane can be described as the result of convolving the focused image $I_f(x, y)$ with the blurring function $h(x, y)$:

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

where

$$h(x, y) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{x^2+y^2}{2\sigma_h}}. \tag{4}$$

The *spread parameter*, σ_h, is assumed to be proportional to the radius r [13]. The constant of proportionality is dependent on the imaging optics and the image sensor. We will see shortly that the value of this constant is not important in our approach. Note that defocusing is observed for both positive and negative sensor displacements.

Now consider the defocusing process in frequency domain (u, v). If $I_F(u, v)$, $H(u, v)$, and $I_D(u, v)$ are the Fourier transforms of $I_f(x, y)$, $h(x, y)$, and $I_d(x, y)$, respectively, we can express (3) as

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

where:

$$H(u, v) = e^{-\frac{u^2+v^2}{2}\sigma_h^2}. \tag{6}$$

We see that $H(u, v)$ allows low frequencies to pass while it attenuates the high frequencies in the focused image. Furthermore, as the sensor displacement δ increases, the defocusing radius r increases, and the spread parameter σ_h increases. Hence defocusing is a *low-pass* filtering process where the bandwidth decreases with increase in defocusing.

From Fig. 1, it is 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 with respect to one another causes the following problems. a) The magnification of the system varies, causing the image coordinates of focused points on the object to change. b) The area on the sensor over which light energy is distributed varies, causing a variation in image brightness. These effects are described in detail in [18]. In order to overcome these problems, we propose varying the degree of focus by moving the object with respect to a fixed configuration of the optical system and sensor (Fig. 2). Object movement is easily realized in industrial and medical applications. This approach ensures that as the object passes through the plane S, surface points that lie on S are perfectly focused onto the image plane with the same magnification. In other words, as the object moves, the magnification of imaging system can be assumed to be constant for image areas that are perfectly focused.

However, from Fig. 2 we see that points that lie outside plane S will be projected onto the image plane with different magnifications. In fact, the magnification of defocused object points will depend on their distance from the plane S. Note that for small displacements Δd, magnification may be assumed to be constant. We will use this assumption while developing the depth estimation algorithm.

III. Shape From Focus

Figure 3 shows a 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 defines a single plane, the "focused plane[2]," 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. Consider the surface element, s, that lies on the unknown surface, S. If the stage is moved towards the focused plane, the image of s 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 s and record the stage displacement $d = \overline{d}$ at the instant of maximum focus, the height d_s of s with respect to the stage can be computed as $d_s = d_f - \overline{d}$. This procedure may be applied independently to all surface elements to obtain the shape of the entire surface S.

To automatically detect the instant of "best" focus, an image focus measure will be developed. 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, only a small number of images are used; 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, an

[1] 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.

[2] The focused plane is the same as the object plane defined in the previous section. A different term is introduced here as the object does not necessarily lie on the focused plane.

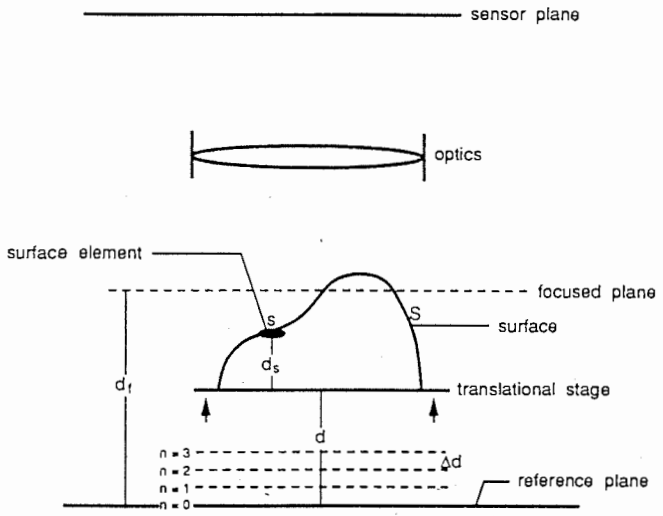


Fig. 3. Shape from focus.

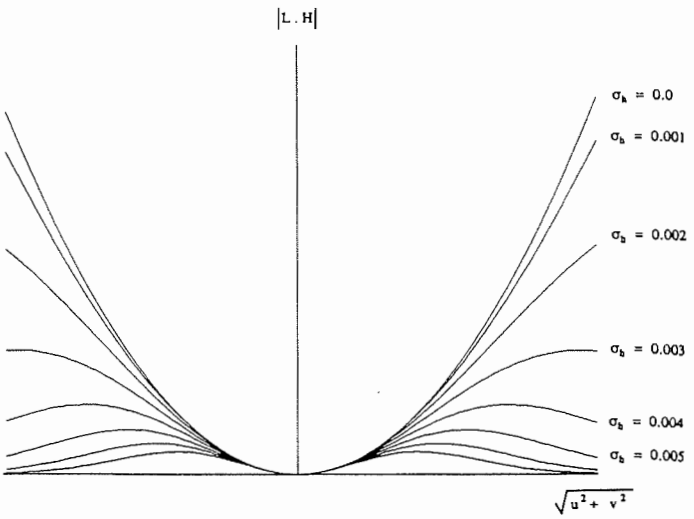


Fig. 4. The effect of defocusing and second-order differentiation in frequency domain.

interpolation method is developed that uses only a small number of focus measures to obtain accurate depth estimates. An important feature of the proposed method is its local nature; the depth estimate at an image point is computed only from focus measures recorded at that point. Consequently, it can adapt well to texture variations over the object surface.

IV. A Focus Measure Operator

To measure the quality of focus in a small image area, we 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. A few focus measure operators have been proposed and used in the past [10], [9]. Generally, the objective has been to find an operator that behaves in a stable and robust manner over a variety of images, including those of indoor and outdoor scenes. Such an approach is essential while developing automatically focusing 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.

An interesting observation can be made regarding the application of focus measure operators. Equation 3 relates a defocused image to a focused image using the blurring function. Assume that a focus measure operator $o(x, y)$ is applied (by convolution) to the defocused image $I_d(x, y)$. The result is a new image $r(x, y)$ that may be expressed as

$$r(x, y) = o(x, y) * (h(x, y) * I_f(x, y)). \tag{7}$$

Since convolution is linear and shift-invariant, we can rewrite the above expression as

$$r(x, y) = h(x, y) * (o(x, y) * I_f(x, y)). \tag{8}$$

Therefore, applying a focus measure operator to a defocused image is equivalent to defocusing a new image obtained by convolving the focused image with the operator. The operator only selects the frequencies in the focused image that will be attenuated due to defocusing. Since defocusing is a low-pass filtering process, its effects on the image are more pronounced and detectable if the image has strong high-frequency content. An effective focus measure operator, therefore, must high-pass filter the image.

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

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}. \tag{9}$$

In frequency domain, applying the Laplacian $L(u, v)$ to the defocused image $I_D(u, v)$ (5) gives

$$L(u, v) . H(u, v) . I_F(u, v) \tag{10}$$

where

$$L(u, v) . H(u, v) = -(u^2 + v^2) . e^{-\frac{u^2+v^2}{2}\sigma_h^2}. \tag{11}$$

Fig. 4 shows the frequency distribution of $| L . H |$ for different values of the defocusing parameter σ_h. For any given frequency (u, v), $| L . H |$ varies as a Gaussian function of σ_h. In general, however, the result would depend on the frequency distribution of the imaged scene. Though our texture is random, it may be assumed to have a set of dominant frequencies. Then, loosely speaking, each frequency is attenuated by a Gaussian function in σ_h and its width is determined by the frequency. Therefore, the result of applying the Laplacian operator may be expressed as a sum of Gaussian functions in σ_h. The result is expected to be maximum when $\sigma_h = 0$, i.e., when the image is perfectly focused. Since the frequency distribution of the texture is random, the widths of the Gaussian functions are also random. Using central limit theorem, the result of applying the Laplacian operator to an image point may be assumed to be a Gaussian function of the defocus parameter σ_h.

This general behavior is expected irrespective of the focus measure operator used. The operator only selects the frequencies that will play a dominant role in this process. Our experiments (in the following section) as well as Krotkov's empirical evaluation of various focus measure operators [9] support the above argument. As seen in [9], image noise and magnification variations will of course degrade the performance any operator.

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. In the case of textured images, this phenomenon may occur frequently and the Laplacian may at times behave in a unstable manner. We overcome this problem by defining the *modified Laplacian* as

$$\nabla^2_M I = \left|\frac{\partial^2 I}{\partial x^2}\right| + \left|\frac{\partial^2 I}{\partial y^2}\right|. \tag{12}$$

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, a discrete approximation to the modified Laplacian is obtained as

$$ML(x,y) = |\, 2I(x,y) - I(x-\text{step},y) - I(x+\text{step},y)\,| + |\, 2I(x,y) - I(x,y-\text{step}) - I(x,y+\text{step})\,| . \quad (13)$$

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

$$F(i,j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} ML(x,y) \text{ for } ML(x,y) \geq T_1. \quad (14)$$

The parameter N determines the window size used to compute the focus measure. In contrast to auto-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). Note that by definition the SML is not a linear operator and cannot be implement as a convolution. However, it can be computed using a simple algorithm.

V. Evaluating The Focus Measure

We evaluate the SML focus measure by analyzing its behavior as a function of the distance between the observed surface and the focused plane. A detailed description of the experimental setup is given in a later section. In the following experiments, texture samples are attached to a translational stage (Fig. 3) and the distance, d_s, of each sample from the stage is known by measurement. Images of the samples are obtained using a microscope and a 512× 480 pixel CCD camera. The complete imaging system has a physical resolution of approximately 1μm per pixel width.

In Fig. 5, 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 1μm, an image of the sample is obtained, and the SML measure is computed using an evaluation window of 10 × 10 pixels. The vertical lines in Fig. 5 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. Though the measure values are slightly noisy, they peak very close to the expected peak positions (vertical lines). We see that the focus measure function peaks sharply for the stronger texture but relatively slowly and with a lower peak value for the weaker texture.

The sharpness of the focus measure function, however, depends not only on the texture strength but also the depth of field of the imaging system. The depth of field, in turn, depends on the magnification and aperture size of the imaging optics as well as the physical resolution of the sensor. A smaller depth of field causes the focus quality of the image to vary more rapidly with object motion, causing a sharper peak. Note that the focus measure functions for both samples in Fig. 5 have Gaussian-like distributions near their peak values. This is consistent with the arguments given in the previous section. The fringes of the focus measure functions are less

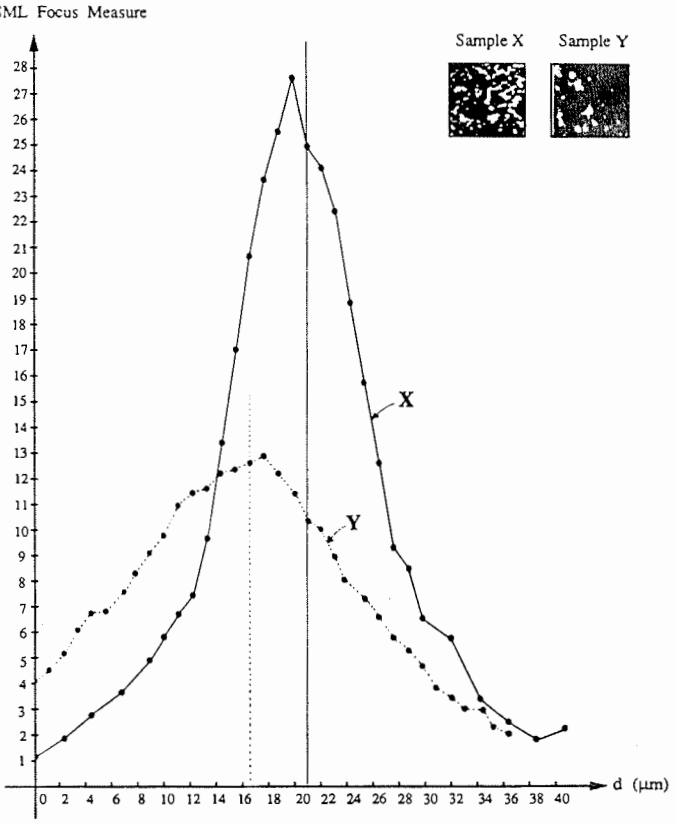


Fig. 5. SML focus measure function computed for two texture samples.

symmetric since magnification can vary substantially from one fringe to the other. As magnification varies, surface elements shift in image space and the focus measures computed within a fixed image window are unreliable. For this reason, we restrict our analysis to the peak region of the focus measure function. Magnification can be assumed to be constant around the peak if the depth of field is small. For a comparison between the SML measure and other focus measures such as *Tenengrad*, *variance*, and *sum-Laplacian* (SL) we refer the reader to [12].

VI. Depth Estimation

We now describe the estimation of depth $\overline{d}$ of a surface point (x,y) from the focus measure set $\{F(d_i) \mid i = 1,2,\cdots,M\}$. For convenience the notation F_i is used in place of $F(d_i)$. A course depth map can be obtained by using an algorithm that simply looks for the displacement value d_i that maximizes the focus measure and assigns that value to $\overline{d}$. A more accurate depth map is obtained by using the Gaussian distribution to interpolate the focus measures. The interpolation is done using only three focus measures, namely, F_{m-1}, F_m, and F_{m+1}, that lie on the largest mode[3] of $F(d)$, such that, $F_m \geq F_{m-1}$ and $F_m \geq F_{m+1}$[4] (Fig. 6(a)).

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

$$F = F_p \exp\left\{-\frac{1}{2}\left(\frac{d-\overline{d}}{\sigma_F}\right)^2\right\} \quad (15)$$

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

[4] Since the Gaussian model is valid only in the peak region of $F(d)$, these three focus measures must be computed in this region. Since the width of the Gaussian is unknown prior to depth estimation and can vary over the surface, it is difficult to select the stage displacement Δd in any systematic manner. Hence, in our experiments Δd was selected empirically.

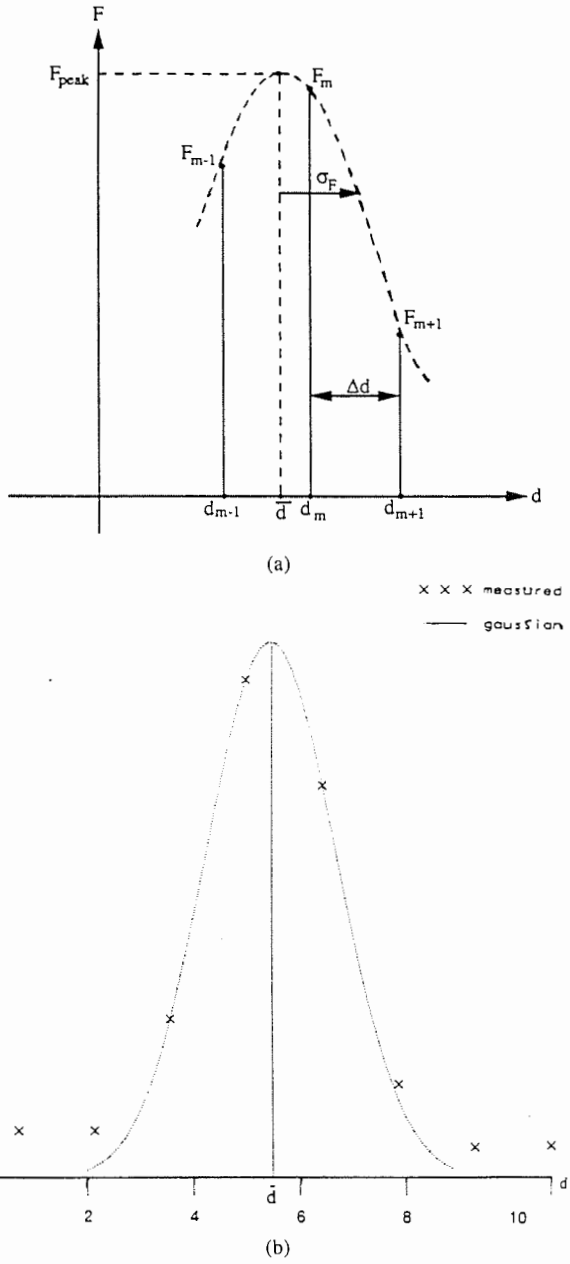


Fig. 6. Depth estimation: (a) Gaussian interpolation of focus measures; (b) experimental result.

where $\overline{d}$ and σ_F are the mean and standard deviation of the Gaussian distribution (Fig. 6(a)). Using natural logarithm, we have:

$$\ln F = \ln F_p - \frac{1}{2}\left(\frac{d - \overline{d}}{\sigma_F}\right)^2. \tag{16}$$

By substituting each of the three measures F_{m-1}, F_m, and F_{m+1}, and its corresponding displacement value in (16), we obtain a set of equations that can be solved for $\overline{d}$ and σ_F:

$$\overline{d} = \frac{(\ln F_m - \ln F_{m+1})(d_m{}^2 - d_{m-1}{}^2)}{2\,\Delta d\,\{(\ln F_m - \ln F_{m-1}) + (\ln F_m - \ln F_{m+1})\}} - \frac{(\ln F_m - \ln F_{m-1})(d_m{}^2 - d_{m+1}{}^2)}{2\,\Delta d\,\{(\ln F_m - \ln F_{m-1}) + (\ln F_m - \ln F_{m+1})\}} \tag{17}$$

$$\sigma_F{}^2 = -\frac{(d_m{}^2 - d_{m-1}{}^2) + (d_m{}^2 - d_{m+1}{}^2)}{2\,\{(\ln F_m - \ln F_{m-1}) + (\ln F_m - \ln F_{m+1})\}} \tag{18}$$

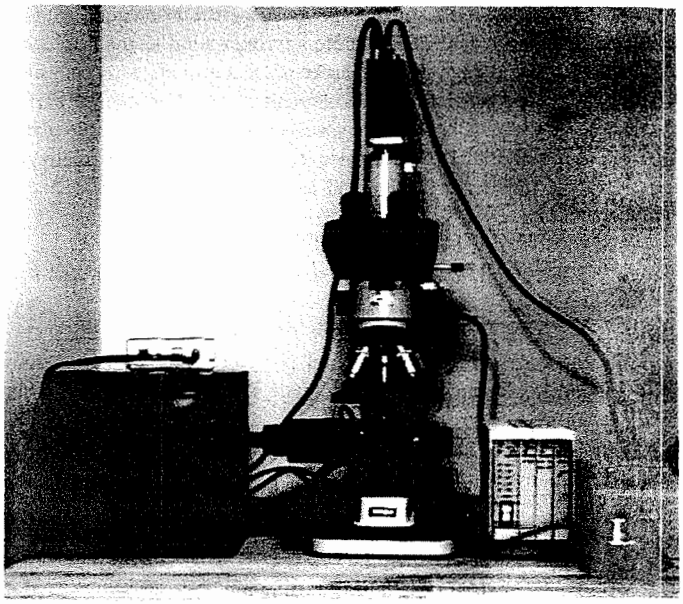


Fig. 7. Automated shape from focus system.

Using (15), we can find F_p from σ_F and $\overline{d}$ as:

$$F_p = F_m \,/\, \exp\left\{-\frac{1}{2}\left(\frac{d_m - \overline{d}}{\sigma_F}\right)^2\right\}. \tag{19}$$

If F_p 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, F_p and σ_F may be used to segment the observed scene into regions of different textures. Fig. 6(b) shows the experimental result of Gaussian interpolation applied to a real sample.

Depth computation can be done very efficiently by the use of look-up tables for each of the unknown parameters. Note that the three focus measure values F_{m-1}, F_m, and F_{m+1} and the corresponding displacements uniquely determine the parameters $\overline{d}$, F_p, and σ_F. Hence, the parameter values corresponding to all possible sets of the three focus measures can be computed a-priori and stored in a table. Given any set of focus measures, the corresponding parameter values are simply read from the table. However, the displacement parameter Δd is not fixed and may vary from one application to another. In order to generate look-up tables that are independent of Δd and the depth $\overline{d}$, we define the normalized focus measures:

$$\tilde{F}_{m-1} = \frac{F_{m-1}}{F_m}, \quad \tilde{F}_m = 1, \quad \tilde{F}_{m+1} = \frac{F_{m+1}}{F_m} \tag{20}$$

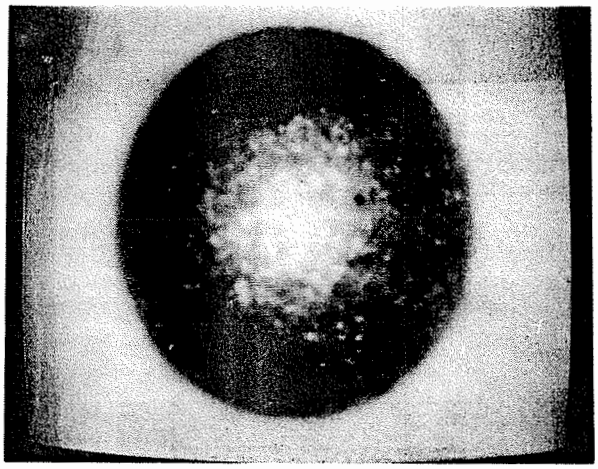
and the corresponding displacement values:

$$\tilde{d}_{m-1} = -1, \ \tilde{d}_m = 0, \ \tilde{d}_{m+1} = 1. \tag{21}$$

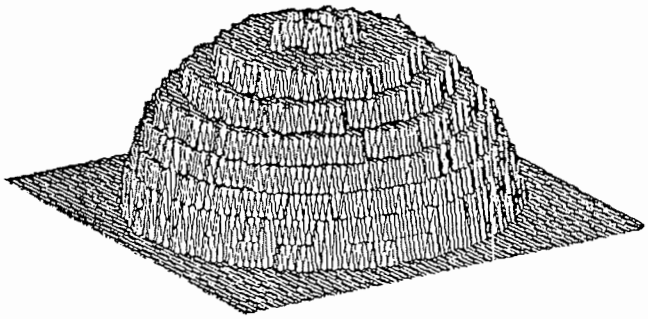
The above set of normalized measures and displacements uniquely determine a Gaussian function that passes through them. The parameters of this Gaussian are the mean value $\tilde{d}$, standard deviation $\tilde{\sigma}_F$, and peak value $\tilde{F}_p$. The correspondence between the normalized measures $\tilde{F}_{m-1}$ and $\tilde{F}_{m+1}$ and the parameters $\tilde{d}$, $\tilde{\sigma}_F$, and $\tilde{F}_p$ are precomputed using (17)–(19), respectively, and stored as three two-dimensional look-up tables. During depth estimation, the computed focus measures F_{m-1} and F_{m+1} are normalized to determine $\tilde{F}_{m-1}$ and $\tilde{F}_{m+1}$. These normalized measures are used to index the look-up tables and determine the parameters $\tilde{d}$, $\tilde{\sigma}_F$, and $\tilde{F}_p$. The parameters of the original Gaussian function are then determined as:

$$\overline{d} = d_m + \Delta d\ \tilde{d}, \quad \sigma_F = \Delta d\ \tilde{\sigma}_F, \quad F_p = F_m\ \tilde{F}_p. \tag{22}$$

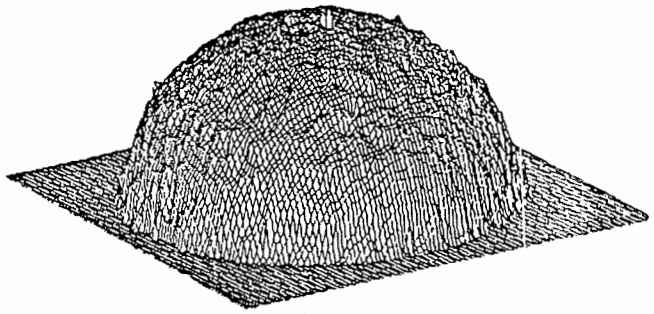
During run time, the use of look-up tables saves the computations involved in the evaluation of $\overline{d}$, F_p, and σ_F using (17)–(19).



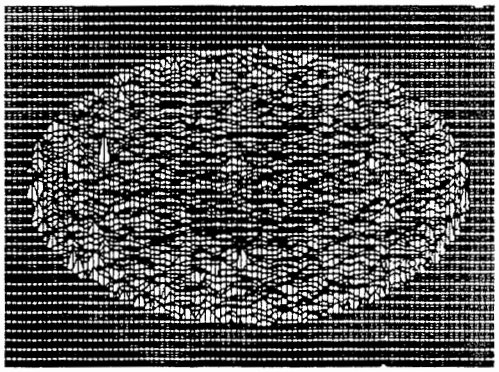
(a)



(b)



(c)



(d)

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)

Fig. 8. Experimental result: Steel ball. The known shape of the ball is used to analyze errors in the depth maps computed using the coarse resolution and Gaussian interpolation algorithms. (a) Camera image; (b) Depth Map: coarse resolution; (c) Depth Map: Gaussian interepolation; (d) Error map: Gaussian interpolation; (e) Error statistics.

VII. Automated Shape From Focus System

We have implemented an automated shape from focus system for recovering microscopic objects. A photograph of the system is shown in Fig. 7. Objects are imaged using a Nikon Alphaphot-2 model microscope and a CCD camera with 512×480 pixels. Magnification of the complete imaging system can be varied using objective lenses of different powers (×10, ×40, or ×100). Bright field illumination is used to illuminate the object; light energy is focused on the object via the same lenses that are used to image it. The z-axis of the microscope stage is driven by a stepper motor and the position of the stage can be computer controlled with a resolution and accuracy of 0.02 μm. The shape from focus algorithm is programmed and executed on a Sun SPARC 2 workstation. The complete recovery process including image acquisition, focus measure computation, and depth estimation takes a total of about 40 seconds for a sequence of 10 input image. We estimate that by using fairly simple customized hardware this recovery process can be accomplished in less than 1 second.

Prior to automating the shape from focus system, experiments were conducted to determine the accuracy and feasibility of the method. In these experiments, the microscope stage was moved manually and a sequence of images obtained and processed using both the coarse resolution and Gaussian interpolation depth recovery algorithms. The first experiment was conducted on a steel ball sample that is 1590 μm in diameter. The ball has a rough surface that gives it a

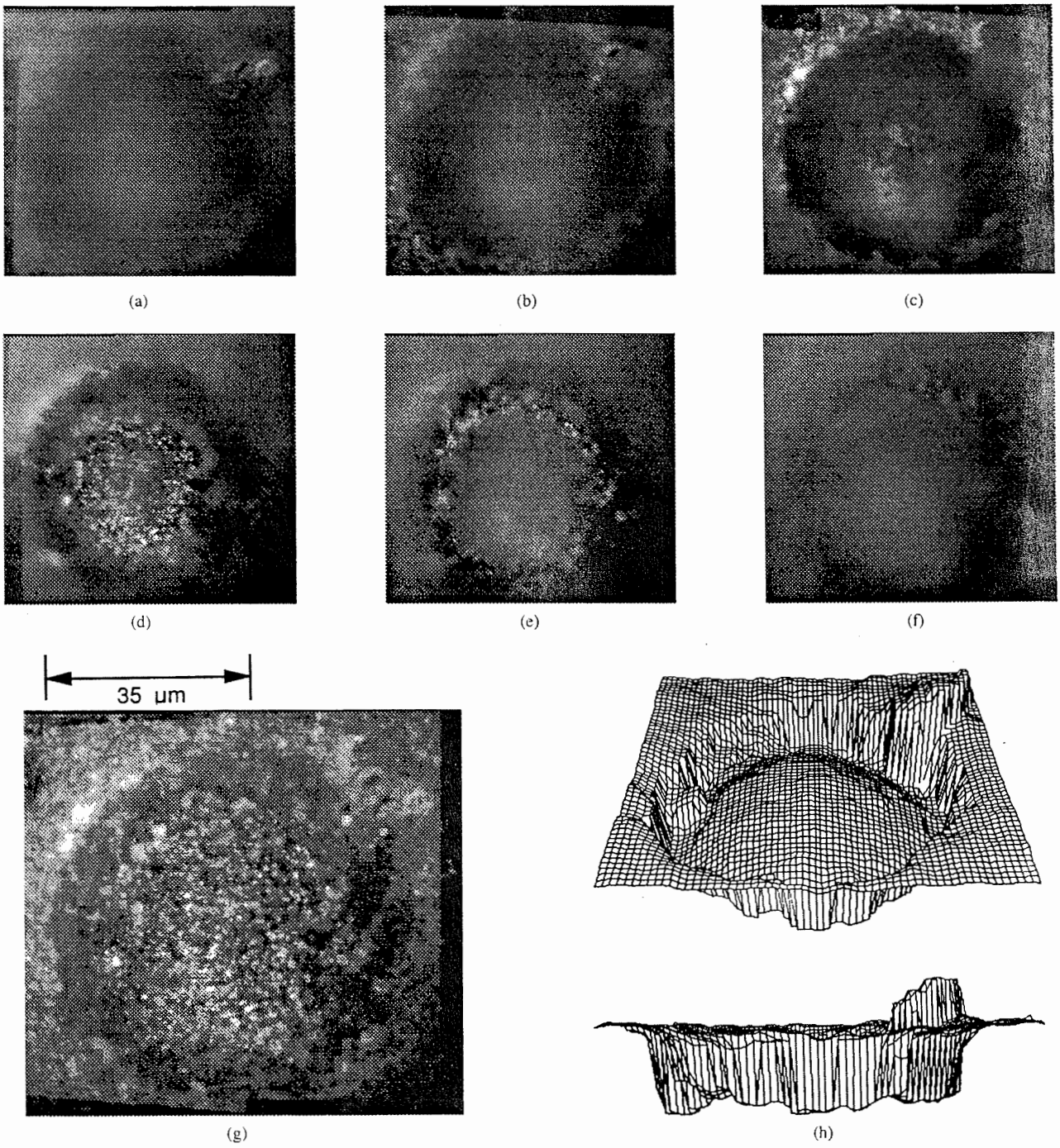


Fig. 9. Result obtained using the automated system: Via-hole filling on a ceramic substrate. The via-hole is approximately 70 μm and has insufficient filling. (a) $i = 2$; (b) $i = 5$; (c) $i = 8$; (d) $i = 11$; (e) $i = 14$; (f) $i = 18$; (g) Reconstructed image; (h) Depth maps.

textured appearance. A camera image of the ball, under bright field illumination, is shown in Fig. 8(a). Due to the small depth of field of the microscope, some areas of the ball are defocused. An incremental displacement of $\Delta d = 100$ μm was used to take 13 images of the ball. A 5×5 SML operator (with $T_1 = 7$ and step = 1) was applied to the image sequence to obtain focus measures. The course resolution depth map in Fig. 8(b) is computed by simply assigning to each surface point the depth value corresponding to the stage position that produced the maximum focus measure. Fig. 8(c) shows the depth map obtained using Gaussian interpolation.

The known size and location of the ball were used to compute error maps from the two depth maps. The error map for the Gaussian interpolation algorithm is shown in Fig. 8(d). The accuracy of the method depends on several factors: surface texture, depth of field of the imaging system, and the incremental displacement Δd. However, we see that there is no obvious correlation between the depth errors and surface orientation. The table in Fig. 8(e) shows error statistics computed from the error maps corresponding to the two algorithms. Clearly, the Gaussian interpolation algorithm produces far superior results than the course resolution one.

The automated system has been used to recover the shapes of a variety of industrial as well as biological samples. Here, we present one of these results. Fig. 9 shows a tungsten paste filling in a via-hole on a ceramic substrate. These fillings are used to

establish electrical connections between components on multilayered circuit boards. Conditions such as excess filling and lack of filling cause electrical defects. The sample shown in Fig. 9 has a cavity, indicating lack of filling. The specular reflectance and variable size of the tungsten particles gives the surface a random texture. In this case, a total of 18 images were taken using stage position increments of 8 μm. Some of these images are shown in Fig. 9(a)–(f). Fig. 9(g) and Fig. 9(h) show a reconstructed image and two views of the depth map, respectively. The image reconstruction algorithm simply uses the estimated depth to locate and patch together the best focused image areas in image sequence. The depths maps have been filtered using a 5 × 5 median filter to remove a few scattered erroneous depths that result from the lack of texture in some image areas.

VIII. Discussion

The above experiments demonstrate the effectiveness of the shape from focus method. The results show that the Gaussian interpolation algorithm performs stably over a wide range of textures. No assumptions are made regarding the type of the textures. Small 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 focus measures.

The above experiments were conducted on microscopic surfaces that produce complex textured images. Such images are difficult, if not impossible, to analyze using recovery techniques such as shape from shading, photometric stereo, and structured light. These techniques work on surfaces with simple reflectance properties. Since the samples are microscopic in size, it is also difficult to use binocular stereo. Methods for recovering shape by texture analysis have been researched in the past. Typically, these methods recover shape information by analyzing the distortions in image texture due to surface orientation. The underlying assumption is that the surface texture has some regularity to it. Clearly, these approaches are not applicable to surfaces that produce random and spatially varying textures. For these reasons, shape from focus may be viewed as an effective method for objects with complex surface characteristics.

Acknowledgment

The authors would like to thank U. Shah for his help in implementing the automated system. This work has benefited from discussions with R. Willson of the VASC Group at CMU.

References

[1] M. Born and E. Wolf, *Principles of Optics*. London: Pergamon, 1965.
[2] T. Darrell and K. Wohn, "Pyramid based depth from focus," in *Proc. CVPR*, 1988, pp. 504–509.
[3] J. Ens and P. Lawrence, "A matrix based method for determining depth from focus," in *Proc. CVPR*, 1991, pp. 600–606.
[4] P. Grossmann, "Depth from focus," *Pattern Recognit. Lett.*, vol. 5, pp. 63–69, 1987.
[5] B. K. P. Horn, "Focusing," MIT Artificial Intell. Lab., Memo no. 160, May, 1968
[6] ——, *Robot Vision*. Cambridge, MA: MIT Press, 1986.
[7] R. A. Jarvis, "Focus optimization criteria for computer image processing," *Microscope*, vol. 24, no. 2, pp. 163–180, 1976.
[8] J. F. Koretz and G. H. Handelman, "How the human eye focuses," *Scientific American*, pp. 92–99, July 1988.
[9] E. Krotkov, "Focusing," *Int. J. of Comput. Vision*, vol. 1, pp. 223–237 1987.
[10] G. Ligthart and F. Groen, "A comparison of different autofocus algorithms," in *Proc. of ICPR*, 1982, pp. 597–600.
[11] H. N. Nair and C. V. Stewart, "Robust focus ranging," in *Proc. CVPR*, 1992, pp. 309–314.
[12] S. K. Nayar and Y. Nakagawa, "Shape from focus," Tech. Rep., Dept. of Comput. Sci., Columbia Univ., CUCS 058-92, Nov. 1992.
[13] A. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 9, no. 4, pp. 523-531, July 1987.
[14] J. F. Schlag, A. C. Sanderson, C. P. Neumann, and F. C. Wimberly, "Implementation of automatic focusing algorithms for a computer vision system with camera control," Tech. Rep., Carnegie Mellon Univ., CMU-RI-TR-83-14, Aug. 1983.
[15] M. Subbarao, "Efficient depth recovery through inverse optics," *Machine Vision for Inspection and Measurement*, H. Freeman, Ed.. New York: Academic Press, 1989.
[16] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," Tech. Rep. 92.12.03, SUNY, Stony Brook, Dec. 1992.
[17] J. M. Tenenbaum, "Accommodation in computer vision," Ph.D. thesis, Stanford Univ., 1970.
[18] R. G. Willson and S. A. Shafer, "Dynamic lens compensation for active color imaging and constant magnification focusing," Tech. Rep., Carnegie Mellon Univ., CMU-RI-TR-91, Oct. 1991.

Gibbs Random Field Model Based Weight Selection For The 2-D Adaptive Weighted Median Filter

Levent Onural, M. Bilge Alp, and Mehmet İzzet Gürelli

***Abstract*—A generalized filtering method based on the minimization of the energy of the Gibbs model is described. The well-known linear and median filters are all special cases of this method. It is shown that, with the selection of appropriate energy functions, the method can be successfully used to adapt the weights of the adaptive weighted median filter to preserve different textures within the image while eliminating the noise. The newly developed adaptive weighted median filter is based on a 3×3 square neighborhood structure. The weights of the pixels are adapted according to the clique energies within this neighborhood structure. The assigned energies to 2- or 3-pixel cliques are based on the local statistics within a larger estimation window. It is shown that the proposed filter performance is better compared to some well-known similar filters like the standard, separable, weighted and some adaptive weighted median filters.**

***Index Terms*—Gibbs random field model, adaptive filtering, weighted median filter, image noise filtering.**

I. Introduction

Median filters have the capability of preserving edge information in images while eliminating the noise components. This brings an advantage over linear filters which suffer from edge smoothing and blurring. Especially with impulsive noise (in general, if the noise distribution has long tails), median filters outperform their linear

Manuscript received December 17, 1992; revised September 10, 1993. Recommended for acceptance by Associate Editor T. Caelli.

L. Onural and M. B. Alp are with the Department of Electrical and Electronics Engineering, Bilkent University, Bilkent, TR-06533, Ankara, Turkey; e-mail: onural@ee.bilkent.edu.tr.

M. İ. Gürelli is with the Signal Processing Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089 USA.

IEEE Log Number 9403148.