TR-92-006

Depth Recovery From Defocused Images

Jih-Jian Leu, Yi-Ping Hung, Chin-Hsing Chen

TR–92–006

Depth Recovery From Defocused Images

Jih–Jian Leu, Yi–Ping Hung, Chin–Hsing Chen

Insititute of Information Science
Academia Sinica
NanKang, Taipei, Taiwan, 115
Republic of China

Depth Recovery From Defocused Images

Jih-Jian leu[†‡] Yi-Ping Hung[†] Chin-Hsing Chen[‡]

[†] Institute of Information Science, Academia Sinica,
Nankang, Taipei, Taiwan, 11529 R.O.C.

[‡] Department of Electrical Engineering,
National Chen Kung University, Tainan, Taiwan, R.O.C.

Abstract

Depth perception is a very important task for recovering three dimensional geometry of the scenes. Pentland and other researchers had proposed several algorithms for estimating depth by measuring the amount of defocus (blurring), caused by inexact focusing. These algorithms could avoid correspondence problem, which had been recognized as the most difficult problem in stereo vision. To estimate the depth of the scenes from the amount of blurring, it is necessary to have knowledge about certain intrinsic camera parameters, e.g., focal length, distance between image plane and lens center, and aperture diameter, etc, which are difficult to be measured accurately. Here, we propose a new method for estimating the depth map of the scenes without measuring explicitly intrinsic parameters mentioned above. Instead, these parameters are composed into two composite parameters which can be calibrated easily. Once the composite camera parameters are calibrated off-line, the results can then be used for the frequency-domain approach (e.g. by Subbarao [6]) and the image-domain approach (e.g. by Hwang [8]) to obtain the depth of the scene. Experiments with real images show that these methods lead to good depth recovery.

Keyword : 3D Depth Estimation, Depth from Defocus, Focus Parameter Calibration

CHAPTER ONE
Introduction

1.1 Introduction

In the category of computer vision, depth perception is a very important task for enabling a mobile robot system to understand the three–dimensional relationship of the world space objects. There are many different methods to solve the depth perception problem , e.g. stereo and shape from shading. Different methods have different constraints and assumptions. Most of the research for recovering the scene geometry is based on a pin–hole camera model (e.g. : Ballard and Brown, 1982; Rosenfeld and Kak , 1982; Horn, 1986). But practical camera systems, including the human eyes, are not pin–hole camera but consist of convex lens. Pentland [4] had derived an algorithm to recover depths of the scene from defocused images (that is, objects in image are not in focus) based on convex lens model. He noticed the fact that most biological lens systems are exactly focused at only one distance along each radius from the lens into the scene; as the distance between the imaged point and the surface of exact focus increases or decreases, the imaged objects become progressively more defocused, and it was feasible to find the depth at given point in the scene by measuring the amount of blurring at the point in the image. Some researchers have used this phenomenon to derive several algorithms for recovering depth informations. We collectively call these techniques depth–from–defocus.

The depth along a given direction can also be obtained by some active ranging devices such as a sonar and laser range finder. In these methods the scene is scanned sequentially along different viewing directions to obtain a complete depth–map. In comparison with active ranging techniques, depth–from–defocus can obtain the depth map of the entire scene at once, irrespective of whether any part of the image is in focus or not, and the depth–map recovery process is parallel and involves only simple local computations. In comparison with some shape recovery processes such as stereo vision and motion analysis, depth–from–defocus are direct in the sense that three–dimensional scene geometry is recovered directly from intensity images of the scene and the correspondence problem does not arise.

To estimate the depth of the scenes by depth–from–defocus, it is necessary to measure blurring parameter and camera intrinsic parameters; however, intrinsic parameters are difficult to be measured accurately. Here, we propose a new method for estimating the depth of the scenes without measuring these parameters directly. Instead, these parameters are composed into two composite parameters which can be calibrated easily. Once these parameters are calibrated, the results can then be used for the frequency–domain approach (e.g. by subbarao [6]) and the image domain approach (e.g. by Hwang [8]). The combined methods could lead to a good depth recovery according to real experiments.

Depth–from–defocus is very different from the autofocusing technique. Autofocusing technique search for the lens setting that gives the best focus at a particular point and use the lens setting to recover depth at this particular point in the scene. The limitations of autofocusing method are that it estimates depth at only one point at a time and it requires to change the lens setting in order to search for the setting that yields the best focused image. In general, autofocusing method requires taking ten or even more images to estimate the depth at one point in the scene, while depth–from–defocus method can estimate the depth of the whole scene by taking only two defocused images.

1.2 Mathematical model for Depth–from–defocus

A camera system can be thought of as a thin lens model. For a thin lens, as shown in figure 1–1, we can obtain the following two equations according to the well–known lens formula.

$$\frac{1}{v_0} + \frac{1}{u_0} = \frac{1}{F} \tag{1}$$

$$\frac{1}{v} + \frac{1}{D} = \frac{1}{F} \tag{2}$$

where u_0, D are the distance between lens and the point source P2, P1, respectively; v_0 is the distance between the lens and image plane, r is the radius of the lens, and F is the focal length of the lens; v is the distance between the lens and the position at which P1 will focus.

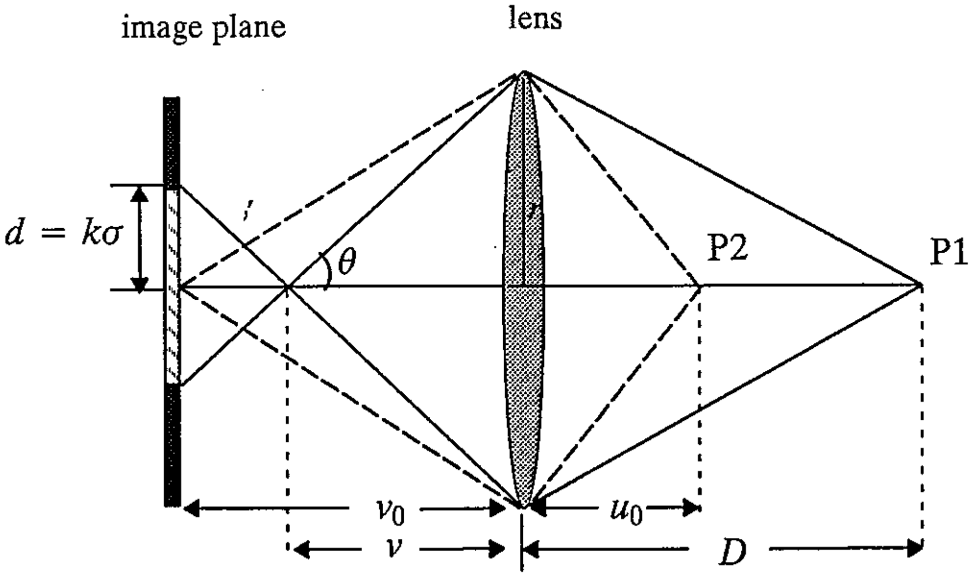

Figure 1–1: Geometry of Imaging

From the geometry of the Figure 1–1, we obtain

$$\tan\theta = \frac{r}{v} = \frac{d}{v_0 - v} \tag{3}$$

Using equations (1), (2) and (3), we can obtain

$$D = \frac{Frv_0}{rv_0 - F(r+d)} \qquad D \geq u_0 \tag{4}$$

Equation (4) can be rewritten as

$$D = \frac{Fv_0}{v_0 - F - 2df} \qquad D \geq u_0 \tag{5}$$

where $f = \frac{F}{2r}$ is the f–number of the lens. If $D \leq u_0$, we can derive

$$D = \frac{Fv_0}{v_0 - F + 2df} \qquad D \leq u_0 \tag{6}$$

The point P1 is not in focus so that it gives rise to a circular image called blur circle on the image plane. According to geometric optics, the indensity within the blurred circle is approximately constant and can be thought of as the point spread function $h1(x,y)$

$$h1(x,y) = \begin{cases} \frac{4}{\pi d^2} & \text{if } x^2 + y^2 \leq \frac{d^2}{4} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $\int\int h1(x,y)\ dxdy = 1$. But due to diffraction, aberration effects and other nonideal conditions, the point spread function will not be of the form of equation (7); the net effect is almost certainly best described by a two dimensional Gaussian function with a spatial constant σ [1][4] (we call σ the blurring parameter), i.e.,

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

where $r^2 = x^2 + y^2$. We assume that k is the proportionality constant between d and σ, that is,

$$d = k\sigma \tag{9}$$

The actual value of k depends on the characteristic of a given camera and is determined by an appropriate calibration procedure. Equations (5), (6), therefore, can be rewritten as

$$\begin{cases} D = \dfrac{Fv_0}{v_0 - F - 2k\sigma f} & D \geq u_0 \\ D = \dfrac{Fv_0}{v_0 - F + 2k\sigma f} & D < u_0 \end{cases} \tag{10}$$

We also assume defocusing process to be a linear shift–invariant process (Rosenfled and Kak, 1982); therefore, a blurred (defocused) image acquired by this camera system can be thought of as the result of convolving a focused image with a point spread function. Let $E(x,y)$ be a defocused local image patch. It can be expressed as

$$E(x,y) = G(r,\sigma) \otimes E_0(x,y) \tag{11}$$

where $E_0(x,y)$ is the corresponding focused image patch, $G(r,\sigma)$ is the point spread function which represents defocus operator, $\otimes$ represents the convolution operator. From equation (11), we see that the blurring parameter σ is covered in the defocused $E(x,y)$. If we find the value σ from $E(x,y)$ and measure the intrinsic parameter F , v_0 , f and k of the camera and then substitute these values into equation (10), we can obtain the distance D between lens and the part of scene corresponding to the image patch $E(x,y)$. Notice that the volumes of both $h1(x,y)$ and $G(r,\sigma)$ can be shown to be unity (The volume of the point spread function of a non–absorbing lens is unity irrespective of the form of point spread function).

1.3 Previous work

Pentland [4] was perhaps the first researcher to investigate depth recovery from the defocused images. Pentland assumed point spread function to be a two dimensional Gaussian function and proposed two methods for finding the depth–map of a scene. The first method was based on measuring the blurring of edges which are step discontinuities of intensity in the focused image. This method requires the knowledge of locations and magnitudes of step edges in the focused image. This information is rarely available in practical situations. As far as arbitrary scenes are concerned, Pentland proposed the second method which was based on comparing two different degrees of defocus of images caused by different known aperture diameter settings. He employed the Fourier transform to estimate the blurring circle radius in his mathematical derivation, but he simplified his algorithm by using Laplacian of Gaussian filter to estimate local high frequency contents in his implementation. He also assumed that value k of equation (10) is constant 1, but it may not be the value for the practical camera systems. Therefore he could only get very rough depth estimates.

Subbarao [6] proposed a general algorithm for depth recovery using defocused information. The general algorithm was based on changing camera focal length, aperture diameter and the distance between image plane and lens,then measuring the blurring parameter to estimate depth of the scene. Pentland's method can be thought of as the special case of Subbarao's method. But Subbarao could not measure the distance between image plane and lens in his camera system, he didn't provide actual experiment results .

These methods mentioned above are based on the Fourier domain analysis of an image. Hwang [8] derived an algorithm based on the spatial domain analysis. He proposed a two–phase algorithm where the point spread function is also modeled as a two dimensional Gaussian point spread function. In the first phase, a camera system parameter k of equation (10), is calibrated. In the second phase, depth was estimated by analyzing and comparing two image of the same scene but with a different amount of defocus caused by changing the distance between image and lens. The algorithm gives only relatively poor depth estimates.

These methods mentioned above must first estimate blurring parameter σ and intrinsic camera parameters f, F, v_0 and k, and then find depth of the scenes using equation (10). However, it is difficult to measure these intrinsic camera parameters accurately. Here, we proposed a new method for estimating the depth map of scene without measuring explicitly intrinsic parameters f, F, v_0 and k. Experiments with real images show that our method leads to a good depth–map recovery of the scene. Lai [14] also use similar calibration technique to avoid direct mearsuring of intrinsic camera parameters. But, in their method, the objects projecting to image must be of step edges. The condition is rarely met in practical situations.

CHAPTER TWO

Depth–from–Defocus Using frequency–domain and Image–domain approaches

2.1 Introduction

Assume that a defocused local image patch $E(x,y)$ is projected from a local portion of the scene with a constant depth, it can be expressed as

$$E(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \otimes E_0(x,y) \tag{12}$$

where $E_0(x,y)$ is the corresponding focused image patch, $\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$ is a point spread function, σ is the blurring parameter satisfying equation (10). The new method consists of two phases, the calibration phase and the depth–recovery phase. The first phase calibrate composite camera parameters by analyzing a simple known picture at different camera settings. In the second phase, the depth recovery is mainly based on both Subbarao's frequency–domain approach[6] and Hwang's image–domain approach[8]. Once we have determined these parameters off–line, we can then start to recover the depth map of arbitrary scenes.

2.2 Composite parameters calibration for Depth–from–Defocus

Considering only the case $D \geq u_0$, we can express equation (10) as

$$D = \frac{a}{b-\sigma} \tag{13}$$

$$\text{where } a = \frac{Fv_0}{2kf}, \quad b = \frac{v_0 - F}{2kf}$$

Suppose the values of a and b can be calibrated in advance, the depth D can be computed using equation (13) after σ is determined from the defocused image.

To calibrate the composite camera parameters a and b, we first rearrange equation (13) to

$$Db - a = D\sigma$$

For a fixed camera setting (i.e., fixed a and b), we can measure the blurring parameter σ_i, $i = 1, \ldots, N$, corresponding to different distance D_i by placing, at different D_i, a piece of paper having a sharp black–to–white transition shown in Figure 2–1. That is,

$$D_ib - a = D_i\sigma_i \qquad i = 1 \ \ldots \ N$$

The matrix form of the above equation is

$$\begin{bmatrix} D_1 & -1 \\ D_2 & -1 \\ \cdot & \cdot \\ \cdot & \cdot \\ D_N & -1 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} D_1\sigma_1 \\ D_2\sigma_2 \\ \cdot \\ \cdot \\ D_N\sigma_N \end{bmatrix}$$

This equation can be denoted by

$$Ax = B \tag{14}$$

$$\text{where } A = \begin{bmatrix} D_1 & -1 \\ D_2 & -1 \\ \cdot & \cdot \\ \cdot & \cdot \\ D_N & -1 \end{bmatrix}, \quad B = \begin{bmatrix} D_1\sigma_1 \\ D_2\sigma_2 \\ \cdot \\ \cdot \\ D_N\sigma_N \end{bmatrix}, \quad x = \begin{bmatrix} b \\ a \end{bmatrix}$$

We can solve equation (14) by using pseudo inverse of A, that is,

$$x = (A^T A)^{-1} A^T B$$

For different camera settings (changing the distance v_0 between image plane and lens center), we can determine the corresponding parameters a and b by the same calibration procedure.

The blurring parameter σ_i's corresponding to different D_i's can be measured as follows. Suppose the vertical transition between the black region and the white region is at x_0. Under proper lighting conditions, the ideal projected image (being in focus) is

$$E_0(x,y) = \begin{cases} g_1, & \text{if } x < x_0 \\ g_2, & \text{if } x \geq x_0 \end{cases}$$

By equation (12), the defocused image $E(x,y)$ can be shown to be

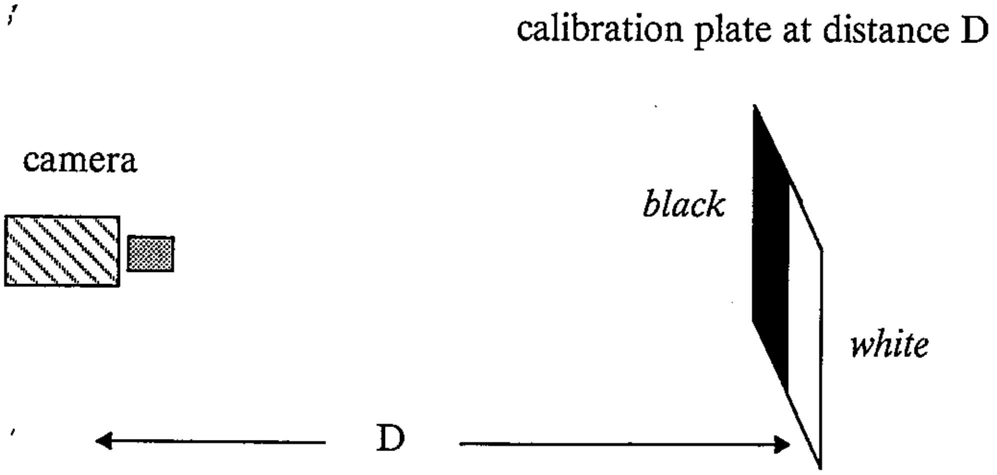

Figure 2–1: Calibration Setup

$$E(x,y) = g_1 N\left(\frac{-(x-x_0)}{\sigma}\right) + g_2 N\left(\frac{(x-x_0)}{\sigma}\right) \tag{15}$$

where $N(x)$ is the standard Gaussian distribution function $\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$.

Ideally, the blurring parameter σ can be determined easily by equation (15) if g_1, g_2, and $x - x_0$ are known. In practice, the true edge location x_0 is not known, and we can estimate it by finding the position of the maximum first directive along x–direction. Since the observed intensity data $E(x,y)$ always contains noise, we must use some techniques to get a better estimate of σ.

Consider only one scan line of the image, we can rewrite equation (15) as

$$E(x,y) = g_1 + N\left(\frac{x-x_0}{\sigma}\right) \cdot (g_2 - g_1) \tag{16}$$

Assume that x_0 and $\hat{x}_0$ are the x–coordinates of the correct and estimated transition (edge) positions, respectively. On the left side of estimated edge location $(\hat{x}_0, y)$, we select M points $\{(x_1,y), \ldots, (x_M,y)\}$ as shown in figure 2–2. If $\hat{x}_0$ is the true edge location, we can solve the following equation for the blurring parameter $\hat{\sigma}$:

$$E(x_j,y) = g_1 + N\left(\frac{x_j-\hat{x}_0}{\hat{\sigma}}\right) \cdot (g_2 - g_1)$$

However, the true x_0 may not equal to $\hat{x}_0$. For the true x_0, the true blurring parameter σ_t can be solved by

$$E(x_j,y) = g_1 + N\left(\frac{x_j-x_0}{\sigma_t}\right) \cdot (g_2 - g_1)$$

From the two equations mentioned above, we can obtain

$$\frac{x_j-\hat{x}_0}{\hat{\sigma}} = \frac{x_j-x_0}{\sigma_t} \equiv z_j \tag{17}$$

On the right side of the estimated edge location $(\hat{x}_0,y)$, we also select M points $\{(x_1',y), \ldots, (x_M',y)\}$ which are the mirror points of $\{(x_1,y), \ldots, (x_M,y)\}$ with respect to $\hat{x}_0$ as shown in figure 2–2. Similarly, we have

$$E(x_j',y) = g_1 + N\left(\frac{(x_j'-\hat{x}_0)}{\hat{\sigma}'}\right)\cdot(g_2-g_1)$$

$$E(x_j',y) = g_1 + N\left(\frac{(x_j'-x_0)}{\sigma_t}\right)\cdot(g_2-g_1)$$

From the two equations mentioned above, we can obtain

$$\frac{x_j'-\hat{x}_0}{\hat{\sigma}'} = \frac{x_j'-x_0}{\sigma_t} \equiv z_j' \tag{18}$$

Combine equations (17) and (18), we have $|z_j| = \frac{x_0-x_j}{\sigma_t}$, $|z_j'| = \frac{x_j'-x_0}{\sigma_t}$, and then

$$|z_j| + |z_j'| = \frac{|x_j'-x_j|}{\sigma_t}$$

Finally, the true blurring parameter can be calculated by

$$\sigma_t = \frac{|x_j'-x_j|}{|z_j| + |z_j'|} \tag{19}$$

We can take the average of it over the selected points in order to obtaining a better estimate of the blurring parameter, that is,

$$\sigma = \frac{1}{M}\sum_{j=1\ldots M}\frac{|x_j-x_j'|}{|z_j| + |z_j'|} \tag{20}$$

where x_j, $j = 1 \ldots M$ are the points along x–direction near the black–and–white transition. By repeatedly compute equation (20) on each scan line and taking the average over them, we can obtain an accurate blurring parameter σ.

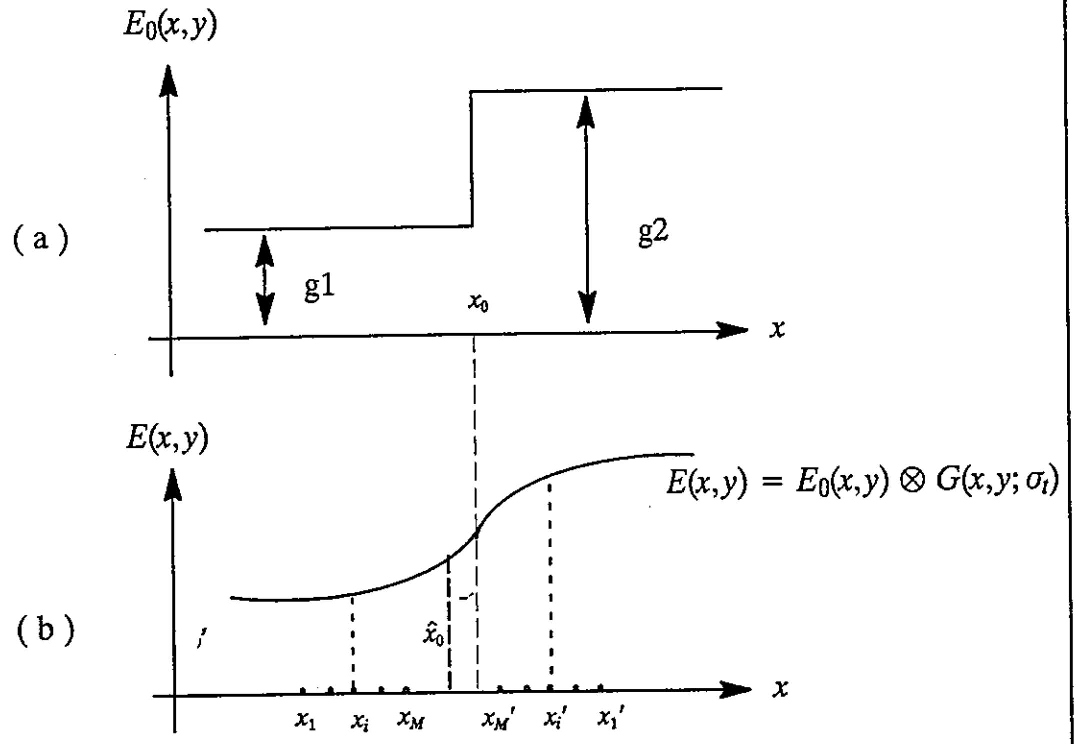

Figure 2–2: A step edge (a), and its defocused intensity (b)

2.3 Depth–from–Defocus using the frequency–domain approach

Let $E(x,y)$ be a small defocused patch on the image. $E_0(x,y)$ is the corresponding focused patch on the image. From equation (11), we obtain

$$E(x,y) = G(x,y;\sigma) \otimes E_0(x,y) \tag{21}$$

where σ is assumed to be a constant over the small patch. Notice that because different patches in the image result from different local portions of the scene may be of different depths, the parameter σ's of different image patches is usually different.

Now, we take two images $I_1(x,y)$ and $I_2(x,y)$ of the same scene with different focus by adjusting two camera settings to composite camera parameter (a1, b1) and (a2, b2), respectively. For the corresponding image patches $E_1(x,y)$ on $I_1(x,y)$ and $E_2(x,y)$ on $I_2(x,y)$, we can write down the following two equations :

$$E_1(x,y) = G(x,y;\sigma_1) \otimes E_0(x,y) \tag{22}$$

$$E_2(x,y) = G(x,y;\sigma_{2)} \otimes E_0(x,y) \tag{23}$$

where $E_0(x,y)$ is the exactly focused image patch of $E_1(x,y)$ and $E_2(x,y)$.

Taking Fourier transform on two equations (22) and (23), we obtain

$$\mathrm{E}_1(\omega_1,\omega_2) = e^{-\frac{1}{2}\sigma_1^2(\omega_1^2+\omega_2^2)}\ \mathrm{E}_0(\omega_1,\omega_2) \tag{24}$$

$$\mathrm{E}_2(\omega_1,\omega_2) = e^{-\frac{1}{2}\sigma_2^2(\omega_1^2+\omega_2^2)}\ \mathrm{E}_0(\omega_1,\omega_2) \tag{25}$$

where $\mathrm{E}_1(\omega_1,\omega_2)$, $\mathrm{E}_2(\omega_1,\omega_2)$ and $\mathrm{E}_0(\omega_1,\omega_2)$ are the Fourier transform of $E_1(x,y)$, $E_2(x,y)$ and $E_0(x,y)$, respectively; $\omega 1$, $\omega 2$ are spatial frequencies in radius per unit distance. The Fourier transform of $G(x,y;\sigma)$ is $e^{-\frac{1}{2}\sigma^2(\omega_1^2+\omega_2^2)}$.

Divide equation (24) by equation (25), we get

$$\frac{\mathbf{E}_1(\omega_1,\omega_2)}{\mathbf{E}_2(\omega_1,\omega_2)} = e^{\frac{1}{2}(\sigma_2^2-\sigma_1^2)(\omega_1^2+\omega_2^2)} \tag{26}$$

Taking the logrithm on either side of equation (26) and rearranging its terms, we have

$$\sigma_2^2 - \sigma_1^2 = \frac{2}{(\omega_1^2 + \omega_2^2)} \ln\left(\frac{\mathbf{E}_1(\omega_1,\omega_2)}{\mathbf{E}_2(\omega_1,\omega_2)}\right) \tag{27}$$

From equation (27) (refer to Subbarao[6]), we can calculate $\sigma_2^2 - \sigma_1^2$ from the observed images. Because σ_1, σ_2 are constants, the right hand side of equation (27) is ideally independent of frequency variable ω_1, ω_2 and we assume it to be C. That is,

$$\sigma_2^2 - \sigma_1^2 = C \tag{28}$$

The discrete formulation of the right hand side of equation (27) is

$$C = \frac{2}{(\frac{2\pi i}{NT_x})^2 + (\frac{2\pi j}{NT_y})^2} \ln\left(\frac{\mathbf{F}_1(i,j)}{\mathbf{F}_2(i,j)}\right) = \left(\frac{NT_y}{2\pi}\right)^2 \frac{2}{(\frac{i}{\varrho})^{2+}(j)^2} \ln\left(\frac{\mathbf{F}_1(i,j)}{\mathbf{F}_2(i,j)}\right) \tag{29}$$

$$\forall i,\ j = 1 \ldots N$$

where the size of the FFT is N by N; $\mathbf{F}_1(i,j)$ and $\mathbf{F}_2(i,j)$ are the magnitudes of the FFT's of $E_1(x,y)$ and $E_2(x,y)$; T_x, T_y are sampling intervals in the x, y directions, respectively; ϱ is the ratio $\frac{T_x}{T_y}$.

From equation (13) , we know that

$$\sigma_1 = \frac{Db_1 - a_1}{D} \tag{30}$$

$$\sigma_2 = \frac{Db_2 - a_2}{D} \tag{31}$$

Substituting equations (30) and (31) into equation (28), and rearranging it, we can obtain a quadratic equation

$$D^2(b_2^2 - b_1^2 - C)^2 \ + \ 2D(a_1b_1 - a_2b_2) \ + a_2^2 - a_1^2 \ = 0$$

The roots of this equation are

$$D = \frac{a_2b_2 - a_1b_1 \pm \sqrt{(a_1b_2 - a_2b_1)^2 + C(a_2^2 - a_1^2)}}{(b_2^2 - b_1^2 - C)} \tag{32}$$

One of the roots is close to F, the focal length. The other one is the depth of the scene that produces the image patch $E_1(x,y)$ and $E_2(x,y)$.

2.4 Depth-from-Defocus using the image-domain approach

Subbarao use Fourier transform to compute depth from defocus. T. L. Hwang proposed a differential approach in the spatial domain. Let $E(x,y)$ be a small projected patch centered at (x_0y_0). Again we know that

$$E(x,y) = G(x,y;\sigma) \otimes E_0(x,y) \tag{33}$$

In typical imaging system, the F is fixed whereas the v_0 and f can be changed by turning the respective rings on the camera lens. Hwang fixed f in the camera system, and v_0 was the only changeable parameter. He took derivatives with respect to v_0 on both sides of equation (33) and get

$$\frac{dE(x,y)}{dv_0} = \sigma\frac{d\sigma}{dv_0}\nabla^2 G(x,y;\sigma) \otimes E_0(x,y) \tag{34}$$

In our approach, we take derivatives with respect to b on both sides of equation (33)

$$\frac{dE(x,y)}{db} = \sigma\frac{d\sigma}{db}\nabla^2 G(x,y;\sigma) \otimes E_0(x,y) \tag{35}$$

Taking ∇^2 on both side of equation (33) and then divide equation (35) by it at $(x_0 y_0)$ if $\nabla^2 E(x_0 y_0)$ is not zero, we get

$$t(x_0,y_0) \equiv \left.\frac{\frac{dE(x,y)}{db}}{\nabla^2 E(x,y)}\right|_{(x,y)=(x_0 y_0)} = \left.\frac{\sigma\frac{d\sigma}{db}\nabla^2\{G(x,y;\sigma) \otimes E_0(x,y)\}}{\nabla^2\{G(x,y;\sigma) \otimes E_0(x,y)\}}\right|_{(x,y)=(x_0,y_0)}$$

$$= \sigma\frac{d\sigma}{db} \tag{36}$$

From equation (13), we have $\sigma = \frac{Db-a}{D}$ and $a = F\frac{v_0}{2kf} = bF + \frac{F^2}{2kf}$ and hence

$$\frac{d\sigma}{db} = \frac{d}{db}\left[\frac{(D-F)b - \frac{F^2}{2kf}}{D}\right] = \frac{D-F}{D}$$

Thus

$$\sigma\frac{d\sigma}{db} = \frac{Db-a}{D} \times \frac{D-F}{D}$$

$$\sigma\frac{d\sigma}{db} = \frac{D^2 b - D(bF+a) + aF}{D^2} \tag{37}$$

By equating $t(x_0,y_0)$ and right hand side of (37), we can obtain a quadratic equation

$$D^2(b-t) - D(bF+a) + aF = 0$$

The roots of this equation are

$$D = \frac{(bF+a)^2 \pm \sqrt{(bF+a)^2 - 4(b-t)aF}}{2(b-t)} \tag{38}$$

Suppose we have two image patches $E_1(x,y)$ and $E_2(x,y)$ obtained from the same scene but with a small different amount of defocus caused by changing the camera settings (different system parameters a, b). Equation (36) can be approximated by

$$t(x_0,y_0) \approx \frac{\frac{E_2(x_0y_0)-E_1(x_0,y_0)}{b_2-b_1}}{\frac{1}{2}\left(\nabla^2 E_1(x_0,y_0) + \nabla^2 E_2(x_0y_0)\right)} \tag{39}$$

The larger $\nabla^2 E_1(x_0y_0)$ and $\nabla^2 E_2(x_0y_0)$ are, the more textured the image patch would be, and the more reliable the recovered depth map would be. Substituting $t(x_0,y_0)$ in equation (39) into equation (38), we can obtain the depth of the scene at point (x_0,y_0), where $a \approx 0.5 \cdot (a_1 + a_2)$, $b \approx 0.5 \cdot (b_1 + b_2)$.

The two roots in equation (38) are positive if t is within a finite interval $(-B_l, B_h)$. One of them is close to F, the focal length . The other root (the larger one) is the scene depth corresponds to the pixel (x_0,y_0).

CHAPTER THREE
Verification of Depth–from–Defocus Model

3.1 Introduction

In the depth–from–defocus model, we assumed that a blurred (defocused) image acquired by the camera system can be thought of as the result of convolving a focused image with a Gaussian point spread function. That is, the defocus process is considered as a linear shift–invariant process and the point spread function is a two dimensional Gaussian function. Here, we use a method to verify these assumptions. According to the experimental results, we find that the characteristic of our camera close to the assumptions of depth–from–defocus model.

3.2 The method of verification

The main assumption of the depth–from–defocus is that the camera system is a linear–shift–invariant system and the point spread function is a two dimensional Gaussian function. We checked these assumptions by some experiments and found that it is a good approximation to our FUJINON cameras.

If $E_1(x,y)$, $E_2(x,y)$ are two images of the same scene taked by a camera at different settings with system parameters (a_1, b_1) and (a_2, b_2), then the two images must satisfy equations (22) to (28). For some (ω_1, ω_2), the right hand side of equation (27), namely $\frac{2}{(\omega_1^2 + \omega_2^2)} \ln\left(\frac{\mathrm{E}_1(\omega_1,\omega_2)}{\mathrm{E}_2(\omega_1,\omega_2)}\right)$, where $\mathrm{E}_1(\omega_1, \omega_2)$ and $\mathrm{E}_2(\omega_1, \omega_2)$ are the Fourier transform of $E_1(x,y)$ and $E_2(x,y)$, can be directly computed from the given image pair. In principle, compute its value at a specific frequency (ω_1, ω_2) is sufficient to obtain the value of $\sigma_2^2 - \sigma_1^2$, which should be a constant. But a more robust estimate can be obtained by taking the average over some sample points in the frequency space. Let the estimated average of $\sigma_2^2 - \sigma_1^2$ be C, thus

$$C = \frac{1}{L} \oint \frac{1}{\omega_1^2 + \omega_2^2} \ln \frac{\mathbf{E}_1(\omega_1, \omega_2)}{\mathbf{E}_2(\omega_1, \omega_2)} \, dl$$

$$C = \frac{1}{2\pi r} \int_0^{2\pi} \frac{1}{\omega_1^2 + \omega_2^2} \ln \frac{\mathbf{E}_1(\omega_1, \omega_2)}{\mathbf{E}_2(\omega_1, \omega_2)} \, r d\theta$$

$$C = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{\omega_1^2 + \omega_2^2} \ln \frac{\mathbf{E}_1(\omega_1, \omega_2)}{\mathbf{E}_2(\omega_1, \omega_2)} d\theta \tag{40}$$

where L is the perimeter of a circle with radius r in frequency domain.

From equation (29), the discrete formulation of equation (40) is given by

$$C(r) = \left(\frac{NT_y}{2\pi} \right)^2 \frac{1}{M} \sum_{k=0}^{M-1} \frac{2}{r^2} \ln \left\{ \frac{\mathbf{F}_1\left([\varrho r \cos \frac{2\pi k}{M}], [r \sin \frac{2\pi k}{M}] \right)}{\mathbf{F}_2\left([\varrho r \cos \frac{2\pi k}{M}], [r \sin \frac{2\pi k}{M}] \right)} \right\}; \; r = 1 \ldots N \tag{41}$$

where the size of FFT is N by N; $\mathbf{F}_1(i,j)$, $\mathbf{F}_2(i,j)$ are the magnitudes of the FFT's of $E_1(x,y)'$, $E_2(x,y)$; T_x, T_y are sampling intervals in the x,y directions; ϱ is the ratio $\frac{T_x}{T_y}$; M is the number of points locating at a circle of radius r; $[n]$ represents the integer nearest to n.

3.3 Experimental results

If equation (41) is constant in all r, we can make sure that the defocusing process is linear shift–invariant process and the point spread function is a two dimensional Gaussian function. We take the first image $E_1(x,y)$ by using a FUJINON camera with focal length 25mm as shown in figure 3–1a, then take the second image again after changing v_0 arbitrarily as shown in figure 3–1b. The curve $C(r)$ of equation (41) is shown in figure 3–2. The size of FFT been used is 256 X 256.

From figure 3–2, we can see that $C(r)$ is roughly constant in the period P1 and it descends gradually in P2. The main reason for this phenomenon is that the higher the frequency is, the more severe the influence of the quantization noise is. From figure 3–2, we see that the characteristics of our camera obey the assumptions, that is, the defocusing process is a linear shift–invariant process and the point spread function is a two dimensional Gaussian function.

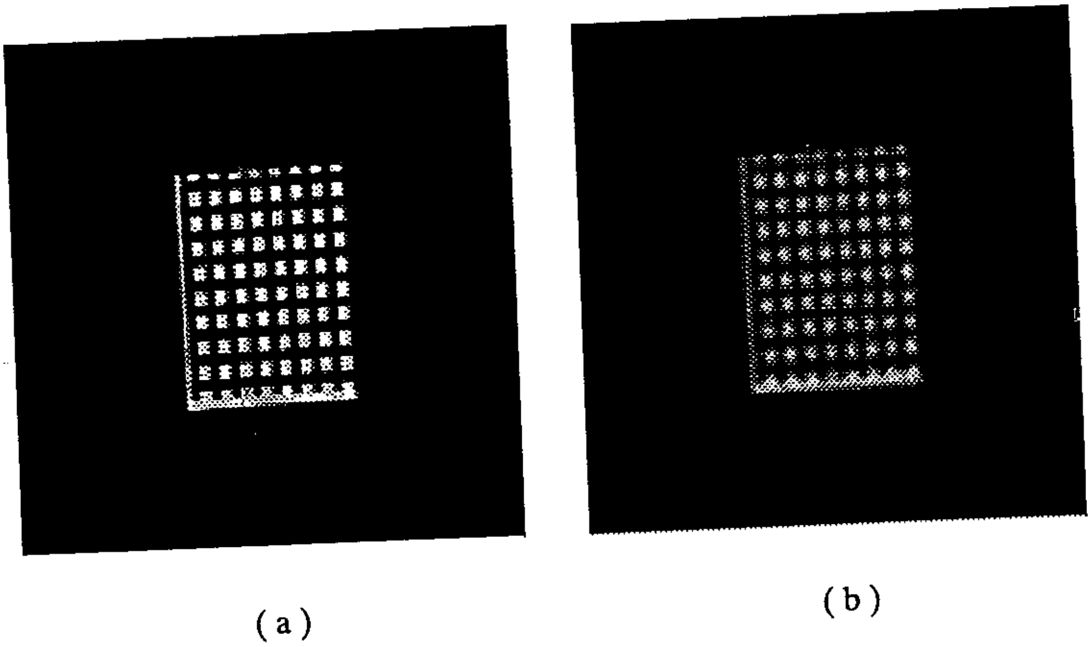

Figure 3–1 : Two defocused images taken by different v_0

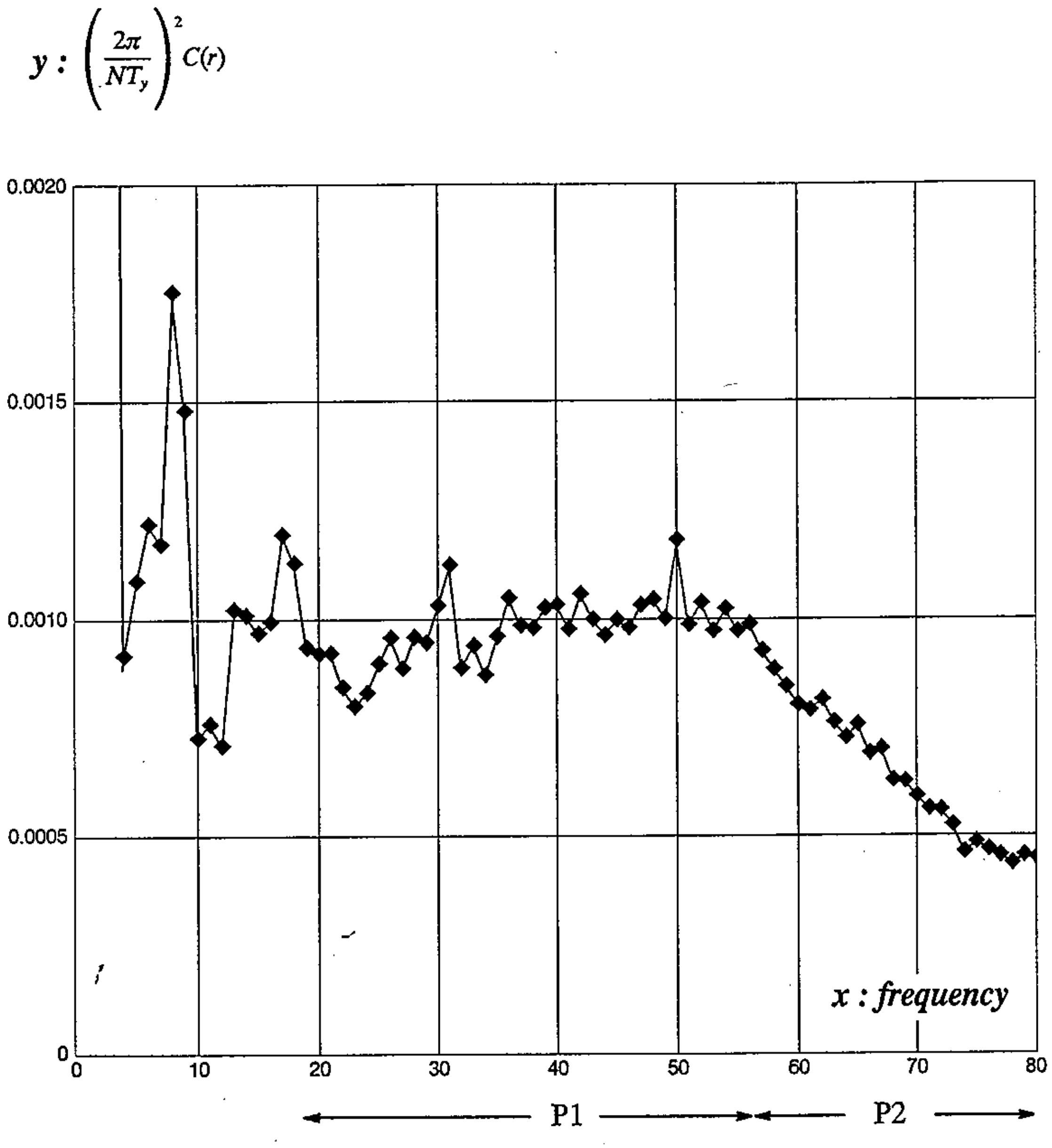

Figure 3–2 : The curve C(r); it is roughly constant in the period P1.

CHAPTER FOUR

Experimental results of Depth–from–Defocus

4.1 Calibration for Depth–from–Defocus

In this section, some real experiments are presented. All images are taken by a vision system called IIS–head (see figure 4–1). Although the IIS–head is a binocular vision system with two CCD camera on the top of it, we only use one of its camera in the experiments. The CCD camera is a PULNiX TM–745E. The image grabber is an ITI series 151. And the lens is an FUJINON lens 1:1.4/25mm.

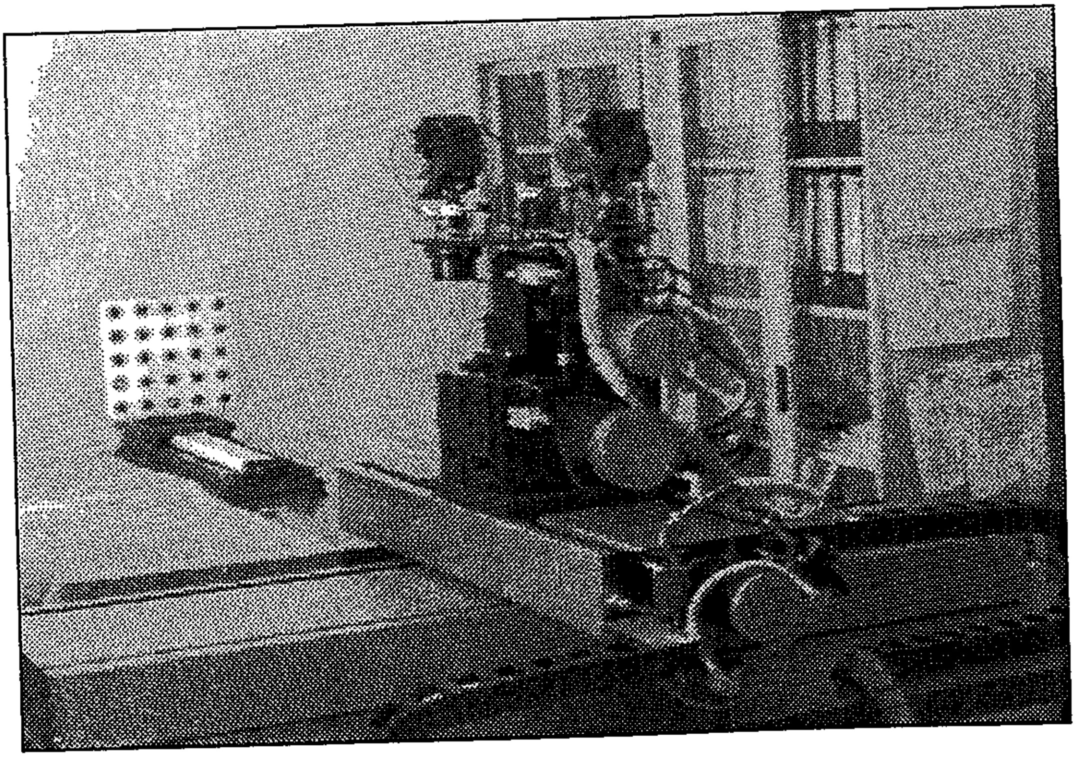

Figure: 4–1 A picture of the IIS–head

We calibrated three different camera settings, with $u_0 = 110$ cm, 130 cm, and 150 cm, to get three sets of system parameters . Table 1 show the calibration results.

4.2 The overlap problem in the frequency–domain approach

In the frequency–domain method, we divide an observed image into smaller subimages within which the depths of the scene are nearly constant. If this assumption is not valid inside a subimage, then this method gives an averaged depth of objects in that subimage which is still an useful piece of information.

Dividing an image into subimages introduces some errors due to border effects. A subimage cannot be analyzed in isolation because, due to blurring, the inten-

setting 1: $u_0 = 110cm$

D	σ (pixels)
75 cm	–3.7
65 cm	–5.3
55 cm	–7.7
45 cm	–11.1

a (cm)	1.322988
b (cm^2)	0.011996

setting 2: $u_0 = 130cm$

D	σ (pixels)
90 cm	–3.1
80 cm	–4.1
70 cm	–5.5
60 cm	–7.7

a (cm)	1.294038
b (cm^2)	0.009864

setting 3: $u_0 = 150cm$

D	σ (pixels)
90 cm	–3.7
80 cm	–5.1
70 cm	–6.5
60 cm	–8.5

a (cm)	1.294000
b (cm^2)	0.008623

Table 1: Three sets of system parameters for three different camera settings

sities near, and within the border of a subimage is affected by the intensities just outside the region. Subbarao calls it the image overlap problem. In the experiment in section 3–3, we painted the border of the images dark in order to avoid the overlap problem. In this section, the image overlap problem may be reduced as follows. The image intensity is multiplied by a suitable center weighted mask (e.g. a Gaussian function) centered at the region of interest. Because the weights are higher at the center than at the periphery, this scheme gives a depth estimate which is approximately the depth along the center of the field of view. Here, we choose the Hanning window to be the weighting function.

Figure 4–2 shows two image patches and the size of each patch is 64 by 64. Figure 4–3 shows the C(r) curve of the image patches without multiplying weighting function. Figure 4–4 shows the C(r) curve of the image patches multiplied with Hanning window function. We can see that in Figure 4–4, C(r) is roughly constant over a period.

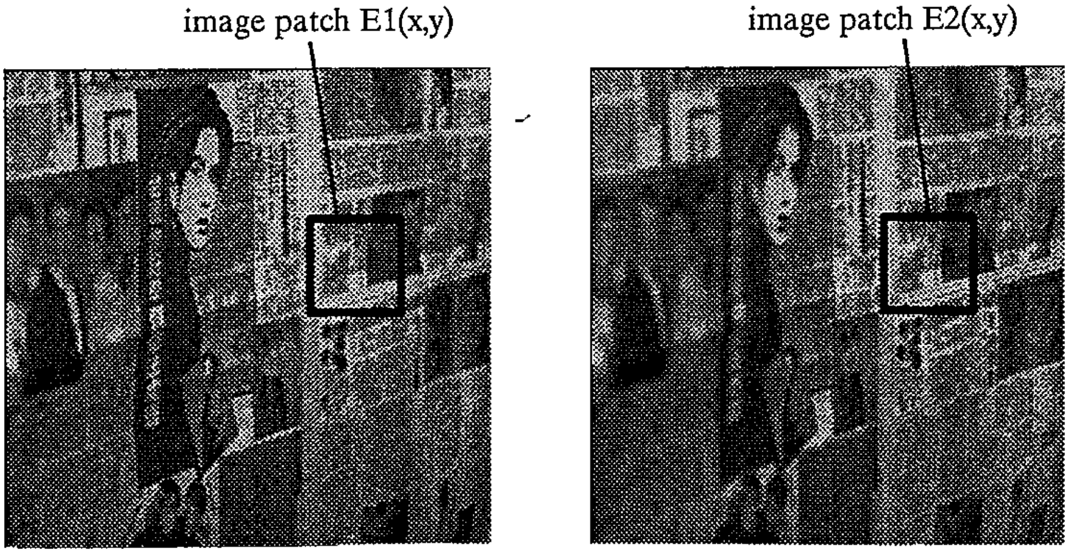

Figure 4–2: Two defocused image patches E1(x,y), E2(x,y)

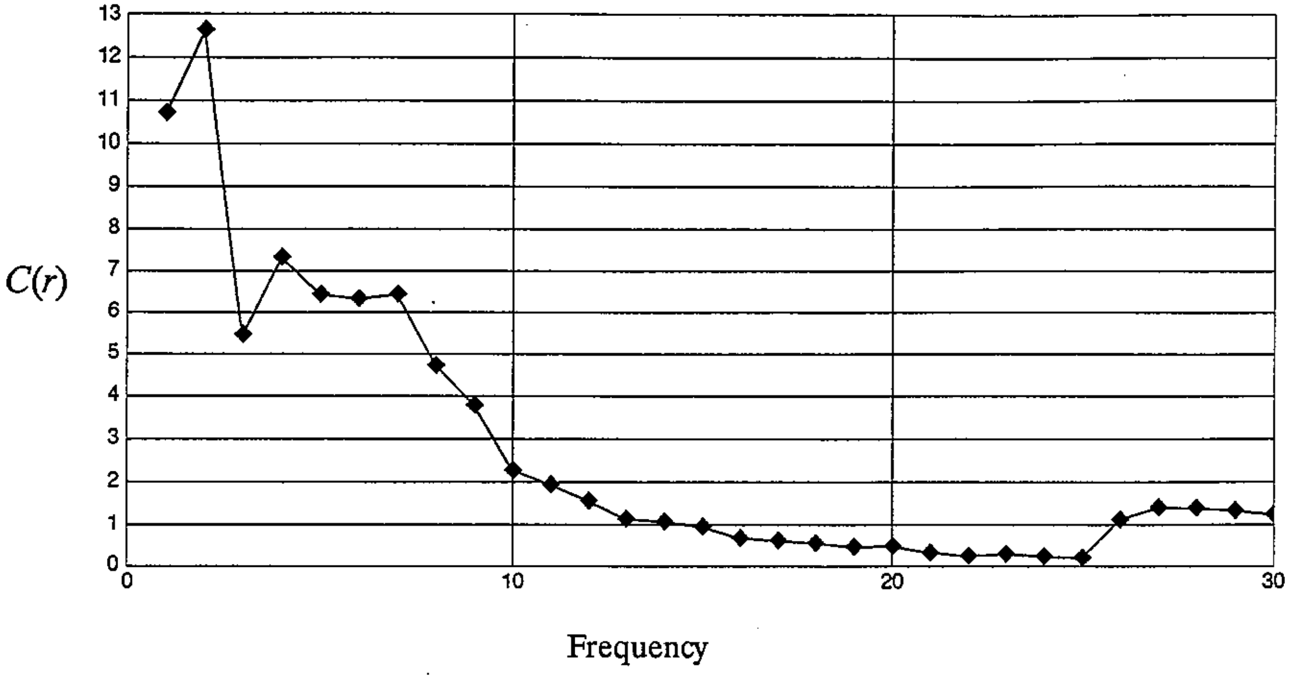

Figure 4–3: The curve C(r) of an image patch without multiplying weighted function

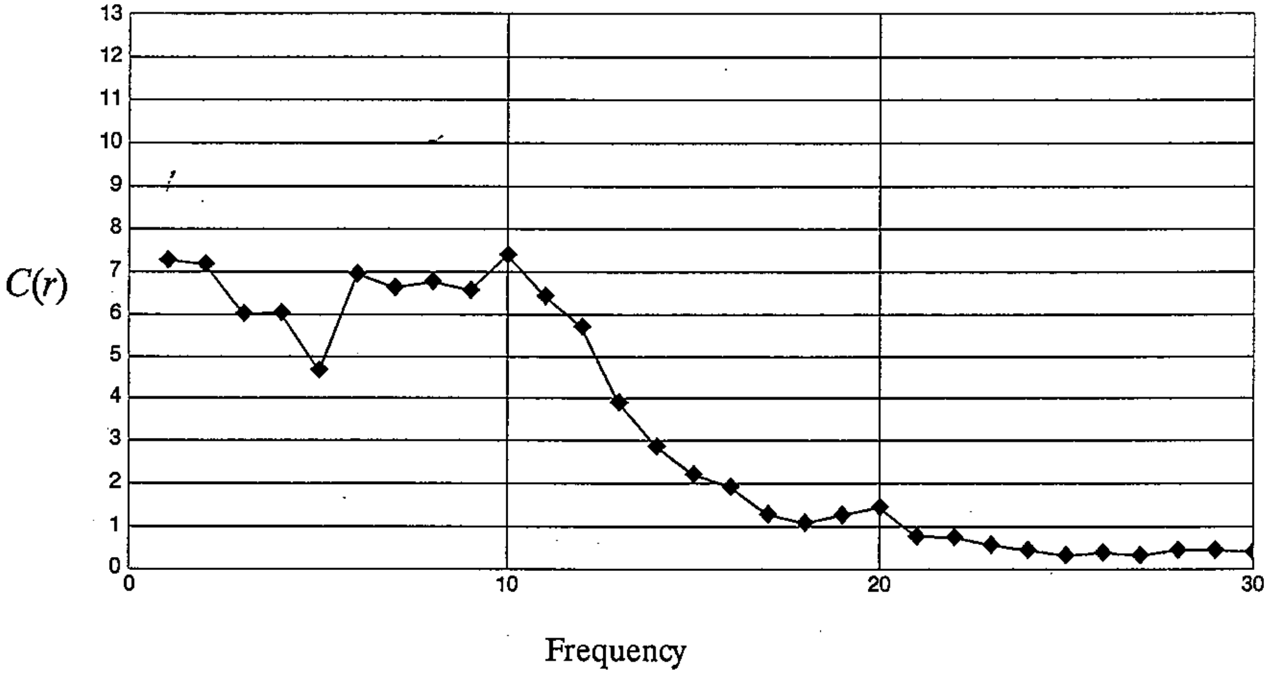

Figure 4–4: The curve C(r) of an image patch multiplied with weighted function

4.3 Asymmetric Laplacian template in the image domain approach

The horizontal and the vertical sampling intervals of our camera system are known to be $T_x =$ 0.001569 cm and $T_y = 0.0013$ cm by static camera calibration. Thus, the sampled points of our digital images are distributed on $(x,y) = (T_x \cdot l_x, T_y \cdot l_y)$ where l_x and l_y are integers. The image–domain approach involves the Laplacian operator. Due to the difference in the resolutions along the horizontal and the vertical directions, the Laplacian template is not symmetric, namely

$$\nabla^2 E(x,y) = \frac{\partial^2 E(x,y)}{\partial x^2} + \frac{\partial^2 E(x,y)}{\partial y^2} = \frac{1}{T_x^2}\frac{\partial^2 E(x,y)}{\partial l_x^2} + \frac{1}{T_y^2}\frac{\partial^2 E(x,y)}{\partial l_y^2}$$

The Laplacian template been used is :

$$\frac{1}{T_x^2}\begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} + \frac{1}{T_y^2}\begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

4.4 Experimental results

Figure 4–5 shows two defocused images of a same scene taken by the camera with setting 1 and 3, respectively. The scene consists of a book and a sheet of newspaper. The book on the left side is 100 cm far from the camera while the newspaper on the right side is 140 cm away. Figure 4–6 shows the resultant depth–map using the frequency–domain approach. Figure 4–7 shows the depth–map of the scene in figure 4–5 generated by first finding the depth–map using the image–domain approach, then averaging the depths of all points in each image patch. The size of an image patch is 64 by 64.

Figure 4–8 shows another two defocused images taken by the camera with settings 1 and 3. The scene of the image is an inclined plane covered with a sheet of newspaper which ranges from 105 cm to 145 cm away. Figure 4–9 is the depth–map of this scene obtained by the frequency–domain approach and the size of the FFT is 64 by 64. Figure 4–10 is the smoothed depth–map of the scene in figure 4–8 obtained by the image–domain apprpach.

The scene in figure 4–11 is also an inclined plane covered with a sheet of newspaper which ranges from 105 cm to 145 cm away, but the two defocused images are taken by the camera with settings 1 and 2, respectively. Figure 4–12 is the depth–map of the scene in figure 4–11 obtained by the frequency–domain approach and the size of the FFT is 64 by 64. Figure 4–13 is the depth–map obtained by the image–domain approach.

Figure 4–14 shows two defocused images taken by the camera with settings 1 and 2. The scene consists of two objects. The nearer object is a bottle which is 108 cm from the camera. The farther object is a book which is 145 cm from the camera. Figure 4–15 is the depth–map of the scene in figure 4–14 obtained by the frequency–domain approach and the size of the FFT is 64 by 64. Figure 4–16 are the smoothed depth–map obtained by the image–domain approach.

Experimental results show that the errors of the estimated depth for both the frequency–domain and the image–domain approach are about 10 percent. The frequency–domain approach is time–consuming because Fourier transform must be applied twice for each image patch; whereas the image–domain approach involves only simple Laplacian operator in the spatial domain. Thus, the image–domain approach is faster than the frequency–domain approach.

Also, we found that the estimated depth of a near object is more accurate than that of a far object. The reason is that the farther the object is, the larger the depth of field is. The depth of field is the range of distance over which objects are focused sufficiently well in the sense that the diameter of the blurred circle is smaller than the spatial sampling period of the imaging device. That is, the resolution of the imaging device is not higher enough to distinguish a sharply focused point from its blurred image. Hence, when we calibrate the composite camera parameters a and b, we choose to use the aperture diameter as large as possible such that the depth of field of the camera is as small as possible.

Since the depth recovery of autofocusing algorithm [3] is quite accurate (2.5 percent precision), our combined approaches can be used as a preprocessing stage of the autofocusing algorithm in choosing the initial search interval of the criterion function. They can also provide an initial depth to guide a binocular stereo matching process. see [15].

CHAPTER FIVE

Conclusion

We have introduced two new depth recovery methods by measuring the amount of defocus (blurring) in the image without calibrating the intrinsic camera parameters. These methods is comprised of two phase. The first phase is calibration phase. In this phase, two composite parameters are calibrated, instead of focal length F, f-number f, v_0, and k. The second phase is depth recovery phase. The depth recovery algorithm used in this phase can be either a modified version of Subbarao's frequency–domain approach or Hwang's image–domain approach. Once we have calibrated the composite camera parameters off–line, we can start to recover the depth of arbitrary scene.

The assumptions in our proposed methods is that the defocusing process is linear shift–invariant process and the point spread function is a two dimensional Gaussian function. We have run some experiments to check these requirements and found that our camera characteristics obey these assumptions.

Experimental results indicate that the estimation errors are about 10 percent for our proposed methods when the object is within the distance of 1.5 meters from the camera. But, our methods do not have any assumption about the scene and involve no correspondence problem which has been recognized as the most difficult problem in stereo vision. Our methods fail for "smooth" or "textureless" objects because the frequency–domain approach must have the spatial frequency contents and the image–domain approach involves Laplacian operator. However, we can introduce "texture" by projecting an arbitrary light pattern (e.g. a random dot pattern) onto the surface of the objects.

Our methods can also be used to provide an initial depth to guide a binocular stereo matching process. In addition, they can provide global depth estimates so that the traditional autofocusing algorithms can use these estimates to reduce the time for focusing at some specific point in the scene.

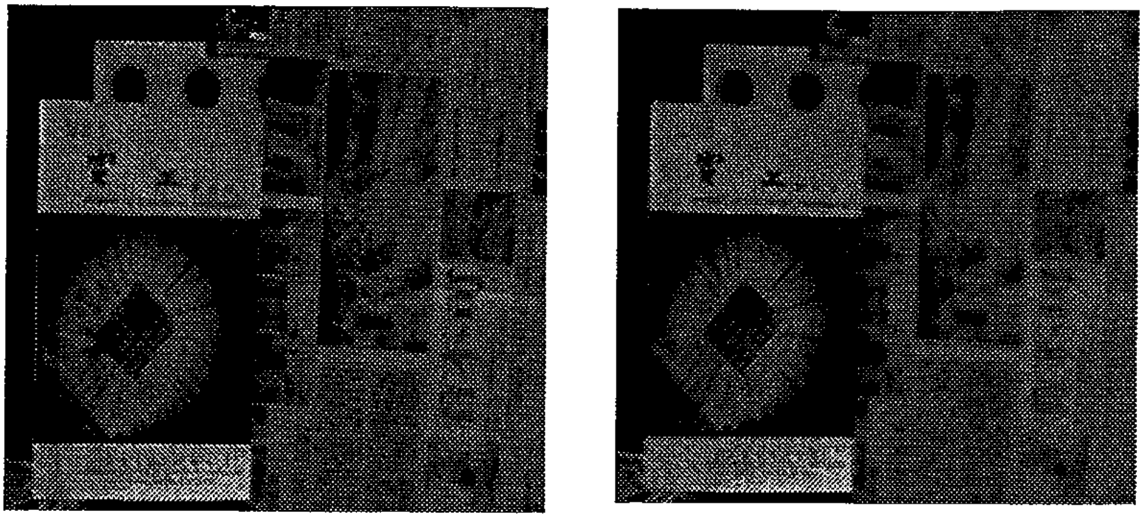

Figure 4–5: Two defocused images taken by the camera with setting 1 and 3. The scene consists of a book of 140 cm and a sheet of newspaper 100 cm away.

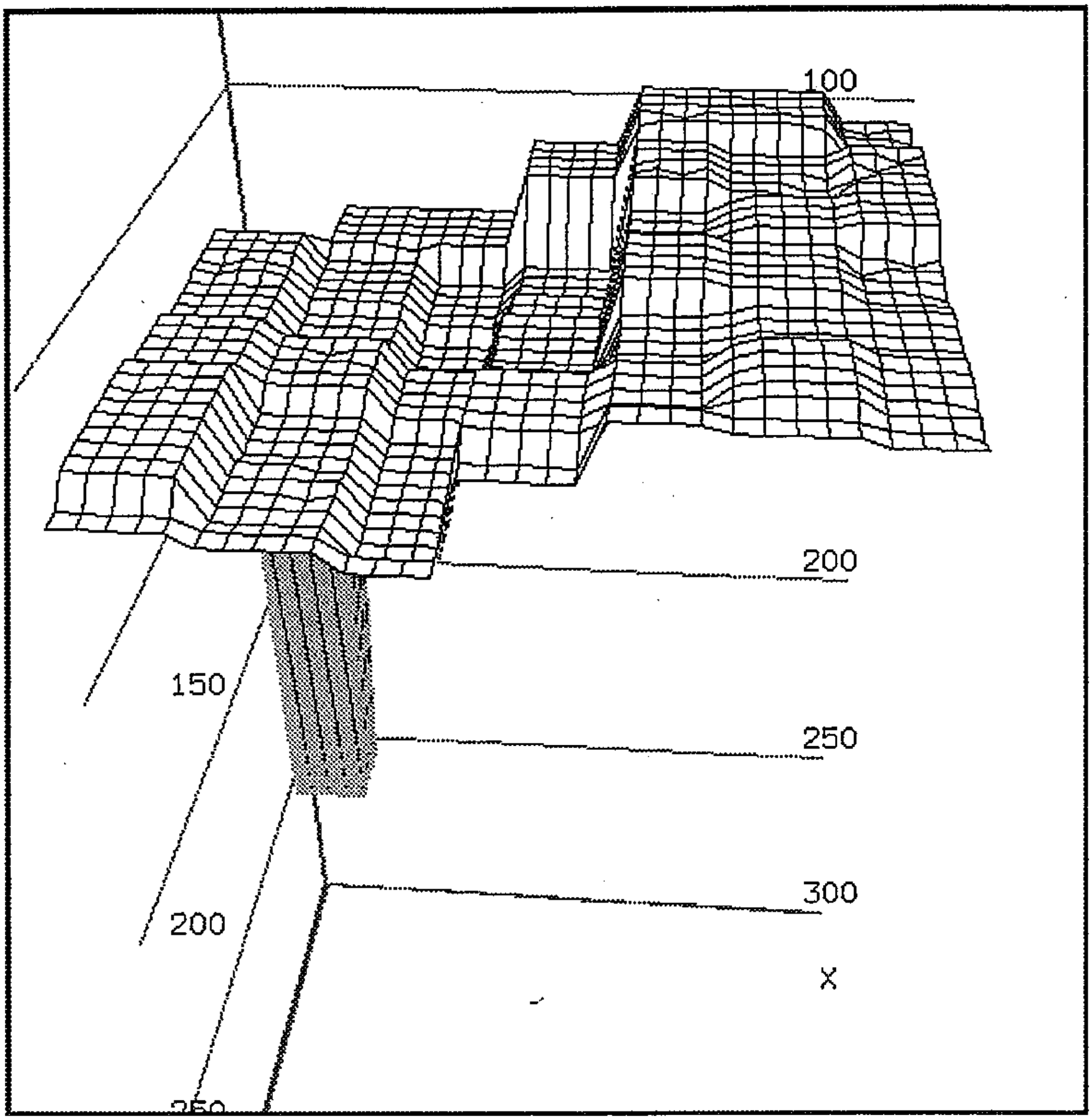

Figure 4–6: The depth–map of the scene in figure 4–5 obtained using the frequency–domain approach.

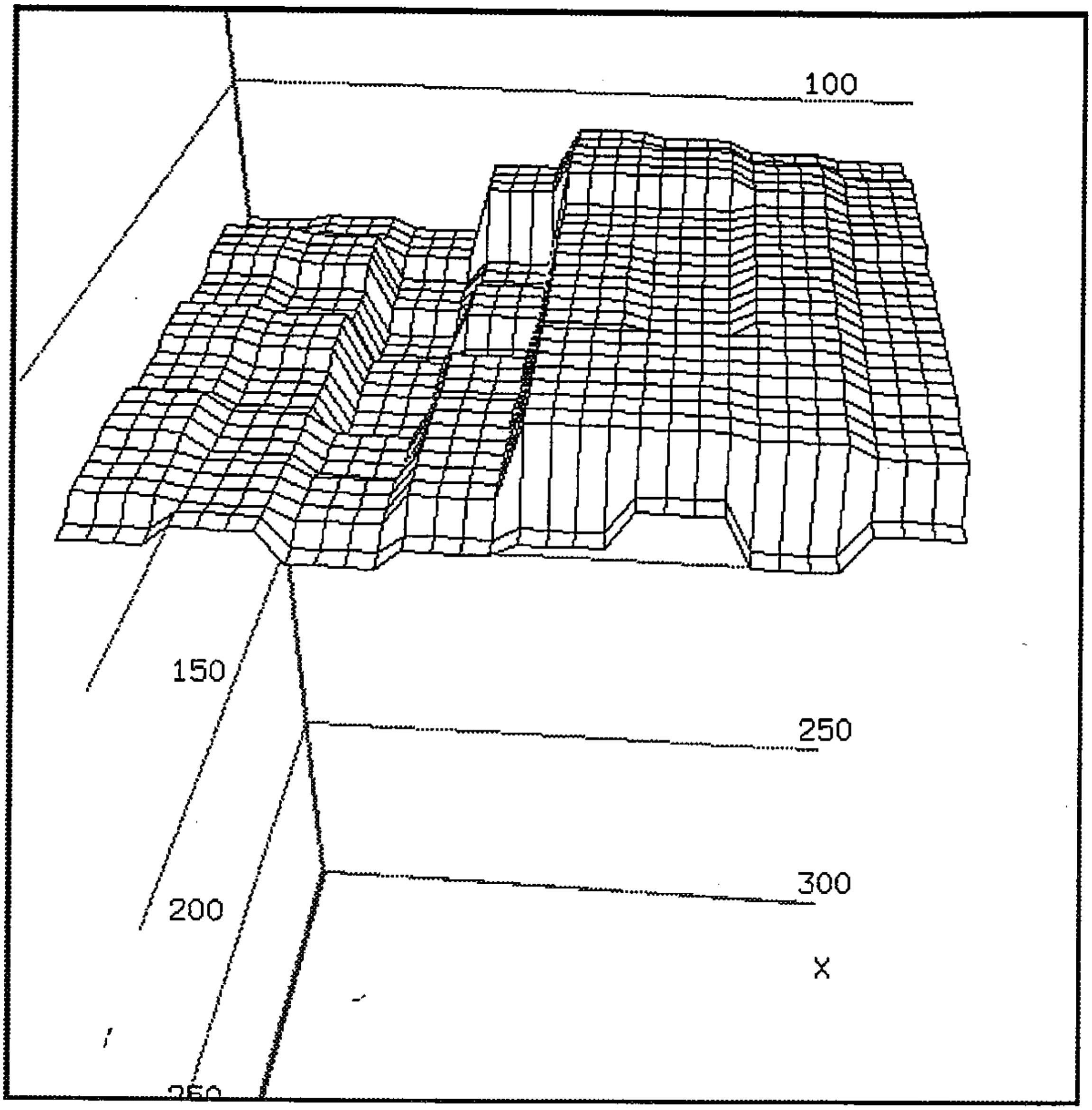

Figure 4–7: The smoothed depth–map of the scene in figure 4–5 using the image–domain approach.

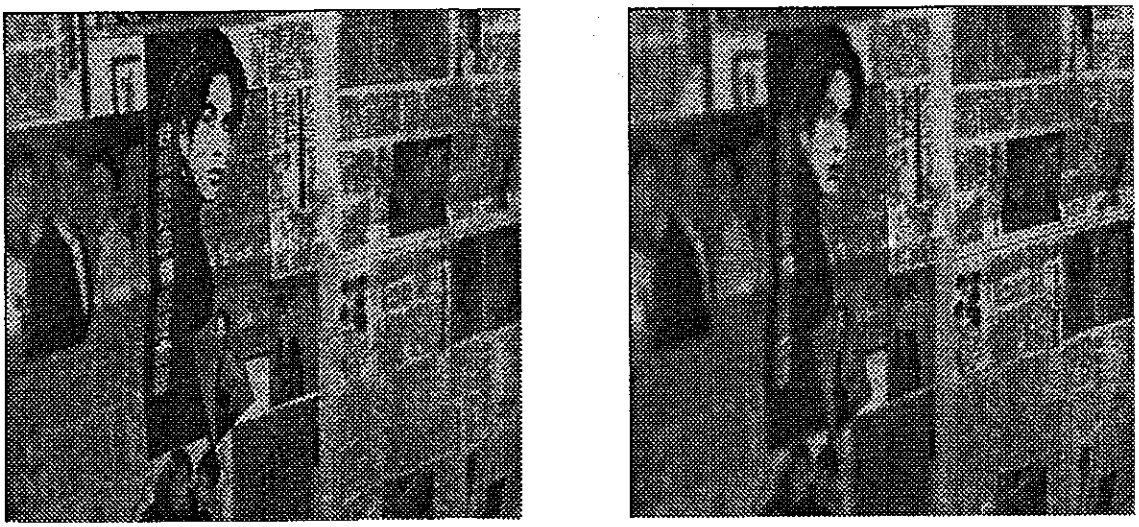

Figure 4–8: Two defocused images taken by the camera with settings 1 and 3. The scene is an inclined plane covered with a sheet of newspaper, distance from 105 cm to 150 cm away.

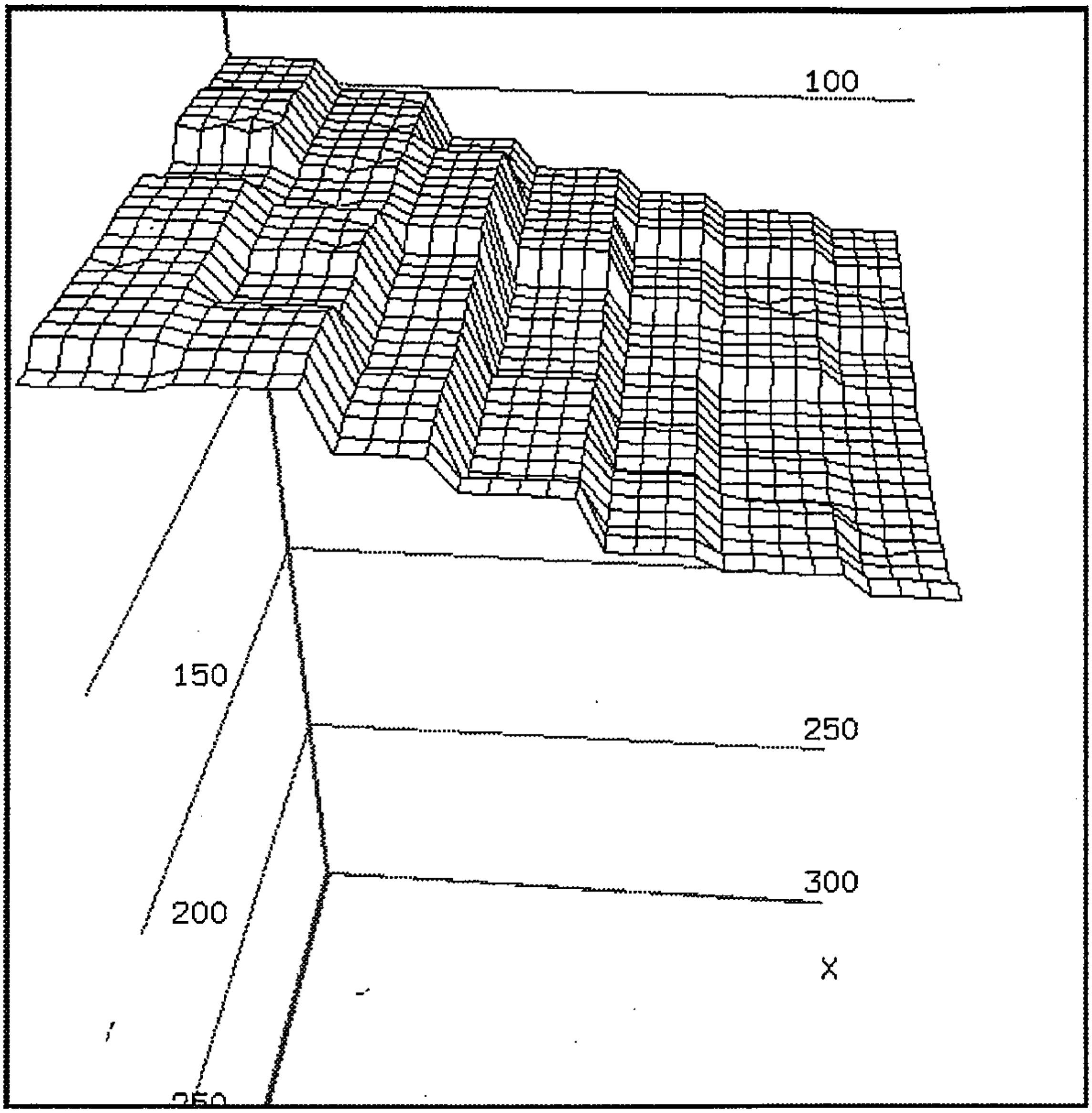

Figure 4–9 : The depth–map of the scene in figure 4–8 obtained using the frequency domain approach.

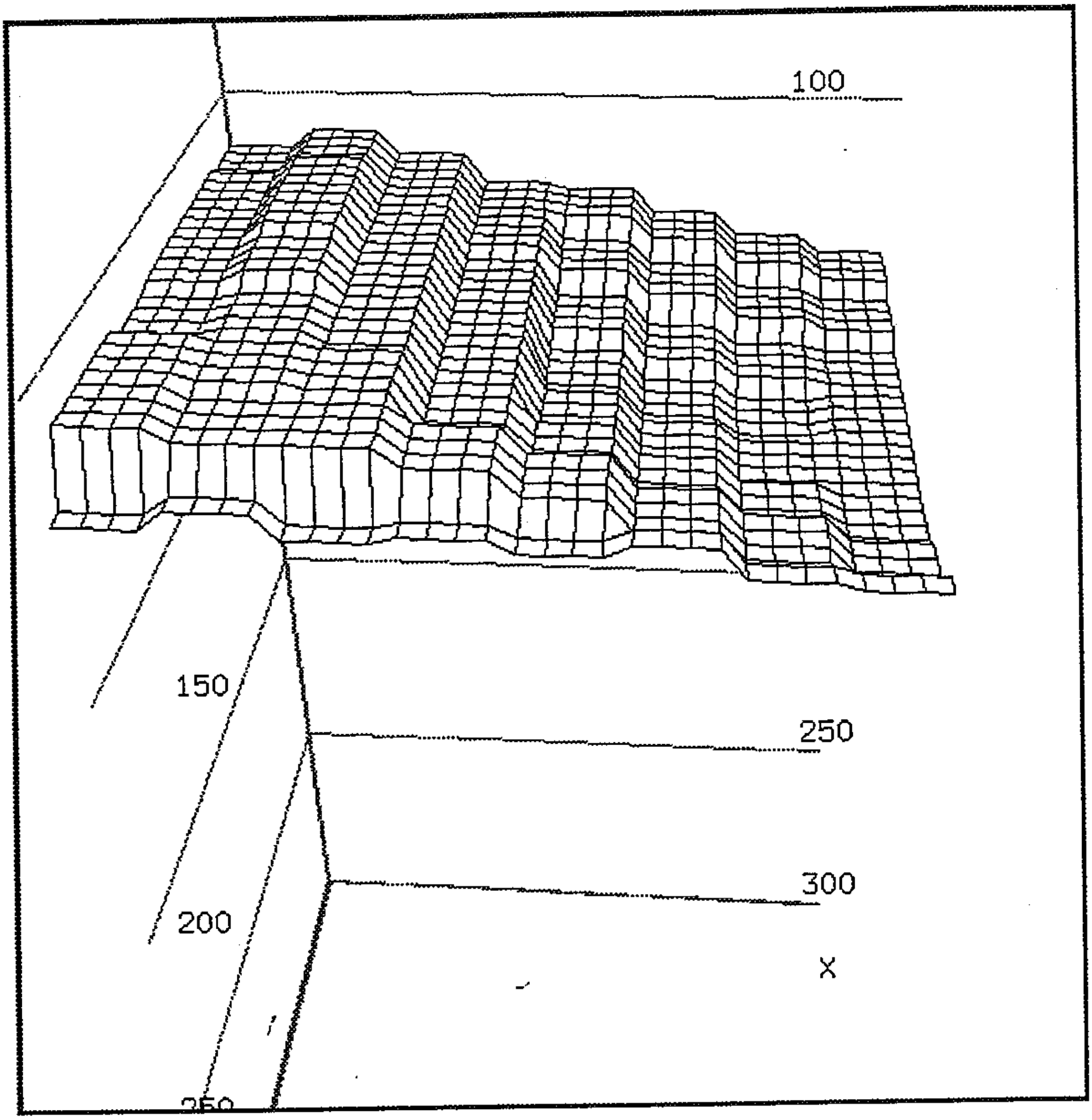

Figure 4–10 : The smoothed depth–map of the scene in figure 4–8 using the image–domain approach.

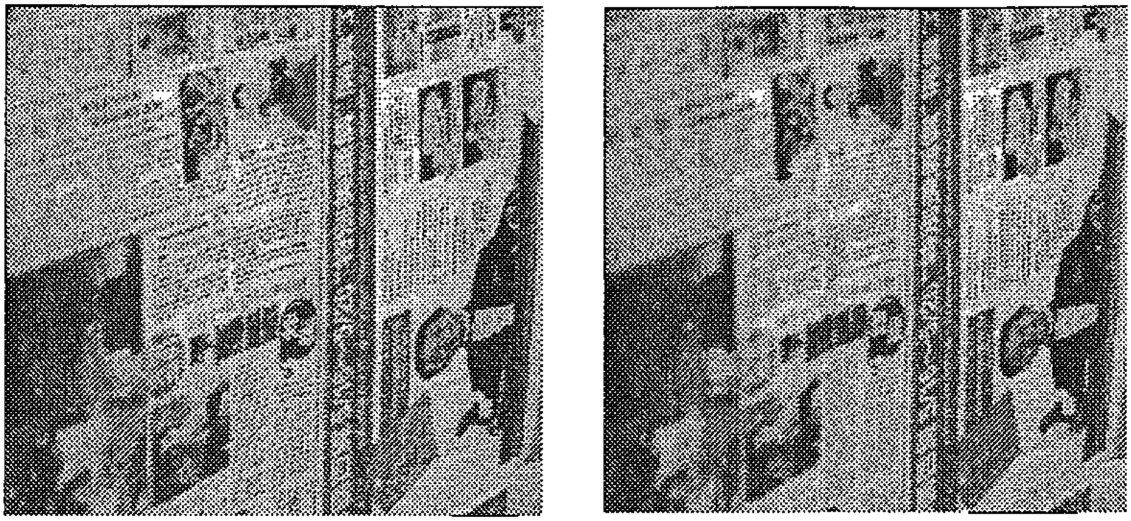

Figure 4–11: Two defocused images taken by the camera with settings 1 and 2. The scene is an incline plane covered with a sheet of newspaper, which ranges from 95 cm to 130 cm.

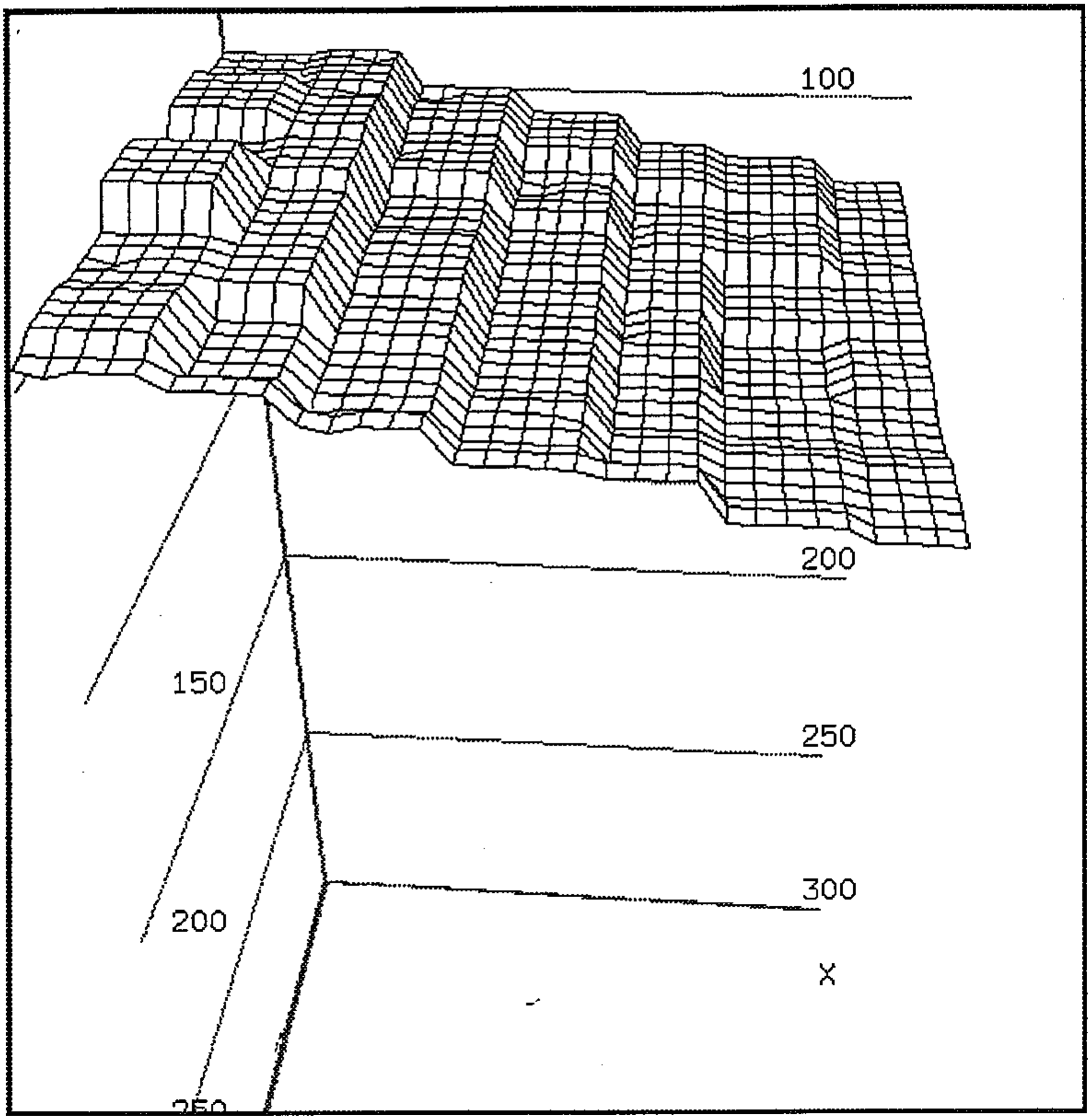

Figure 4–12: The depth–map of the scene in figure 4–11 obtained using the frequency–domain approach.

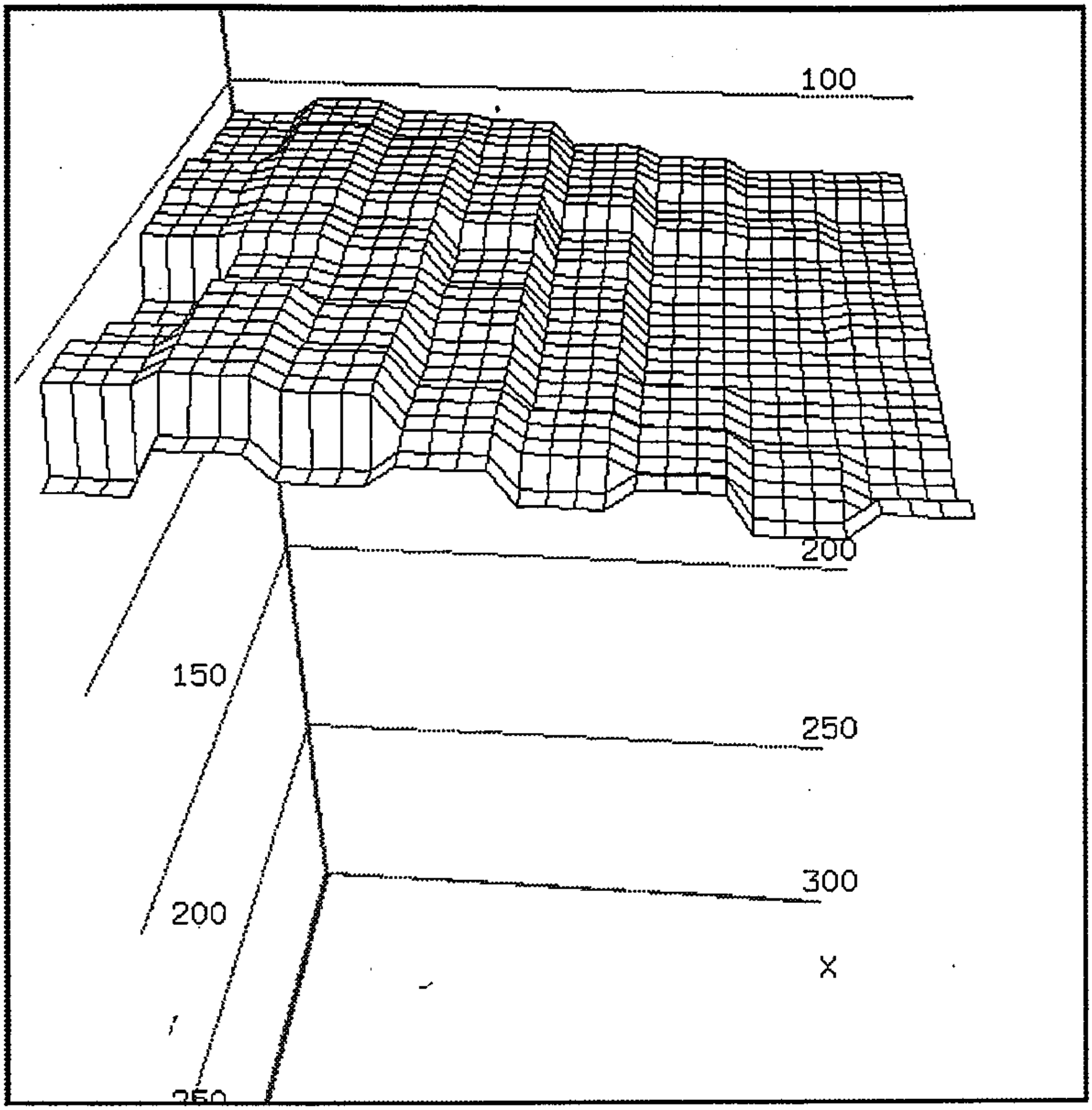

Figure 4–13: The smoothed depth–map of the scene in figure 4–11 using the image–domain approach.

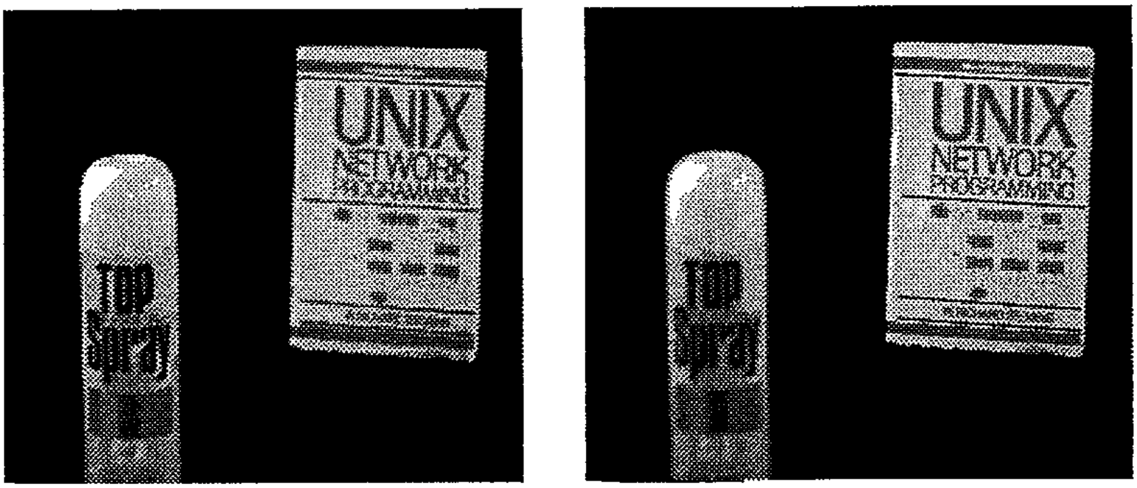

Figure 4–14: Two defocus images taken by the camera with settings 1 and 2. The scene consists of a bottle 108 cm from the camera and a book 145 cm from the camera.

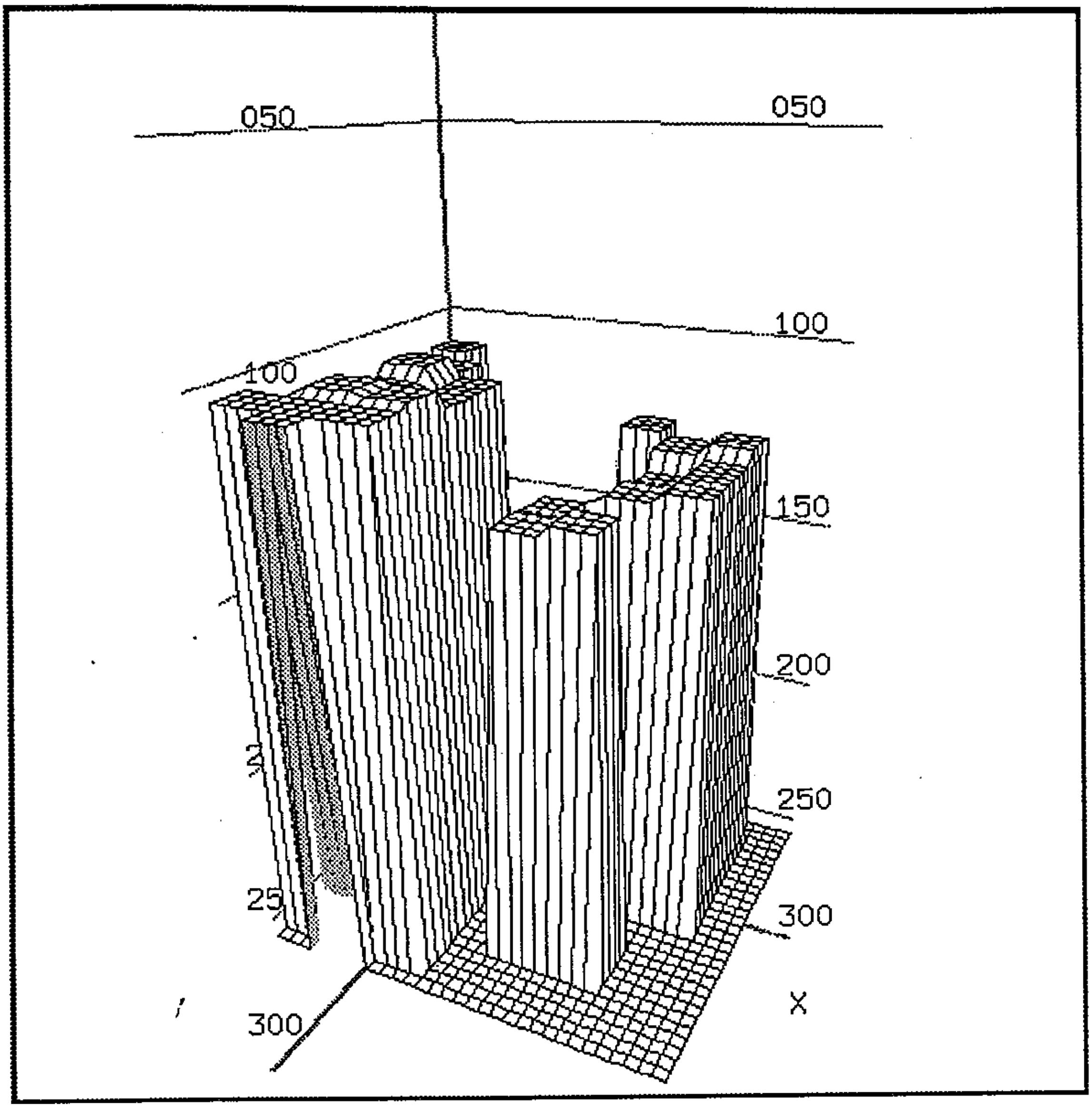

Figure 4–15: The depth–map of the scene in figure 4–14 obtained using the frequency–domain approach.

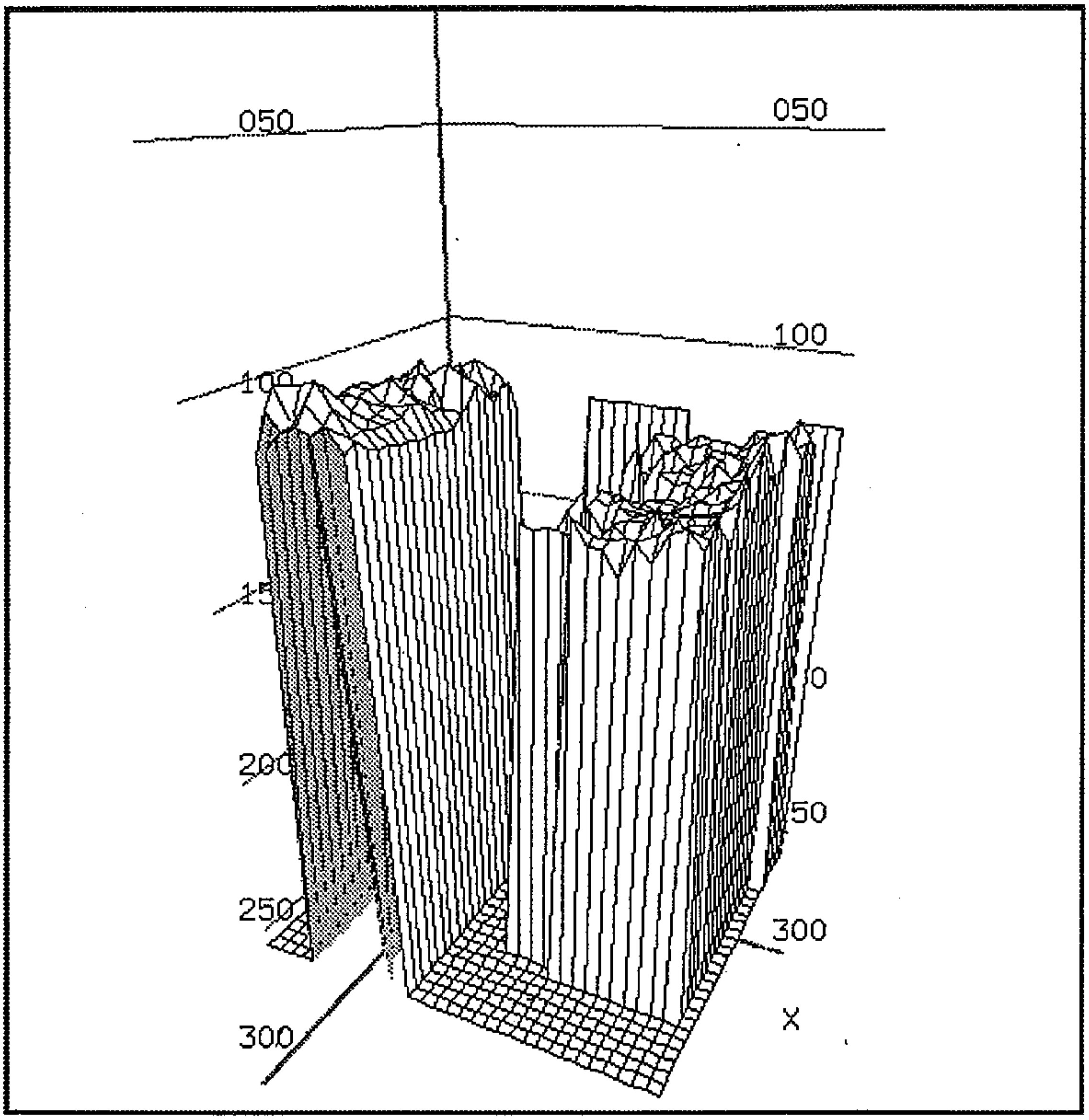

Figure 4–16 : The smoothed depth–map of the scene in figure 4–14 using the image–domain approach.

References

[1] M. Born and E. Wolf, *Principle of Optics*, p434–449, Pergamon Press, 1965.

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

[3] E. Krotkov, " *Focusing* ", International Journal of Computer Vision, vol. 1, pp. 223–237, 1987.

[4] A. Pentland, " *A New Sense for Depth of Field* ", IEEE Transaction on Pattern Analysis and Machine Intelligence,Vol. PAMI–9, No.4, pp.523–531, July 1987.

[5] A. Pentland, T. Darrel, M. Turk and W. Hwang, " A simple, Real–Time Range Camera ", IEEE Conference on Computer Vision and Pattern Recognition, pp. 256–261, 1989.

[6] M. Subbarao, " *Parallel Depth Recovery by Changing Camera Parameters* ", Proceeding of the second International Conference on Computer Vision, pp. 498–503, 1988

[7] M. Subbarao and N. Gurumoorthy, " *Depth Recovery from Blurred Edges* ", IEEE Conference on Computer Vision and Pattern Recognition, pp. 498–503, 1988.

[8] T. L. Hwang, J. J. Clark and A. L. Yuille," *A Depth Recovery Algorithm Using Defocus Information* ", IEEE Conference on Computer Vision and Pattern Recognition, pp. 476–482, 1989.

[9] J. Ens and P. Lawrence, " *A Matrix Based Method For Determining Depth From Focus* ",IEEE Conference on Computer Vision and Pattern Recognition, pp. 600–606, 1991

[10] T. Darrell, " *Pyramid based depth from focus* ", IEEE Conference on Computer Vision and Pattern Recognition, pp. 504–509, 1988.

[11] G. Ligthart and F.C.A. Greon. 1982. " *A Comparison of Different Autofocus Algorithms* ", Proceedings of the International Conference on Pattern Recognition, 1982.

[12] R.A. Jarvis. " *A perspective on range–finding techniques for Computer Vision* ", IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 12–139, 1983.

[13] A.L. Abbott and Ahuja, " *Surface reconstruction by dynamic integration of focus, camera vergence and stereo* ", International Conference on Computer Vision, pp. 532–543, 1988

[14] S.H Lai, C.W Fu and S. Chang, "*A Generalized Depth Estimation Algorithm with a Single Image* ", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 14, No. 4, pp. 405–411, 1992

[15] J.J. Leu, C.J. Tsai, T.P. Hung, C.H. Chen " *Depth Recovery by Integrating Depth–from–Defocus with Stereo* ", to appear in the Proceedings of International Conference on Automation, Robotics and Computer Vision, 1992.