A video-rate range sensor based on depth from defocus

Ovidiu Ghita* and Paul F. Whelan [a]

[a]Vision Systems Laboratory, School of Electronic Engineering, Dublin City University, Dublin 9, Ireland

Recovering the depth information derived from dynamic scenes implies real-time range estimation. This paper addresses the implementation of a bifocal range sensor which estimates the depth by measuring the relative blurring between two images captured with different focal settings. To recover the depth accurately even in cases when the scene is textureless, one possible solution is to project a structured light on the scene. As a consequence, in the scene's spectrum a spatial frequency derived from illumination pattern is evident. The resulting algorithm involves only simple local operations, this assures the possibility of computing the depth at a rate of 10 frames per second. The experimental results indicate that the accuracy of this proposed sensor compares well with that offered by other methods such as stereo and motion parallax, while avoiding the problems caused by occlusion and missing parts.

Keywords: Range sensing; Image blurring model; Active illumination; Focus operator

1. Introduction

Depth information plays a key role in machine vision and has a strong relationship with the real world in robotic applications by allowing 3-D scene interpretation. The 3-D information can be obtained using various techniques [3]. Among other approaches for 3-D estimation, *depth from defocus* (DFD) methods have recently attracted a great deal of interest. Initially developed by Krotkov [4] and Pentland [5], the DFD methods use the direct relationship between the depth, camera parameters and the degree of blurring in several images (usually two as is the case of the present implementation). In contrast with other techniques such as stereo or motion parallax where solving the correspondence between different local features represents a difficult problem, DFD relies only on simple local operators. Also, the DFD techniques are not hampered by problems generated by occlusions or missing parts since the images to be analysed are identical except the depth of filed.

Historically, the DFD techniques have evolved as a passive sensing strategy [1,6,10]. In this regard, Xiong and Shafer [14] propose a novel approach to determine dense and accurate depth structure using the maximal resemblance estimation. Subbarao and Surya

*Corresponding author. Tel.: +353-1-7005869; Fax: +353-1-7005508; E-mail: ghitao@eeng.dcu.ie

[11] reformulate the problem as a one of regularised deconvolution where the depth is obtained by analysing the local information contained in a sequence of images acquired under different camera parameters. Later, Watanabe and Nayar [12] argue that the use of focus operators such as the Laplacian of Gaussian results in poor depth estimation. Thus, they developed a set of broadband rational operators to produce accurate and dense depth maps. However, if the scene under investigation has a weak texture or is textureless, the depth estimation achieved when passive DFD is employed is far from accurate. Fortunately, there is a solution to this problem and is offered by active DFD, when a structured light is projected on the scene. Consequently, a strong texture derived from illumination pattern is forced on imaged surfaces and as an immediate result the spectrum will contain a dominant frequency. The use of active illumination was initially suggested by Pentland *et al* [7] where the apparent blurring of a pattern generated by a slide projector is measured to obtain the depth information. Then, Nayar *et al* [8] developed a symmetrical pattern organised as a rectangular grid which was optimised for a specific camera.

In this paper we describe the implementation of a bifocal sensor able to generate 256 × 256 depth maps at a rate of 10 frames per second. In the past, the performance of the DFD range sensors was limited by the variation in image magnification between images acquired under different focal levels. This forced the researchers to either implement computationally intensive techniques such as image registration and warping [2] or to address this problem on an optical basis by using a telecentric lens [13]. Another difficult problem consists of choosing the illumination pattern. Nevertheless, an optimal pattern is difficult to manufacture and in this paper we proposed a simple solution to address the abovementioned problems by employing image interpolation.

2. Depth from defocus

If the object to be imaged is placed in the focal plane, the image formed on the sensing element is sharp since every point P from the object plane is refracted by the lens into a point p on the sensor plane. Alternatively, if the object is shifted from the focal plane, the points situated in the object plane are distributed over a patch on the sensing element. As a consequence, the image formed on the sensing element is blurred. From this observation, the distance from the sensor to each object point can be estimated by evaluating the degree of blurring which is determined by the size of the patch formed on the sensing element.

This can be observed in Figure 1 where the image formation process is illustrated. Thus the diameter of the patch (blur circle) d is of interest and can be easily determined by the use of similar triangles:

$$\frac{D/2}{v} = \frac{d/2}{s-v} \Longrightarrow d = Ds(\frac{1}{v} - \frac{1}{s}) \tag{1}$$

where v is the focal distance, D is the aperture of the lens and s is the sensor distance. Because the parameter v can be expressed as a function of the focal length f and the object distance u, Equation 1 becomes:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \Longrightarrow d = Ds(\frac{1}{f} - \frac{1}{u} - \frac{1}{s}) \tag{2}$$

Note that d can be positive or negative depending on whether the image plane is behind or in front of the focused image. Consequently, the level of blurring and the resulting images are identical. To overcome this problem, we either have to constrain the sensor distance s to be always greater than the image distance v [5] or to employ two images captured with different focal settings [6,10,12]. It is worth mentioning that for the former case the depth can be determined accurately and uniquely only for places in the image with known characteristics (e.g. sharp edges). The latter is not hampered by this restriction and for this present approach a pair of images, i.e. the near and far focused images, separated by a known distance b are employed to determine the blur circle.

3. The blur model

The blurring effect can be seen as a convolution between the focused image and the blurring function. The blurring function is also referred to as *point spread function* (PSF) and can be approximated by a two-dimensional Gaussian [5,11].

$$h(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

where σ is the standard deviation of the Gaussian. The standard deviation σ (also referred to as spread parameter) is of interest because it indicates the local level of blurring contained in a defocused image. If the brightness is constant over a region of the image projected on the sensing element, the transformation from focused image into defocused image is a linear shift invariant operation. This represents the actual situation and as a consequence we can assume that the blur parameter σ is proportional with d.

4. Estimating the depth of the scene

Since the PSF is a low pass filter, the effect is a suppression of high frequencies. Therefore, to isolate the effect of blurring it is necessary to extract the high frequency information derived from the scene. For this purpose, the near and far focused images are filtered with a 5×5 Laplacian operator, where the output of this operator gives an indication of the focus level. Figure 2 illustrates the relation between the outputs of the focus operator and the depth estimation.

5. Active illumination

The high frequencies derived from the scene determine the accuracy of the depth estimation. If the scene has a weak texture or is textureless (like a blank sheet of plain paper) the depth recovery is very far from accurate. Consequently, the applicability of passive DFD is restricted to scenes with high textural information. To address this limitation, Pentland *et al* [7] proposed to project a known pattern of light on the scene. As a result, an artificial texture is forced on the visible surfaces and the depth can be obtained by measuring the apparent blurring of the projected pattern. The illumination pattern was generated by a slide projector and selected in an arbitrary manner. In Figure 3d it can be observed the textural frequency derived from an illumination pattern organised as a striped grid.

Later, Noguchi and Nayar [9] developed a symmetrical pattern optimised for a specific camera. They use the assumption that the sensing element is organised as a rectangular grid. The optimisation procedure presented in the same paper, consists of a detailed Fourier analysis and the resulting model of illumination is a rectangular cell with uniform intensity which is repeated on a two dimensional grid to obtain a periodic pattern. The resulting pattern is very dense and difficult to fabricate and in our testing we found by use of image interpolation the problems caused by a sub-optimal pattern can be significantly alleviated. This issue will be presented in the next section.

6. Image interpolation

Since active illumination is employed, the depth estimation will have the same pattern. Due to magnification changes between the near and far focused image the stripes does not match perfectly together. As a consequence, the depth estimation is unreliable especially around the stripes' borders. This can be observed in Figure 4 where the depth estimation is not continuous. Also, note the errors caused by changes in magnification.

To compensate for this problem, Watanabe and Nayar [13] proposed to use a telecentric lens. This solution is elegant and effective but since the telecentric lens requires a small external aperture, the illumination source necessary to image the scene has to be very powerful. To avoid this complication we propose to map the dark regions by using image interpolation. Linear interpolation was found to be sufficient in the case where a dense (10 lines per mm) illumination pattern was used. The effect of image interpolation is depicted in Figure 5 where the quality of the depth estimation is significantly improved.

7. Physical implementation

The aim of this implementation is to build a range sensor able to extract the depth information derived from dynamic scenes. Thus, a key issue is to capture the near and far focused images at the same time. For this purpose, two OFG VISION*plus* frame grabbers were utilised. The scene is imaged using an AF MICRO NIKKOR 60 mm F 2.8 lens. Between the NIKKOR lens and the sensing elements a 22 mm beam splitter cube is placed. The sensing elements used for this implementation are two low cost 256 × 256 VVL 1011C CMOS sensors. Nevertheless, the beam splitter introduced a supplementary distance between the CMOS sensors and the lens that image the scene. As a consequence, the images projected on the sensors' active surface will be significantly out of focus. To overcome this problem, the camera head was opened and the first sensor was set in contact with the beam splitter inside the case. The second sensor was positioned with a small gap (approximately 0.8 mm) from the beam splitter surface using a multi-axis translator. The distance between the lens and the beam splitter is about 1mm. The diagram of the developed sensor is illustrated in Figure 6, while Figure 7 depicts the actual set-up.

The structured light is projected on the scene using a MP-1000 Projector fitted with a MGP-10 Moire gratings (stripes with density of 10 lines per mm). The lens attached to the projector is the same type as that used to image the scene. Note that all equipment required by this implementation is low cost.

The software is simple as long as it includes only local operators. The flowchart illustrated in Figure 8 describes the main operations required to compute the depth map of a

256×256 resolution.

As we mentioned earlier, the focus operator is modelled by a 5×5 Laplacian operator. Since the Laplacian operator enhances the high frequency noise, to compensate for this issue a smoothing 5×5 Gaussian operator is applied. The information resulted after image interpolation is used to determine the depth map. As illustrated in Figure 2 the defocus function can be implemented by the ratio between $\nabla^2 g_1$ and $\nabla^2 g_2$. For the sake of computation efficiency, the defocus function is implemented using a look-up table. Finally, the depth map is smoothed with a 5×5 Gaussian operator. The depth map is computed in 95 ms on a Pentium 133, 32 Mb RAM and running Windows 98.

8. Camera calibration

Like for any other range sensor, the calibration procedure represents an important operation. This sensor requires a two-step calibration procedure. The first step involves obtaining a precise alignment between the near and far focused sensing elements. To achieve this goal, the calibration is performed step by step using the multi-axis translator which is attached to one of the CMOS sensors. This procedure continues until the mis-registrations between the near and far focused images are smaller than the errors caused by changes in magnification due to different focal settings. As suggested in [8], the second step carries out a pixel by pixel calibration in order to compensate for the errors caused by the imperfection of the optical and sensing equipment. This operation consists of the following procedure: a planar target is perpendicularly placed to the optical axis of the sensor at a precise known distance. Then, the depth map is computed and the difference between the resulting depth values and the real distances are recorded in a table which defines the offset map. The pixel by pixel calibration consists of subtracting the depth offset values from the detected depth map. Nevertheless, this procedure holds only if the errors introduced by the optical equipment are constant. Our experiments indicated that most of the errors are caused by image curvature, errors that are constant and easy to correct. Consequently, this pixel by pixel gain calibration is appropriate for this implementation.

9. Experiments and results

To evaluate the performance of the developed sensor, it was tested on several indoor scenes. Initially, the sensor was tested on simple targets like planar surfaces, then on scenes with a complex scenario. The accuracy and linearity is estimated when the sensor is placed at a distance of 86 cm from the baseline of the workspace.

Figure 11 illustrates the depth recovery for two textured planar objects (see Figure 10) situated at different distances in front of the sensor. Figure 13 shows the depth map for a slanted planar object illustrated in Figure 12.

Figure 15 shows the depth recovery for a more complex scene defined by various textureless objects and Figure 17 illustrates the depth map for a scene which contains mildly specular LEGO objects with different shapes and a large scale of colours.

For scenes containing non-specular objects the lowest accuracy (see Figure 9) is 3.4% of the overall ranging distance from the sensor. This accuracy is reported for both textured and textureless objects. When the scene contains objects with specular properties the accuracy is affected in relation to the degree of specularity. These results indicate that the developed sensor may provide enough information for a large variety of visual applications including object recognition and advanced inspection.

10. Conclusions

In this paper we described the implementation of a video-rate bifocal range sensor. Since the depth is estimated by measuring the relative blurring between two images captured with different focal settings, this approach in contrast with methods such as stereo and motion parallax is not hampered by problems such as detecting the correspondence between features contained in a sequence of images or missing parts. Also it is worth noting that DFD techniques offer the possibility of obtaining real-time depth estimation at a low cost.

Active DFD is preferred in many applications because it can reliably estimate the depth even in cases when the scene is textureless. The changes in magnification between images captured with different focal settings limited the performance of this technique and forced the researchers to resort to computationally expensive techniques such as image registration and warping. Nevertheless, such implementation is not appropriate for real-time depth estimation, hence we propose to tackle this problem by employing image interpolation. Another advantage is the fact that this approach relaxes the demand of having an optimal illumination pattern. The consistency between theory and experimental results has indicated that our implementation offers an attractive solution to estimate depth quickly and accurately.

REFERENCES

1. Asada N., Fujiwara H. and Matsuyama T., Seeing behind the scene: analysis photometric properties of occluding edges by the reversed projection blurring model, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 1998; 20(2):157-166.
2. Darrell T. and Wohn K., Depth from defocus using a pyramid architecture, Pattern Recognition Letters, 1990; 11(12):787-796.
3. Jarvis R., A perspective on range-finding techniques for computer vision, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 1983; 5(2):122-139.
4. Krotkov E., Focusing, Intl. Journal of Computer Vision, 1987; vol. 1:223-237.
5. Pentland A., A new sense for depth of field, IEEE Trans. on Pattern Analysis and Machine Inteligence (PAMI), 1987; 9(4):523-531.
6. Pentland A., Darrell T., Turk M. and Huang W., A simple, real-time range camera, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 1989; p. 256-261.
7. Pentland A., Scherock S., Darrell T. and Girod B., Simple range cameras based on focal error, Journal of Opt. Soc. of America, 1994; 11(11):2925-2935.
8. Nayar S. K., Watanabe M. and Noguchi M., Real-time focus range sensor, Proc. of Intl. Conf. on Computer Vision, 1995; p. 995-1001.
9. Noguchi M. and Nayar S.K., Real-time focus range sensor, Proc. of International Conference on Pattern Recognition (ICPR 94), October 1994.

10. Subbarao M., Parallel depth recovery by changing camera parameters, Proc. of the IEEE Conf. on Computer Vision, 1988; p. 149-155.
11. Subbarao M. and Surya G., Depth from defocus: a spatial domain approach, Intl. Journal of Computer Vision, 1994; 13(3):271-294.
12. Watanabe M. and Nayar S. K., Rational filters for passive depth from defocus, Technical Report CUCS-035-95, Dept. of Computer Science, Columbia University, New York, USA, September 1995.
13. Watanabe M. and Nayar S. K., Telecentric optics for computational vision, Proc. of Image Understanding Workshop (IUW 96), Palm Springs, February 1996.
14. Xiong Y. and Shafer S. A., Depth from focusing and defocusing, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 1993; p. 68-73.

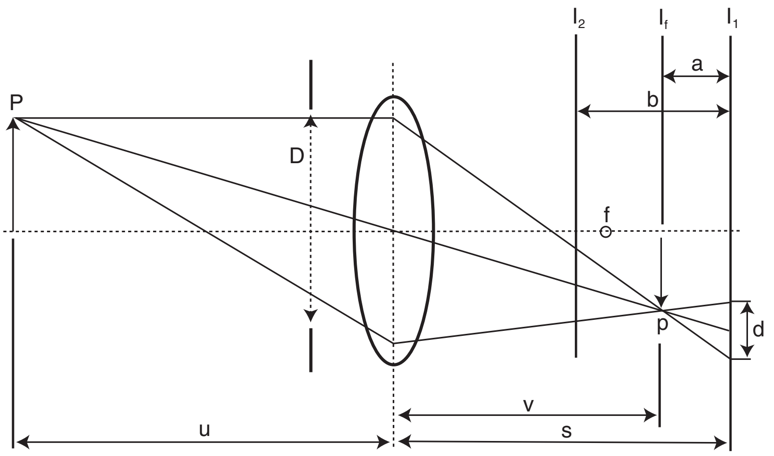


Figure 1. The image formation process. The depth can be determined by measuring the level of blurring.

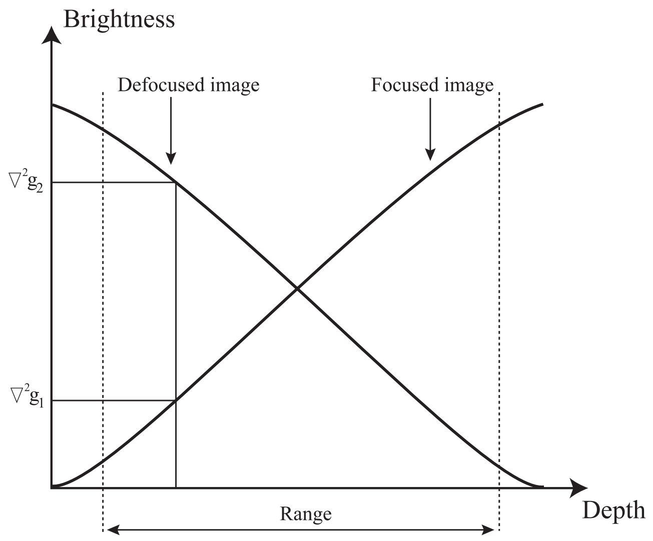
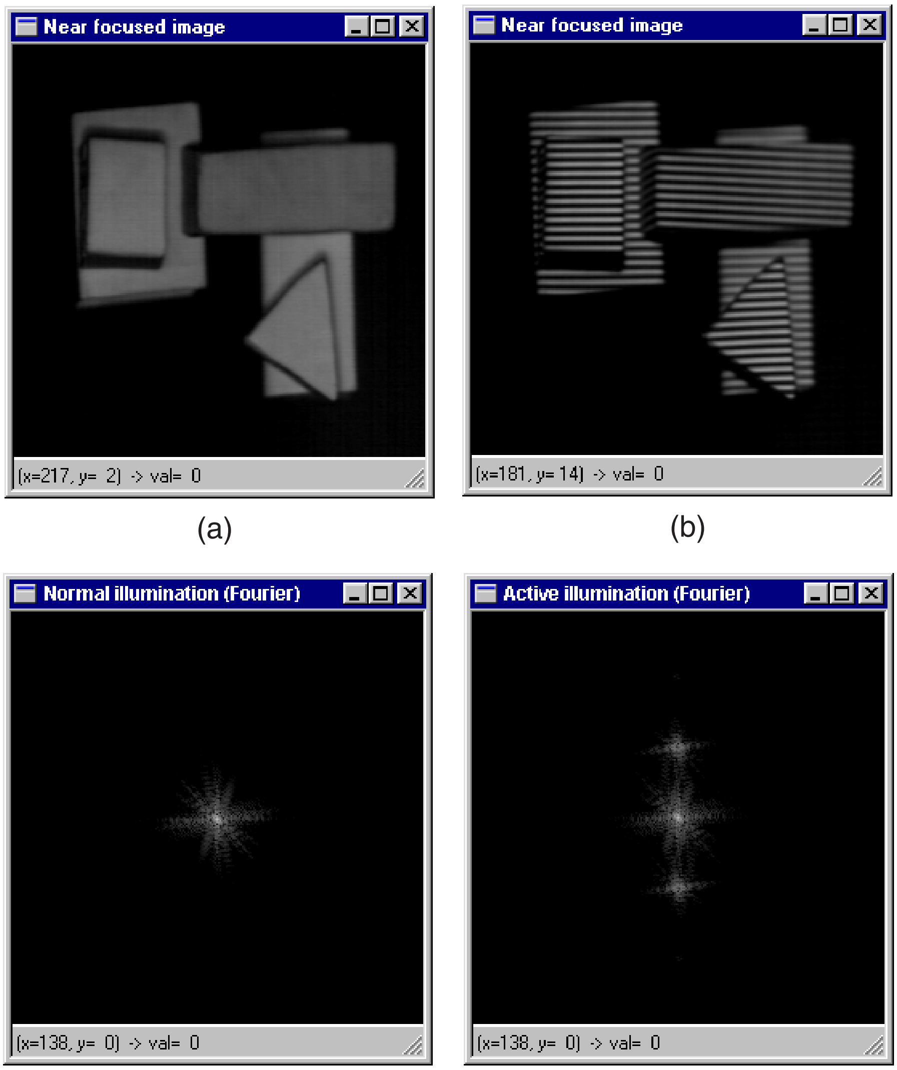


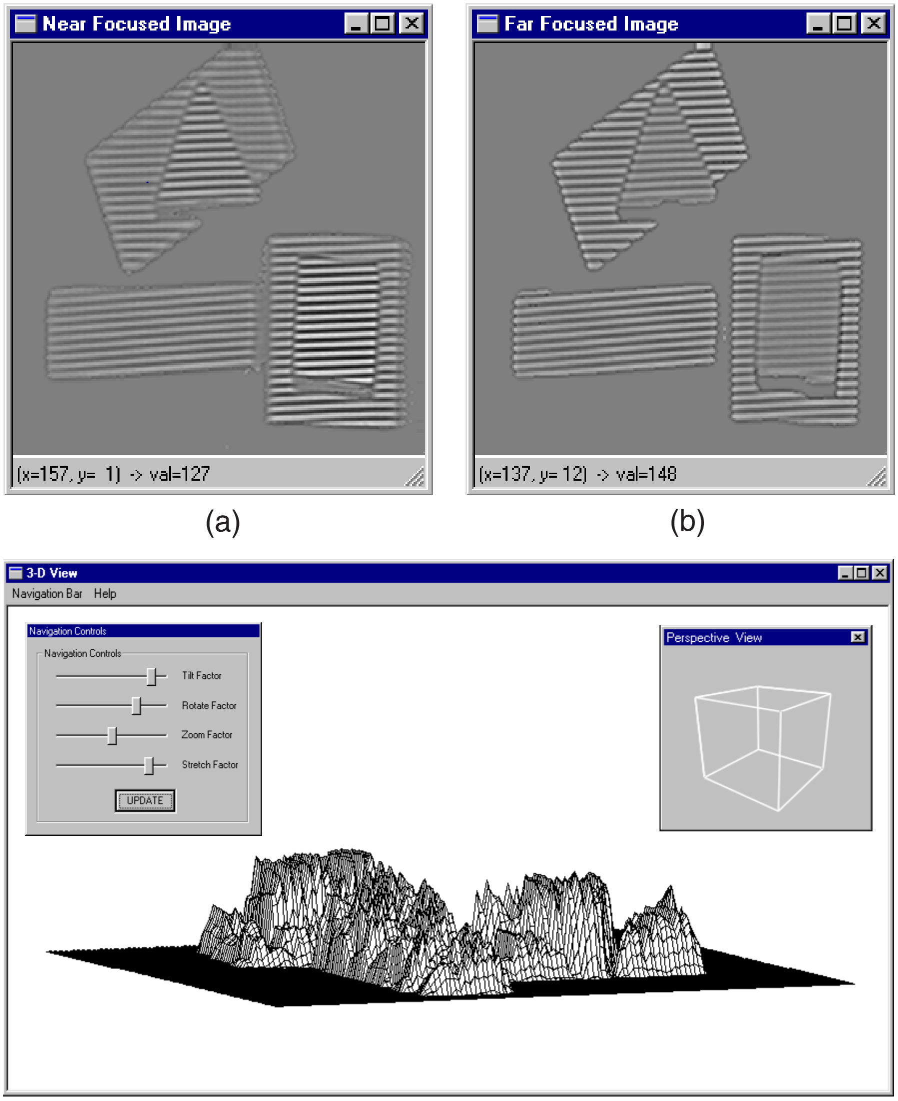
Figure 2. Estimating the depth from two images.



(a) (b)

(c) (d)

Figure 3. Normal illumination versus active illumination. (a) Image captured using normal illumination. (b) Image captured when active illumination is employed. (c) The Fourier spectrum of image (a). (d) The Fourier spectrum of image (b).

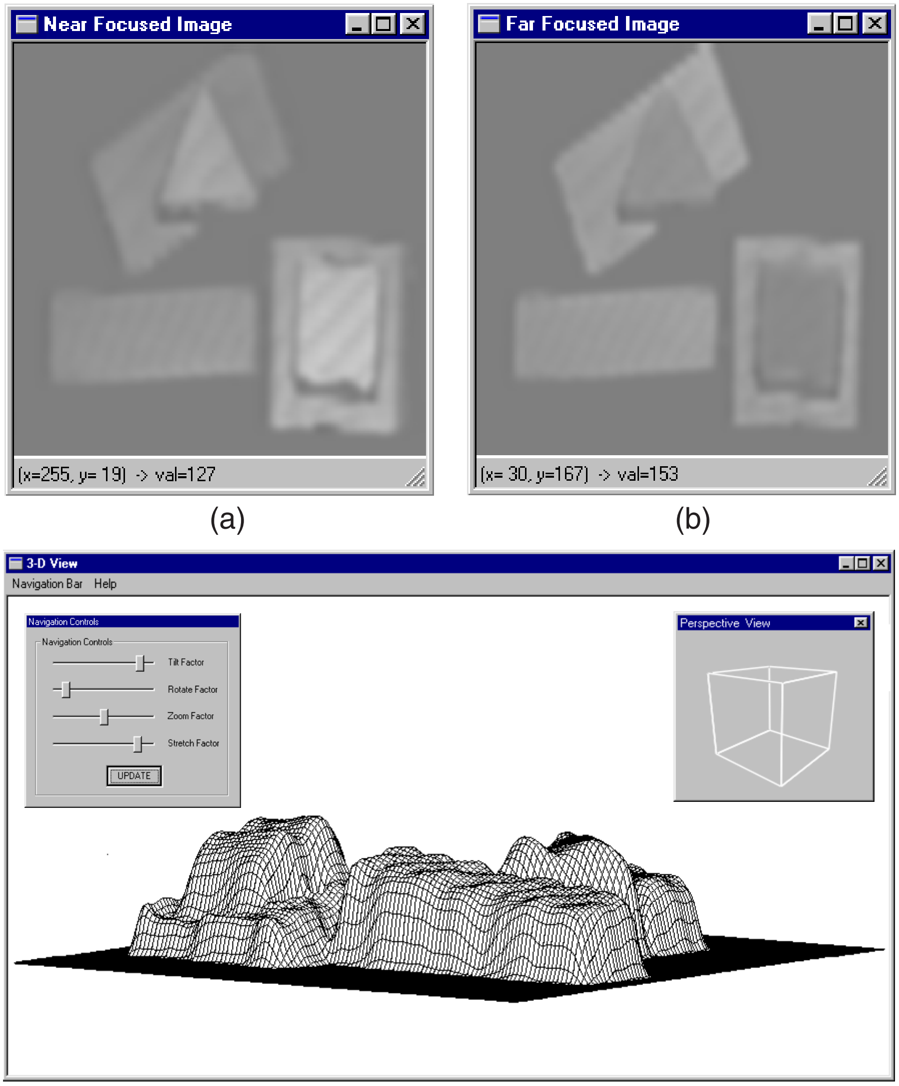


(a)

(b)

(c)

Figure 4. (a,b) The near and far focused images after the application of the focus operator. (c) The resulting depth map.



(x=255, y= 19) -> val=127

(x= 30, y=167) -> val=153

(a)

(b)

(c)

Figure 5. (a,b) The effect of interpolation when this operation is applied to the images illustrated in Figure 4(a,b). (c) The resulting depth map.

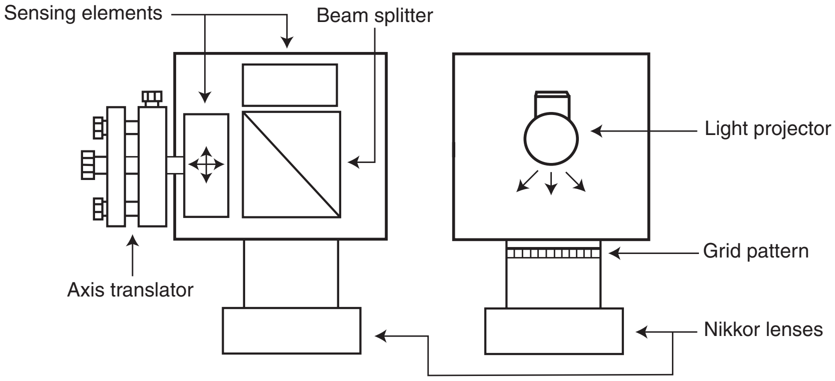
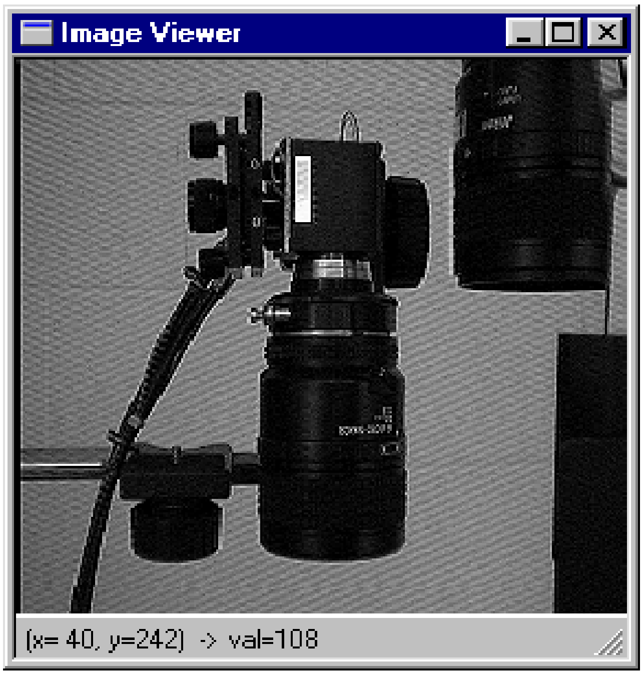


Figure 6. The diagram of the bifocal range sensor.



(x= 40, y=242) -> val=108

Figure 7. A picture of the developed range sensor.

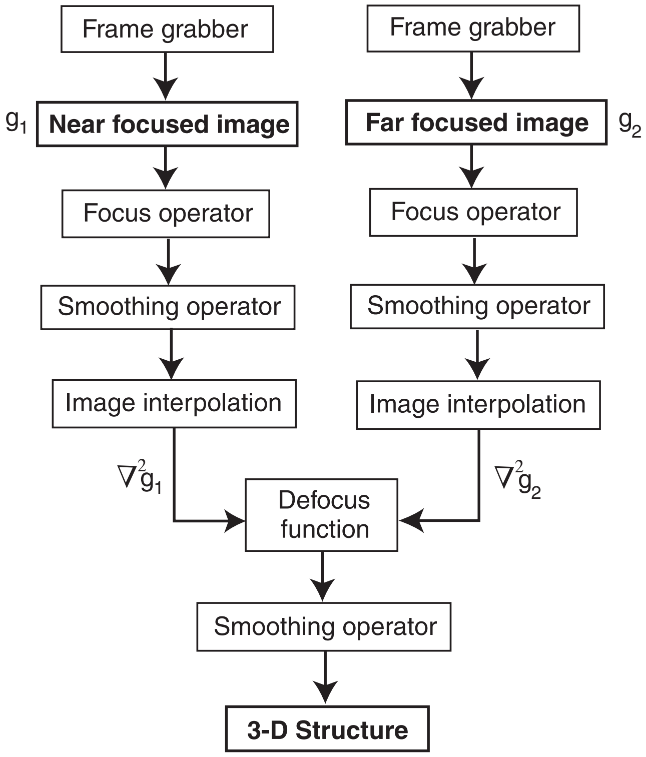


Figure 8. Data flow during the computation process

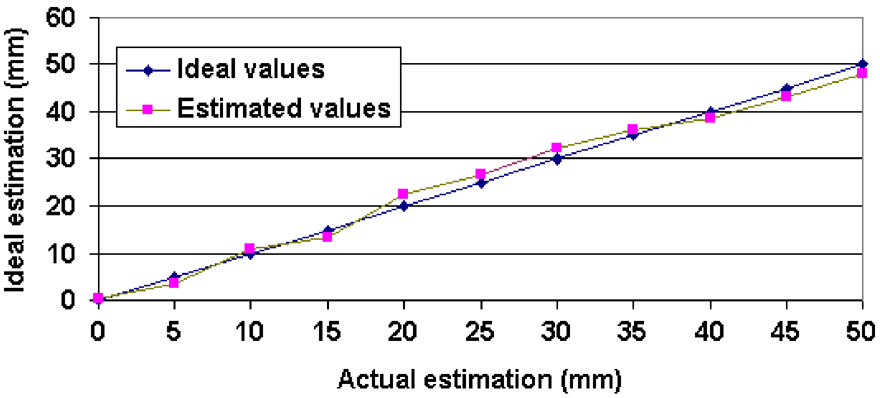
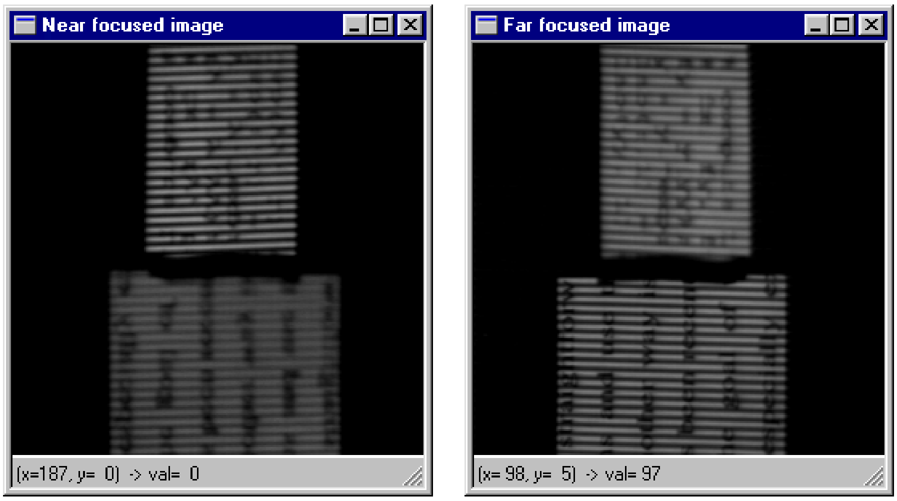


Figure 9. Ideal depth versus estimated depth.



(x=187, y= 0) -> val= 0

(x= 98, y= 5) -> val= 97

Figure 10. The near and far focused images for a scene which contains two planar objects situated at different distances from the sensor.

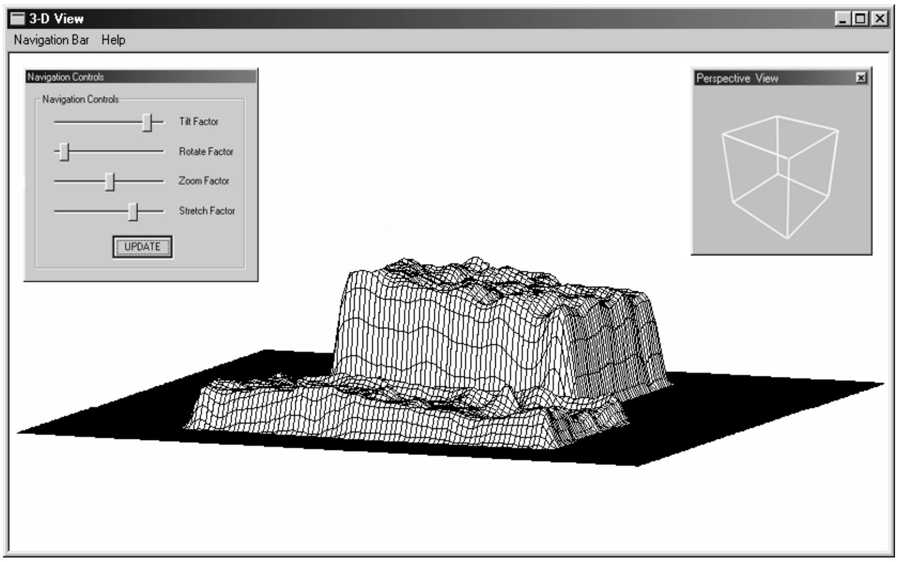
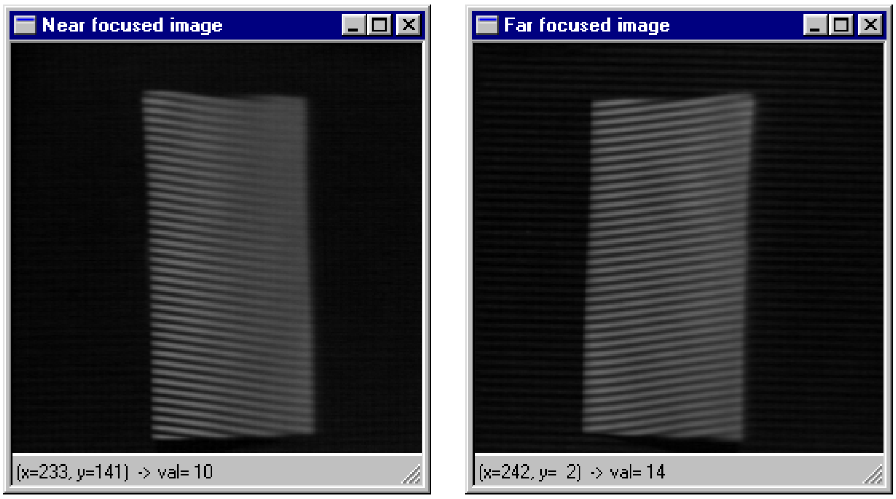


Figure 11. The depth estimated for the scene illustrated in Figure 10.



(x=233, y=141) -> val= 10

(x=242, y= 2) -> val= 14

Figure 12. The near and far focused images for a scene which contains a slanted planar object.

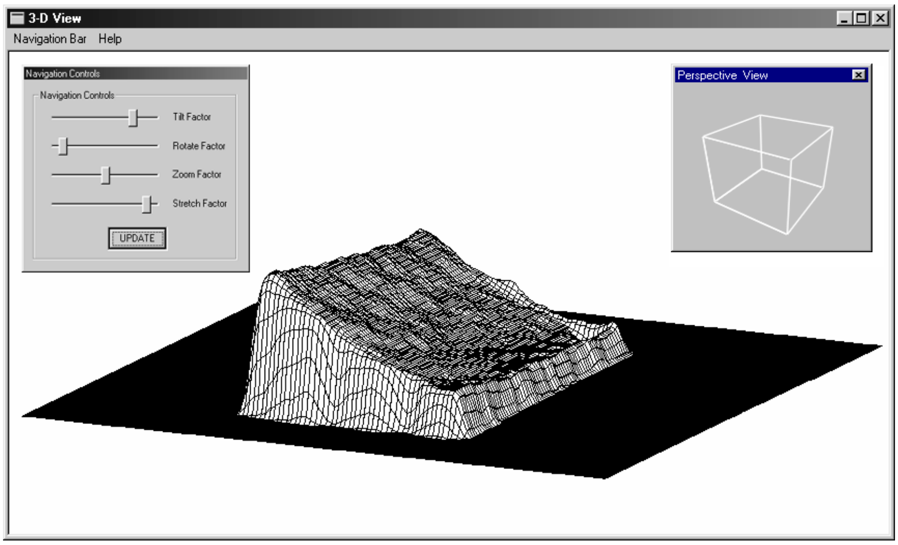


Figure 13. The depth estimated for the scene illustrated in Figure 12.

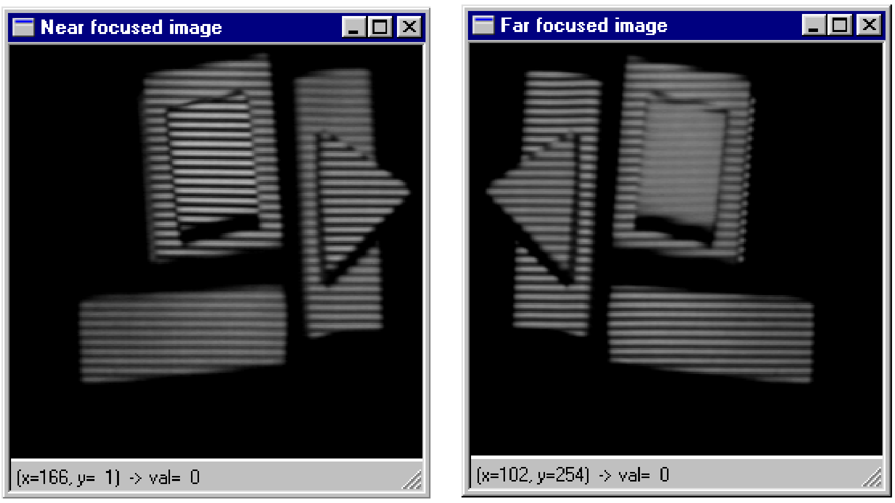


Figure 14. The near and far focused images for a scene which contains various textureless objects.

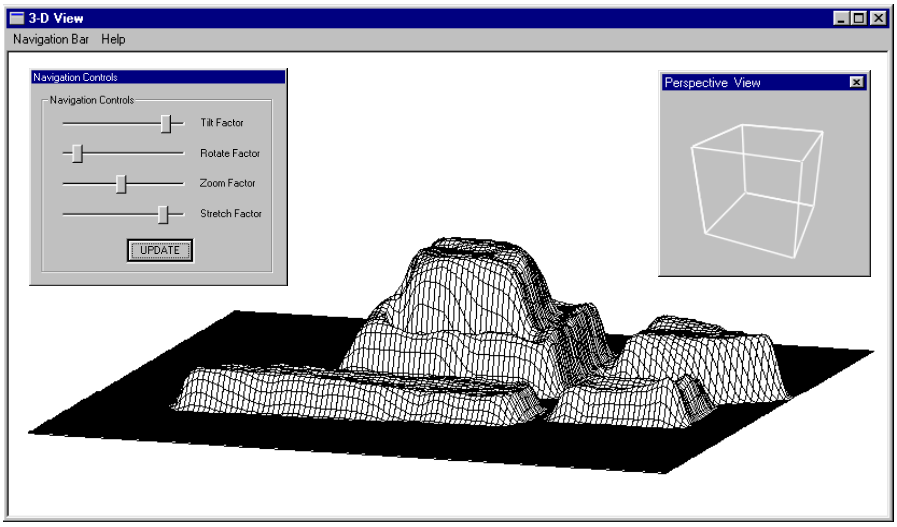
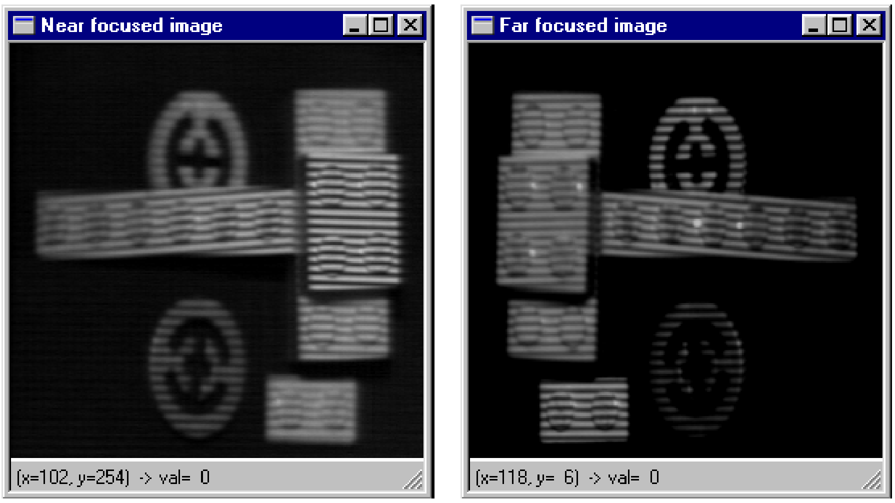


Figure 15. The depth estimated for the scene illustrated in Figure 14.



(x=102, y=254) -> val= 0

(x=118, y= 6) -> val= 0

Figure 16. The near and far focused images for a scene which contains various LEGO objects.

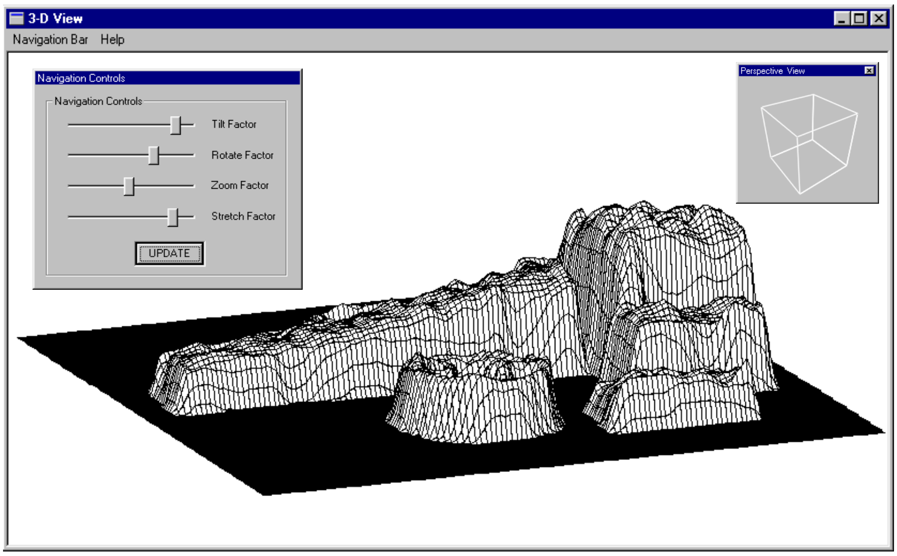


Figure 17. The depth estimated for the scene illustrated in Figure 16.