Pattern Recognition Letters 5 (1987) 63-69
North-Holland

January 1987

Depth from focus

P. GROSSMANN

GEC Research Ltd., Long Range Research Laboratory, East Lane, Wembley, Middx., HA9 7PP, United Kingdom

Received 14 September 1985
Revised 12 June 1986

Abstract: A new method for extracting depth information from 2-dimensional images is propsed. The depth structure of a scene is computed using measurements of the degree of blur in an image which is only in parts sharply focused. Preliminary tests suggest that the method is surprisingly powerful and may have interesting applications.

Key words: Depth map, blur.

1. Introduction

One of the most important stages in a general vision system is a process by which depth information is extracted from 2-dimensional images. This is frequently followed by a 3-D reconstruction of the scene and ultimately by the object recognition and scene interpretation. There are also other applications where the main task may be to generate and display a depth map without further processing (e.g. remotely operated vehicles, underwater inspection systems).

There are many processes which can be used to compute information about depth or other properties of surfaces in the scene in question (see e.g. Marr, 1982, pp. 103 and 265), stereopsis being one of the most important ones. Surprisingly, there exists one phenomenon which offers a clear depth clue but which has until now been largely ignored.

Anyone who has ever seen a photograph taken with a small depth of focus will be familiar with this phenomenon: the objects in focus are sharp and clear while the background is blurred; so may be some objects in the foreground. The further away from the focused plane an object is (in space), the fuzzier (more blurred) its image appears on the photograph.

Some work has been done in the past on various methods of evaluation of the camera-to-object (or rather, focal plane-to-object) distance (=range) for components of the scene in sharp focus (see e.g. Jarvis, 1983). This range evaluation, which has been used in automatic focusing methods, forms only a small part of our 'depth from focus' (DFF) technique which is mainly concerned with the depth structure of the scene and uses all the parts of the image (in and out of focus) to compute it. Ranging merely provides a constant factor which relates the relative depth values to the absolute ones.

2. The DFF method

There are several assumptions which have to be made and tested (no matter how plausible they look) before the method can be formulated with any degree of confidence. The assumptions are:

(i) It is possible, in most situations, to create an image with a useful range of blur.

(ii) There exists a meaningful measure of blur which can be evaluated in a consistent way for important points in the image.

(iii) There exists a simple relationship between

0167-8655/87/$3.50

the blur measure and the distance from the focused plane to the relevant point in space.

The method then consists of the following steps:
- Find a set of image primitives (edges etc.).
- Evaluate the blur measure W for each primitive.
- Convert W to the relative depth d (the distance from the focused plane).
- If necessary, use range information to convert d into the absolute depth z.

The emphasis, at least at this stage, is on empirical verification of the assumptions and feasibility tests of the method rather than development of complex mathematical formulation. The aim is to keep the method and its implementations as simple as possible.

3. Testing the assumptions

Some examples of images containing a significant range of blur clearly exist. For the applicability of the method, however, it is important to determine the general conditions under which a useful range of blur can be achieved with a conventional photographic lens.

Figure 1 illustrates how the depth of field (both near D_n and far D_f) relates to the other parameters of the imaging process. The equations

$$D_n = cs/(A+c), \qquad D_f = cs/(A-c) \tag{1}$$

relate D_n and D_f to the focusing distance s, aperture A and the circle of confusion c (e.g. Kingslake, 1983).

If we require now that D_n is less than $\sim 1/3$ of the relevant depth range in the scene and if we take $A = 3.6$ cm, say, and $c = s/1700$ (Kingslake, 1983), we can use the method for scene depths of >5 cm at 1 m distance, >5 m at 10 m distance and >0.5 mm 10 cm distance. This implies limitations which are probably quite similar to the limitations in stereo, although at very small distances DFF has a clear advantage.

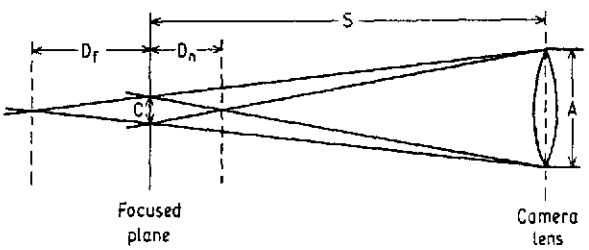

Figure 1. Depth of field diagram.

4. Human vision

Before considering our experimental set up and results it is useful to comment on the DFF phenomenon in human vision. As implied by eqn. (1), the depth of field is smaller for larger apertures. The human eye has aperture $\sim 10\times$ smaller than a good camera lens and correspondingly the depths of field are $\sim 10\times$ greater. Hence any effect associated with the blur in human vision will be much weaker. Secondly, our visual input does not consist of separate snapshots taken with fixed focus; we continuously scan the scene and continuously change focus (accomodate) so that we are only very seldom aware of any blurred objects in our field of vision. It is, however, quite possible that these changes in focus are used as cue for the depth structure of the scene, but we have not been able to find any psychophysical data relevant to this question.

5. Experimental set up and results

While our assumption (i) was dealt with in Section 3, the other basic assumptions of DFF can only be tested by the implementation of the method. The first tests were performed using a specially prepared object as illustrated in Figure 2.

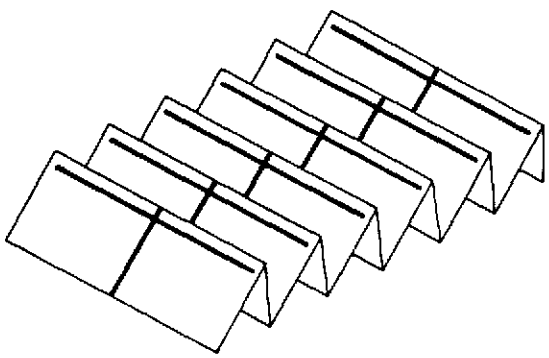

Figure 2. A drawing of the test object.

It is a folded piece of white paper that presents to the camera a series of (almost) parallel surfaces which are marked with thin black strips running horizontally and also in the 'vertical' (on the image) direction. The size of each face is 10 cm × 5 cm and the object was extended to ~35 cm in length. The mean distance from the camera was ~1 m.

The digitised images were created using a vidicon camera connected to the GEC Image Display Processor.

One of the images taken is shown in Figure 3. In this case the third horiontal strip from the bottom is in focus while the top and bottom strips are clearly blurred. The images were stored as 512 × 512 × 8 bit arrays and processed at resolutions 256 × 256 and 512 × 512.

The algorithm used in the first implementation of the method was the following: The standard Marr–Hildreth operator (Marr and Hildreth, 1980) was used to find the edges and their orientation (Lloyd, 1985). For each edge the first derivative of the grey level intensity perpendicular to the edge direction was computed (using the original image) and the width of the distribution peak W was evaluated. To keep the implementation as simple and fast as possible, we measured W by counting the number of pixels in the peak with values greater than half maximum. A correction was made for directions other than horizontal or vertical. Although this method worked well in our case, in general it can be ill defined and sensitive to noise. Smoothing the image prior to differentiation improves the situation but also introduces an unwanted contribution to W. Smoothing at different scales and taking the zero width limit looks attractive, but computationally rather involved. The development of an optimised, robust and fast W estimator is in progress.

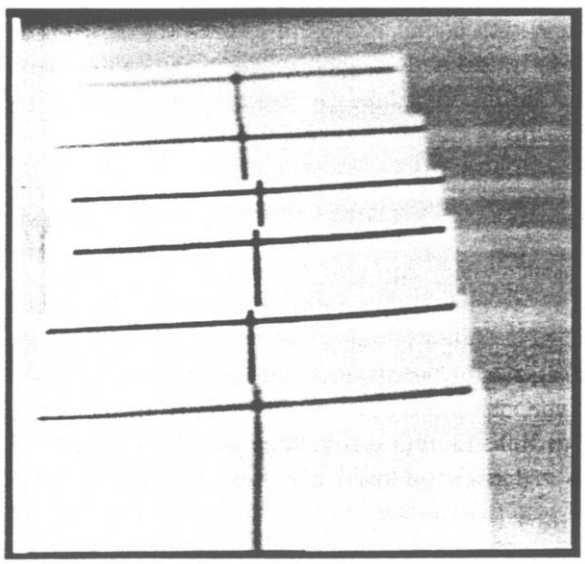

Figure 3. An image of the test object.

At this point scale emerges as an important concept. While a small size edge finder is reasonably efficient in locating both 'narrow' and 'wide' edges, an operator measuring the width of an edge has to adjust its size accordingly. If the region (or window) we consider is too large, it may contain several edges, while a region which is too small may contain only a part of one edge. In both cases the estimate of W would be wrong. Thus, because of the large range of blur, the use of different scales is essential in the DFF method.

The statistical error associated with W is obviously large and so it is necessary to average over a number of edges. This also makes the display of results easier. The results shown here are based on

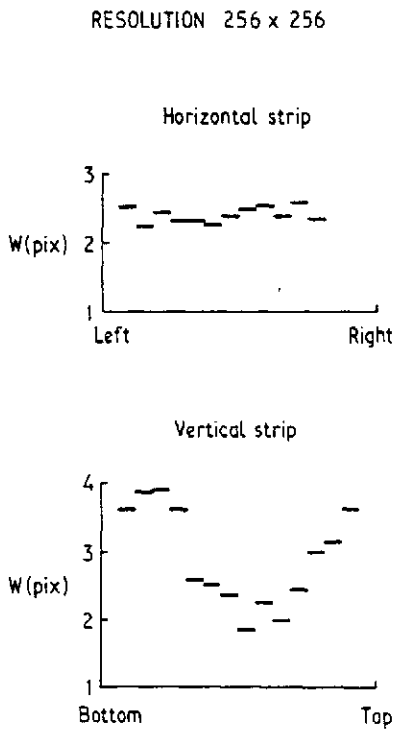

Figure 4. W for different parts of the image at 256 × 256 resolution.

values of W averaged over squares of 16×16 pixels.

Figure 4 shows W values for the image of Figure 3 at resolution 256×256. First, W for the strip in focus is plotted as a function of the position along the strip (i.e. the image column number). As all the points on each strip lie at the same distance from the focused plane (here $d = 0$) no variation in W is expected. Also shown is W for the 'vertical' strips as a function of the image row number. Here one expects W to vary with the distance from the focused plane and indeed a strong effect is observed.

Figure 5 shows the corresponding results at the higher resolution of 512×512. The conversion of W values to depth profile requires two more assumptions. Firstly, the observed first derivative distribution is a convolution of the intrinsic edge shape (characterised by a width parameter W_i) and blur (characterised by a width parameter B). By intrinsic shape we mean here the shape of the edge in sharp focus. It is reasonable to assume that the corresponding widths W_i and B are related to the measured width W as follows:

$$W^2 = W_i^2 + B^2 \tag{2}$$

(i.e. as for a convolution of two Gaussian distributions). The intrinsic width W_i is unknown in general and this could, in principle, be a serious problem. In many applications, however,

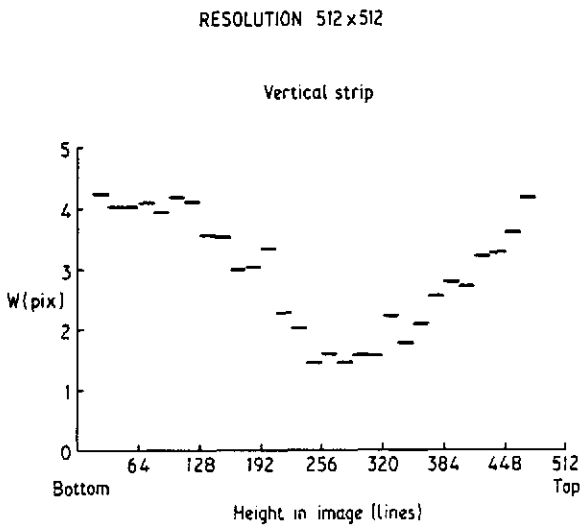

Figure 5. W profile for 512×512 resolution.

it may be reasonable to assume that it the same for all the edges in the image (e.g. in the case of manufactured parts with sharp object edges). In the special case of ideal step edges W_i is theoretically zero but in practice (as a consequence of image discretisation) values between 1 and 2 pixels are expected. When the ideal step edge coincides with the boundary between two pixels, the grey level sequence, say: $\ldots, G_1, G_1, G_2, G_2, \ldots$ gives the following difference distribution: $\ldots, 0, G_2 - G_1, 0, 0, \ldots$, which has, by definition, width of 1 pixel. On the other hand a step edge passing through the middle of a pixel will give a grey level sequence: $\ldots, G_1, G_1, (G_2 + G_1)/2, G_2, \ldots$ giving the difference distribution: $\ldots, 0, (G_2 - G_1)/2, (G_2 - G_1)/2, 0, \ldots$ which is 2 pixels wide. With averaging we can expect a value of ~ 1.5. This is indeed observed in Figure 5 as the minimum value of W which corresponds to the focused plane. One can then define W_i to be the smallest value of W observed and we can use eqn. (2) to extract B.

The next step is the conversion of B into the distance from the focused plane d. Geometrical optics (e.g. Kingslake, 1983) suggests a linear relationship:

$$d = \text{const} \cdot B$$

where the constant clearly depends on the camera setting. While it mights be possible to calculate the value of the constant from the parameters of the optical set up, it may be much simpler to determine it by calibration. From the camera setting we can also compute the distance s from the camera lens to the focused plane (= range) which we need to determine the absolute depth in space z:

$$z = s + d \tag{3a}$$

or

$$z = s - d. \tag{3b}$$

Here we encounter another problem – that of ambiguity. Unless all the relevant parts of the scene lie in the same hemisphere defined by the focused plane, it is not automatically obvious which of the equations (3a), (3b) should be used for which part. In our case the ambiguity has been resolved 'by hand'. It is, however, an important problem and the appropriate strategy has yet to be devised.

Resolution of the ambiguity problem finally

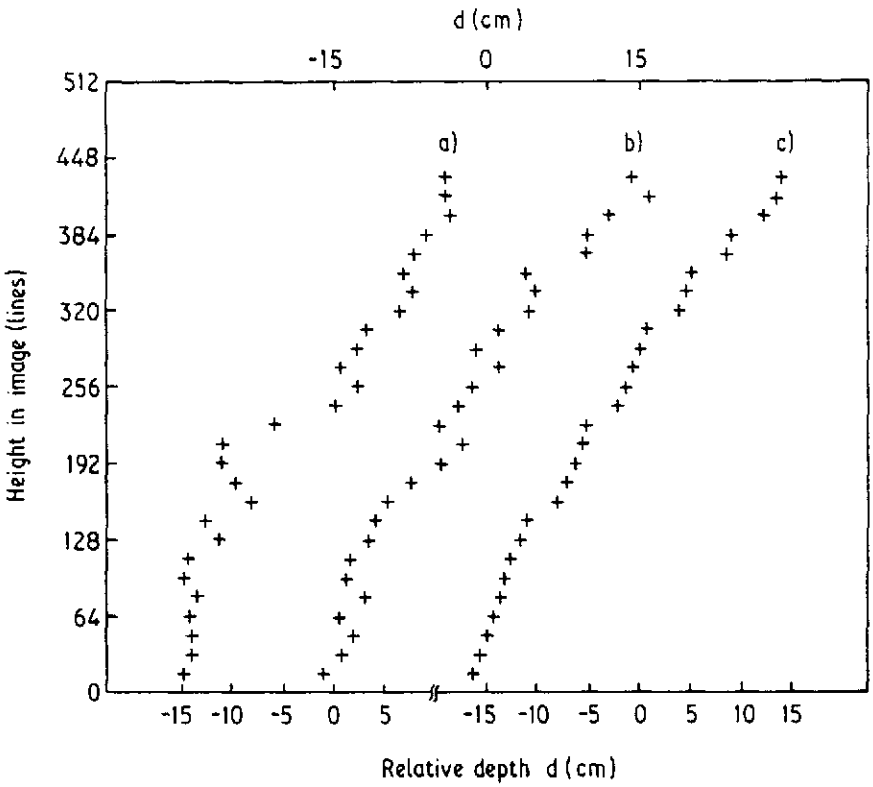

Figure 6. The 'depth profile' of the test object; data (a) and simulation using 6 equidistant line segments with $\sigma_d = 1.25$ cm (b) and $\sigma_d = 2$ mm (c).

yields a set of depth values for our test object (Figure 6a). Although the overall trend is correct, the detailed step structure is not clearly visible as it is obscured by the noise in the data. A quantitative comparison between the data and the test object was performed using a simple model of the object, as its detailed description was not available. The six planar faces (or rather the six corresponding 'vertical' strips) were represented by six parallel lines. This model was used in a least squares fit to the data which was performed for a range of values of σ_d, the relative depth error. The orientation and the distances between the model lines were left as free parameters in the fit. From the values of σ_d tried in the fit we chose the one that gave chi-square per degree of freedom close to 1.0 (expected for a good fit) as our fit estimate of the depth error: $\sigma_d = 1.25$ cm.

We can estimate the purely statistical component of the error as follows: the peak width of 4 pixels can be estimated with ~50% error. By averaging over a 16×16 pixel neighbourhood this can be reduced to ~10% which corresponds to ~1.5 cm for $d = \pm 15$ cm. This is consistent with our fit result, as we expect the statistical component to dominate. The value of σ_d can also be compared with the error on absolute depth in space $\sigma_z \sim 1.5$ cm (1.5% at 100 cm) achieved in stereopsis at resolution 64×64 as reported by Lloyd (1985).

Our fit obviously underestimates the true error as a comparison with the true object profile would probably show. On the other hand these are the first results and we can expect the depth data to improve with better imaging hardware and more sophisticated analysis, e.g. in the estimate of the peak width W.

A failure to reveal the object's step structure is to be expected as the scale of the structure is almost the same as the depth resolution σ_d. In order to detect depth profile details of a particular size, one needs depth resolution several times smaller. A computer simulation of a more regular depth pro-

file using six equidistant line segments and the same depth resolution yielded a pattern very similar to that in our data (Figure 6b). Only when much better resolution ($\sigma_d - 2$ mm) was used, did the step structure clearly emerge (Figure 6c).

6. Present problems and future direction

Some of the problems have already been mentioned. The ambiguity between the parts of the scene in the opposite hemispheres equally distant from the focused plane could be resolved by operator intervention (i.e. additional information), by use of another image taken with a different focused distance or avoided altogether by insuring that all the relevant parts are in the same hemisphere at the focusing stage.

Another difficulty lies in the fact that not all edges have the same intrinsic width W_i. Whether one can distinguish 'wide' edges in focus from narrow but blurred ones has yet to be investigated. At present, however, this somewhat limits the applicability of the method to images where the assumption of a constant W_i is a reasonably good one.

In a more general case one may have to resort to an implementation where a series of pictures is taken each with a different camera setting so that several subsets of edges can be identified as being in focus (i.e. having minimum W) in different focused planes. This would be a generalisation of the simple range finding method as has already been described by e.g. Jarvis (1983).

Other limitations and possible improvements are of a less fundamental nature. An improvement in the quality and resolution of the image capture hardware is desirable. Fast and efficient algorithms for edge detection, blur evaluation and intelligent averaging have to be developed to capitalise on the inherent simplicity of the method.

7. Alternative algorithms

There obviously exists a number of operators that could provide a simpler and faster (and perhaps also better) ways of evaluating grey level gradient in the neighbourhood of every pixel. This may be important in designing a simple and fast implementation. Preliminary studies using combinations of Sobel-like operators produced some encouraging results although the problem of scale again emerged as the dominant one. In general our preference would be for operators that do not introduce unwanted smoothing.

8. Summary

A new method for extracting depth maps from 2-dimensional images has been described. First experiments with the test object demonstrated a clear correlation between the depth in space and a measure of the blur in the image W. This demonstrates feasibility of the DFF approach, although some problems still have to be solved.

9. Postscript

After the work described in this paper has been complete, a paper entitled "A new sense for depth of field" by A.P. Pentland (1985) was published. This paper contains ideas and results very similar to the ideas and results described above. While we place the main emphasis on experimental demonstration of the blur-depth correlation, Pentland concentrated on the mathematical formalism and in particular solved the 'back-front' ambiguity referred to in Sections 5 and 6.

Acknowledgements

The author would like to thank Sheelagh Lloyd and Iain Graydon for their help with the vision programs and hardware. Special thanks are due to Margaret McCabe for her support and encouragement.

References

Marr, D. (1982). *Vision*. Freeman, San Francisco.

Jarvis, R.A. (1983). A perspective on range finding techniques for computer vision. *IEEE Trans. Pattern Anal. Machine*

Intell. 5(2), 122–139.

e.g. Kingslake, R. (1983). *Optical System Design*. Academic Press, New York.

Marr, D. and E. Hildreth (1980). Theory of edge detection. *Proc. R. Soc. Lond.* B207, 187–217.

Lloyd, S.A. (1985). A dynamic programming algorithm for binocular stereo vision. *GEC J. Research* 3(1), 18–24.

Pentland, A.P. (1985). A new sense for depth of field. IJCAI 2, 988–994.