Depth Extraction by Focal/Aperture Variation

S. R. Daniel[1]

IBM UK Scientific Centre, Winchester
Athelstan House, St. Clement Street
Winchester, Hampshire, SO23 9DR

This paper describes a method of extracting depth information about a scene from a single static viewpoint. The approach uses aperture variation to obtain a sequence of images differing only in depth of field, thus avoiding the 'correspondence problem' associated with stereo and optical flow techniques. These images contain the necessary coded depth information. By using one image (that with the smallest aperture) as a reference and using knowledge of the Point Spread function (PSF) of the system, we can compare the images and hence obtain a depth map for all those points in the image which have sufficient grey-level gradient. We propose a realistic model for the system. Several methods of depth extraction from the data are suggested.

The use of focal variation to recover depth information about a scene has been the subject of several recent papers [1,2,3]. We begin with an outline of the method of depth recovery by focus. In section I a new theoretical model for the system is introduced. This is followed in section II by a discussion of how to obtain blurred images. We then outline various methods that are available for solving these systems. We will consider methods which give an estimate of the solution and those which give a 'statistically optimal' solution.

I. THEORY

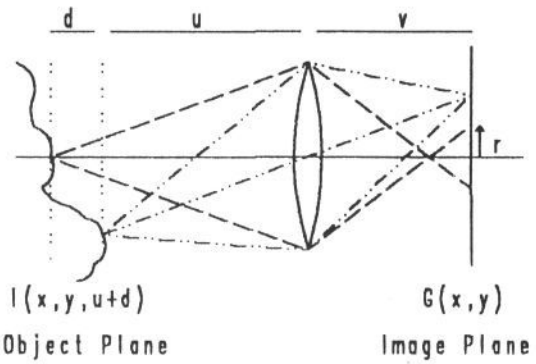
Basic Principles

Let us consider a simple lens system obeying:
$1/u + 1/v = 1/F$

where F is the focal length which is constant for a given lens, v is the distance from the lens to the image plane, and u is the distance from the lens to the focal plane.

All points in the focal plane I(x,y;u) will map to points in the image plane G(x,y); all points in an object plane I(x,y;u+d), displacement d from the focal plane, will map to blur circles in the image plane.

Figure 1



The radius of the blur circle r is related to the distance from the focal plane d by the following equation [1,4]

$$r = 1/(2f)\,(\,v - F - Fv/(u + d)\,) \qquad (1.1)$$

which may be re-written

$$D = u + d = Fv/(v - F - 2rf) \qquad (1.2)$$

where f is the f-number[2] of the lens and D is the overall distance from the lens to the object plane. The above equation demonstrates that there is a gradient of focus, or blur, for points on either side of the focal plane. This can be modelled by a transform function H(), which describes how the camera system blurs a single point source. This is referred to as the Point Spread Function (PSF) and is a characteristic of the lens camera system.

Model of System

If we consider a single object plane I(x,y;u+d) then the output G(x,y) will be the superposition of all the projected blur circles, radius r, from each point in the original input. This output G(x,y) can therefore be related by a convolution of the intensity image of this plane, I(x,y), with the system's PSF H(x,y,σ) :

$$G(x,y) = \iint H(\eta - x, \xi - y\,;\,\sigma)\, I(\eta,\xi)\, \delta\eta\delta\xi \qquad (2.1)$$

---

[1] This research has been conducted while the author was a pre-university Student at the IBM UK Scientific Centre.

[2] The f-number for a lens is the ratio of the focal length (F) of the lens to the diameter of the aperture.

BMVC 1990 doi:10.5244/C.4.50

where the PSF is parameterised by σ which equals the blur radius r and hence the depth of the object plane u + d via equation 1.2.

For a 3-dimensional input I(x,y,z), the output G(x,y) will consist of the superposition of all the blur circles from points at varying depth in the scene, and is therefore scene dependent. We propose to model the output G(x,y) for this superposition as follows:
We assume initially that the shape of the blur circle from a point is unaffected by neighbouring points in the scene. Thus we can write this as a space-variant convolution since the PSF now varies across the image [4].

$$G(x,y) = \int\int H(\eta\text{-}x,\xi\text{-}y\ ;\ \sigma(\eta,\xi))\ I(\eta,\xi)\ \delta\eta\delta\xi \qquad (2.2)$$

where η, ξ are two spatial variables and the PSF H() is now parameterised by the set $\sigma(\eta,\xi)$.

In other words, if we know the intensity map I(x,y) and its corresponding depth map D(x,y) for a given projection of the scene, we can calculate the blur radius map σ(x,y) using equation 1.1. Thus the intensity value at G(x,y) will be the weighted sum of all intensity values in I(η,ξ) whose blur circle overlaps the point at x,y.

Some authors have used the assumption that the blur remains constant over the blur circle [1,2,3,5]. This reduces the region over which this integral is evaluated to being the area of the blur circle centred on the point (x,y). The integral is then reduced to the space-invariant form of convolution:

$$G(x,y) = \int\int H(\eta\text{-}x,\xi\text{-}y\ ;\ \sigma(x,y))\ I(\eta,\xi)\ \delta\eta\delta\xi \qquad (2.3)$$

However, I believe this places too great a restriction on the scene and limits accuracy [4].

Form of Point Spread Function

From geometrical optics the PSF for a simple camera system can be shown to be a uniform cylindrical impulse function. However imperfections in the optical system introduce aberrations (e.g. spherical and chromatic aberration). Most important is the effect of diffraction. The full system can be described by a complex series of Bessel functions with appropriate constants [1]. However, for standard optical systems diffraction effects can be ignored since the aperture diameter is always several orders of magnitude larger than the wavelength of light. Hence the true PSF for the system still has the shape of a cylindrical function. Unfortunately the PSF is therefore a generalised function and lacks a specific functional form. This has perhaps led other authors to adopt the use of a Gaussian PSF which is simpler to manipulate. In fact, the exactness of the PSF probably only becomes significant when high resolution depth recovery is undertaken.

II. DISCUSSION

In order to obtain the unknown set σ(x,y), we can either seek to estimate the parameters by observing some characteristic of grey-level variation in a single blurred image, or by comparing two or more images which differ by some change to the lens system. If we examine just one image we are limited in the accuracy to which we can estimate depth. This accuracy will be highly dependent on the scene; it will introduce ambiguities at rough edges which appear naturally blurred. However, if we look at two or more images, we can look at the change of blur, which allows us to construct a more reliable depth map. We need a method of obtaining a sequence of images which can be easily compared, and we recommend that these are obtained by aperture variation. This gives us a sequence of images which differ only in depth of field and overall intensity, whereas by varying the focus or position of the camera we introduce the correspondence problem common to stereopsis and optical flow methods. The change in overall intensity can be normalised, either when the data is collected, by using different exposure times or mutual density filters, or numerically (though this would affect the dynamic range of the images). Hence introducing the f-number into our equation and writing it in the discrete form gives:

$$G(x,y;f) = \sum_i\sum_j H(i\text{-}x,j\text{-}y,\sigma(i,j;f))\ I(i,j) \qquad (2.4)$$

$$\sigma(i,j;f) = 1/(2f)\ (v - F - Fv/D(i,j)) \qquad (2.5)$$

where I(i,j) is an estimate of the ideal unblurred image and D(i,j) is the unknown depth map.

III. METHODS OF SOLUTION

We have two possible data sets to consider. The first is where we have an image of the source I(x,y) and one or more blurred images G(x,y); the second is where I(x,y) is unknown. The former will be the more useful since, in order to perform any subsequent scene analysis, we will need the unblurred image I(x,y) of the scene.

We can either choose to solve the problem of obtaining the set σ(x,y) using equation 2.2, or using the simpler form of equation 2.3. which approximates 2.2 by assuming local-invariance within a certain region about each point. The space-variant form requires a method of inverse optimisation, whereas the second form can be solved more directly. Previous work has concentrated on solving this simpler model [1,2,3]. The depth maps obtained by this method will be coarse and will give, in effect, the average depth over each region. The practical implementation of this method is straightforward and fast, which makes it attractive to real-time applications. The approach requires estimating the high frequency component of the Fourier spectrum for each region [1,3,4,5]. The main drawback appears to be that in order to obtain a stable solution a large region is re-

quired, but this results in a breakdown of the assumption that blur remains constant over the region. Hence there is a trade-off between the size of the region and the stability of the solution. The ultimate size required to obtain a stable solution therefore constrains the density of the final depth map.

The new model, however, does not require the image to be broken up into regions, and therefore the depth map can be dense. Moreover, it should be reliable, as any errors may be modelled mathematically and it will not be prone to the occasional large mismatch errors common to mistaken correspondence in other techniques. Unfortunately, the solution of the model is nontrivial and requires the use of an inversion technique. This is more computationally expensive and so will be less useful in a real-time application.

Inversion Techniques

Inverting the model requires the deconvolvution of equation 2.2 which has total parametric uncertainty in the model. Similar problems involving incomplete model knowledge also arise in fields such as seismology. The problems are generally considered to be ill-posed, by which we mean that they do not satisfy the three conditions of existence, uniqueness and stability of the solution. Several methods do exist which in some cases exploit parallelism.

A simple approach is the least-squared method of residuals whereby a set $\sigma()$ is "guessed" to minimise the Chi-squared statistic C in the following equation.

$$C(\sigma(1,1),..,\sigma(n,n)) = \chi^2 = \Sigma\Sigma(G(i,j) - G'(i,j))^2/V(i,j) \qquad (3.1)$$

where G(i,j) is the actual noisy data, and G'(i,j) is the predicted data calculated under equation 2.4. using the "guessed" set $\sigma()$. The variance term V(i,j) is related to the stability of the solution G(i,j) = G'(i,j). It is assumed initially to be uniform but it is in some sense related to the image I(i,j) and the depth map which suggests we should use it as a weighting factor to allow stricter control over the chi-squared statistic [4].

Equation 3.1 effectively generates a set of solutions which fit the data. In practice, in order to determine a unique and stable solution for such an ill-posed problem, we need to introduce additional a priori information (sometimes referred to as regularization functions). Any a priori constraints can either be in a deterministic form (arising from physical or geometric considerations such as positivity) or in some stochastic form (direct constraints on probability density functions). A stochastic approach derived from Bayesian probability theory allows us to coherently handle noisy data and is an effective method of solving inverse problems of this nature.

Examples of methods other than the straightforward unconstrained iterative least squares , which allow the incorporation of a priori knowledge, are the Maximum Entropy Method [4], Stochastic relaxation and Simulated Annealing.

It is worth noting that since most methods require iteratively updating a guessed set $\sigma()$, it is quicker to estimate $\sigma()$ using a Fourier transform based method (using the simpler model outlined in the last section), rather than by starting with a uniform guess for $\sigma()$. However, by doing so we introduce a slight bias on these results and slightly effect the convergence of any statistical procedure.

IV. PRELIMINARY RESULTS

Image 1.1 shows the intensity image of a computer generated textured 3-dimensional object. Image 1.2 shows the corresponding depth map.

Image 1.1 - Source Image

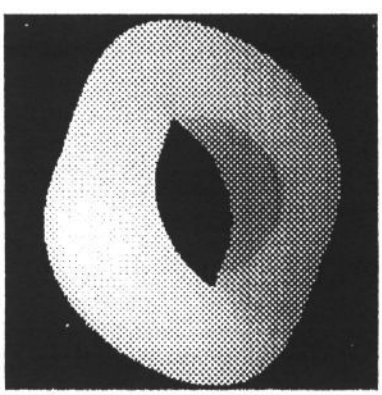


Image 1.2 - Depth Map

These images were then used in equation 2.4 to calculate the predicted blurred image (Image 2) that would be obtained through a camera with a narrow depth of field (i.e. with a small f-number).

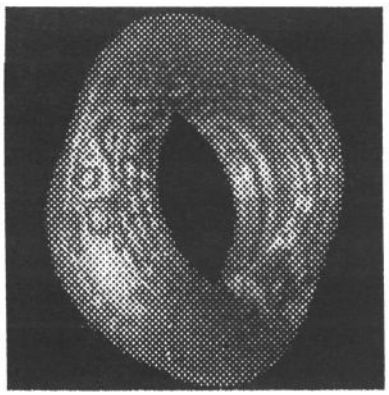


Image 2 - Blurred Image

This image was used as the actual data G(i,j). We then calculated the predicted data G'(i,j) that would be obtained from a uniform depth map using equation 2.4 and 2.5. Hence the magnitude of the chi-squared statistic in 3.1 tells us how good our estimated uniform depth map is by comparing the predicted data G'(i,j) with the actual data G(i,j). The derivative of 3.1 with respect to the estimated depth map gives us a measure on the iterative change to be made to improve the depth estimate at each point. The following image was obtained from image 1.1 and image 2, using constant variance, and shows this derivative.

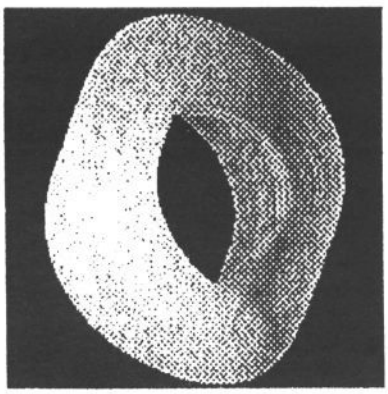


Image 3 - Qualitative Depth Information

This derivative shows qualitative evidence that depth information is recoverable from a blurred image, and it is used in a first stage of the iterative depth extraction algorithm [4].

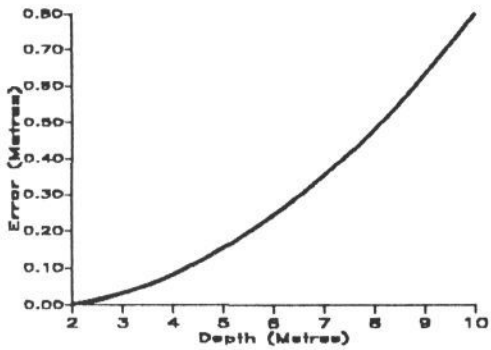
V. DEPTH RESOLUTION

A detailed account of the depth resolution and errors is beyond the scope of this paper (a more detailed analysis can be found in 4). Two general observations are apparent: the depth resolution is directly limited by the accuracy to which we can determine σ; and the error in determining $\sigma(x,y)$ depends on the magnitude of the intensity at I(x,y), the gradient at I(x,y), the magnitude of σ and the accuracy of the algorithm used to determine it. We will only be able to obtain reliable depth estimates for those points in I(x,y) which have sufficient grey-level gradient.

If we assume a constant error $\partial\sigma$ in determining σ then the error in depth ∂D is related by the following equation:

$$\partial D/D = (D/u - 1) \partial\sigma/\sigma \qquad (4.1)$$

which gives the following graph for a 2 % error in σ, and u = 2:



CONCLUSION

The emphasis in this paper has been to outline the theory behind depth extraction by focus. We have proposed a new model which is a more accurate description of the physical system. This model can be reduced under assumptions to the simpler less accurate form adopted by previous authors; it also allows us to investigate the theoretical potential of depth information obtained from blurred images.

REFERENCES

1. **Pentland, A. P.** "A New Sense for Depth of Field" *IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-9 No.4 July 87*

2. **Pentland, A. P.** "A Simple, Real-time Range Camera" *IEEE Comp. Soc. Conf. on Comp. Vision and Pattern Recognition San Diego,Ca June 4/89*

3. **Subbarao, M.** "Parallel Depth Recovery by Changing Camera Parameters" *CH2664-1/88/IEEE*

4. **Daniel, S.R.** "Review on Depth from Focusing Methods", *IBM UKSC Technical report 1990 no 230.*

5. **Bove, V.M.** "Discrete Fourier Transform Based Depth-from-Focus", *Proceedings OSA Topical Meeting on Image Understanding and Machine Vision, Falmouth,MA, June 12-14, 1989.*

Acknowledgements:

I would like to thank Mike Ibison of the IBM UK Scientific Centre for invaluable advice at various stages of this work.