Scale Space Surface Recovery Using Binocular Shading and Stereo Information

A.G. Jones and C.J. Taylor,
Department of Medical Biophysics,
University of Manchester,
Oxford Road, Manchester M13 9PT,
United Kingdom.

Shape-from-shading algorithms that use a single image often find incorrect solutions because of ambiguity in the image shading. We describe a robust shape-from-shading algorithm using *scale space tracking* that can resolve most of these ambiguities by using two images taken from slightly different positions. Further improvements are obtained by combining stereo with shape-from-shading. Results are shown for synthetic test images (with and without added noise) for a Scanning Electron Microscope stereo pair, and for a stereo pair taken using a video camera.

1 Introduction

Recovering depth from the shading information in a single image is a difficult problem that has interested computer vision researchers for many years [1]. Although significant progress has been achieved using synthetic image data or unrealistically high quality real data, the results obtained with more general images have so far been disappointing. A survey of the literature reveals that even for those real images where shape-from-shading might be expected to work well, the results obtained are typically only qualitatively correct (e.g. [1,2,3]).

By using a stereo pair of images, the shape-from-shading problem can be greatly constrained and the problems with ambiguity in the shading reduced [4]. Having a stereo pair also allows us to combine stereo with shape-from-shading. Shape-from-shading and stereo are complementary, since shape-from-shading works best where the surface is smooth and featureless, whereas stereo works well where the surface is rough and contains interesting structure.

The work described here extends our monocular scale space shape-from-shading algorithm (see [5]) to use the information available in two views of the surface. The algorithm is robust, provably convergent and does not require prior decisions to be made about the smoothness of the solution. The algorithm works by first finding an approximate solution at a coarse scale and then tracking this solution through scale until a final solution is obtained at full resolution. Scale space behaviour is achieved by constructing the solution at each scale from a set of Gaussian basis functions of appropriate width. Results are shown for several synthetic images (both with and without added noise), for a real Scanning Electron Microscope (SEM) image pair, and for a real image pair taken with a video camera.

2 Related Work

Most shape-from-shading algorithms are derived from the early work of Horn [1], where the Calculus of Variations was used to obtain an iterative scheme. These algorithms typically first recover the surface gradients and then at a later stage cast these onto the nearest integrable surface [1,6]. More recently, Horn devised an algorithm that recovers

BMVC 1995 doi:10.5244/C.9.8

surface height directly, avoiding the problem of ensuring integrability in the gradient field [7]. Szeliski [8] improved upon the slow convergence of Horn's relaxation schemes by minimizing the integral equation directly using conjugate gradient descent. Ron and Peleg [9] developed a multiresolution shape-from-shading algorithm which also offers improved convergence rates.

Progress has also been made with local shape-from-shading algorithms, most notably the recent work of Oliensis and Dupuis [2] which casts shape-from-shading as an optimal control problem. Their algorithm is fast and robust, but does not cope adequately with inflections and plateaux in the surface and requires prior knowledge about singular points (points of zero gradient).

Several ways of using stereo images for shape-from-shading have been suggested in the literature. Shao *et al.* [10] used a global co-ordinate system and the Calculus of Variations to derive a non-linear system of equations for solution by an iterative method, but no results are given in their paper. Ellison and Taylor [11] used a similar approach, and although results were obtained, the method often failed to converge [12]. Lee and Brady [13] used stereo matching (by hand) to achieve approximate correspondences between the images and then used an algorithm similar to photometric stereo to recover the surface gradient in the direction of the stereo epipolar lines. Hartt and Carlotto [14] applied the theory of Markov Random fields to an energy functional consisting of a brightness error term for each image and a smoothness term. Energy was minimized using the Metropolis algorithm with a coarse-to-fine strategy.

Another approach is to combine stereo with shape-from-shading. Ikeuchi [15] proposed that shading information could be used for interpolating between stereo matches. Grimson [16] combined shading information with stereo data to determine surface orientation along feature point contours, giving better surface reconstructions than could be obtained using stereo data alone. Hougen and Ahuja [17] used separate stereo and shape-from-shading modules and combined their results, although it is not clear in their paper exactly how this was done.

3 SEM Imaging

Scanning Electron Microscope images are formed by raster scanning the sample with an electron beam and detecting the electrons emitted at the point where the beam strikes the surface [18]. The number of electrons emitted is related to the angle between the electron beam and the surface normal, thus providing shading in the image. The images obtained are of high resolution and good contrast, with a large depth of field.

The SEM provides a controlled environment that is particularly appropriate for shape-from-shading since it allows the following assumptions to be made:

- The "light source" may be approximated as a point source at infinity and its position in relation to the surface is known.
- The projection can be considered orthographic.
- There is no significant mutual illumination.

Stereo SEM images can be obtained either by tilting the electron beam or by tilting the specimen stage. We favour tilting the stage, since it is more accurate.

The way in which the electron beam interacts with the specimen can be modelled using the simple reflectance function [19]:

$$R = a/\left(b+\cos\theta\right) = a/\left(b+\hat{\mathbf{n}}\cdot\hat{\mathbf{l}}\right)$$

where $\hat{\mathbf{l}}$ is the unit illumination vector, $\hat{\mathbf{n}}$ is the unit surface normal, θ is the angle between these two vectors, a is the surface albedo and b is a parameter.

4 Binocular Shape-From-Shading

For binocular shape-from-shading, we use a Cyclopean co-ordinate system. To transform co-ordinates from the Cyclopean co-ordinate system to image co-ordinates we define the rotation matrix

$$\mathbf{S}(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

representing the stage rotation. The left image co-ordinates are given by the transformation $\left[X_L, Y_L, Z_L\right]^{\mathrm{T}} = \mathbf{S}(\theta)\left[x, y, z(x,y)\right]^{\mathrm{T}}$, and similarly for the right-hand co-ordinates.

Using two images of the surface allows us to eliminate the albedo from the objective function. For a surface point (x,y), the albedo can be estimated as:

$$a_L(x,y) = E_L\left(X_L(x,y), Y_L(x,y)\right)/R_L\left(\mathbf{S}(\theta)\mathbf{n}\right)$$

where E_L is the left image, and R_L is the reflectance function for the surface as viewed from the left viewpoint. A second estimate is obtained for the right image. Since the albedo of a surface point is invariant to the viewing direction, we should have $a_L(x,y)=a_R(x,y)$ for the correct solution, suggesting an objective function of the form:

$$F = \sum_{x,y}\left(a_L(x,y) - a_R(x,y)\right)^2$$

In the following section we will use this idea to obtain an objective function which combines binocular shape-from-shading with correlation stereo.

5 Obtaining an Objective Function

Combining correlation stereo with the binocular shape-from-shading algorithm can improve the solution by resolving ambiguities in the shading.

We have used a straightforward template matching stereo algorithm to obtain a sparse array of good matches to pixel accuracy. The algorithm is run once on the original images and the matches are incorporated into the objective function by applying Bayes's theorem. For stereo, the probability of a surface z given the disparity evidence can be written:

$$P_S(z|\psi) = \frac{P(\psi|z)P(z)}{P(\psi)}$$

where $\psi(x,y)$ are the disparities found by the stereo algorithm. Assuming the prior probabilities to be approximately constant and assuming the distribution of disparity measurements to be approximately normal we obtain:

$$P_S(z|\psi) \propto \exp(-2(2z\sin\theta - \psi)^2)$$

where we assume the standard deviation of the distribution is *0.5* pixel. Using Bayes's theorem for binocular shape-from-shading (for albedo measurements) we obtain:

$$P_{SFS}(z|E_L, E_R) \propto \exp(-((E_L / R_L)^2 + (E_R / R_R)^2)/2s^2)$$

where *s* is an estimate of the standard deviation of the albedo measurements. Since the largest errors in the albedo measurements are caused by inaccuracies in the estimated correspondences between image points, the standard deviation of the albedo measurements can be estimated by considering corresponding image points:

$$s = \sqrt{\sum_{x,y} \frac{(a_L(X_L, Y_L) - a_R(X_R, Y_R))^2}{4n^2}}$$

where $(X_L(x,y), Y_L(x,y))$ and $(X_R(x,y), Y_R(x,y))$ are the co-ordinates of corresponding image points and the images are n x n pixels. As the scale decreases during the solution process, the solution surface determines an increasingly accurate map between corresponding image points, so *s* should reduce with scale. This makes it necessary to recompute *s* periodically.

Treating the binocular shape-from-shading probability and the stereo probability as (approximately) independent, the combined probability of a surface is:

$$\begin{aligned} P(z|E_L, E_R, \psi) &= P_S(z|\psi) P_{SFS}(z|E_L, E_R) \\ &\propto \exp(-2(2z\sin\theta - \psi)^2)\exp(-((E_L / R_L)^2 + (E_R / R_R)^2)/2s^2) \end{aligned}$$

Using the maximum likelihood estimator then gives the following objective function:

$$F = \sum_{x,y} 2(2z\sin\theta - \psi)^2 + \sum_{x,y} \frac{(E_L / R_L)^2 + (E_R / R_R)^2}{2s^2} \qquad (1)$$

6 Scale Space Tracking

Scale space tracking has proved a useful technique for solving many of the large-scale, non-convex minimization problems which commonly occur in computer vision, such as the problem just described. An approximate solution is first found at a coarse scale and then "tracked" as the scale is gradually reduced [20,21]. At a sufficiently large starting scale, the problem will be convex so the first solution is easily found, and although local minima reappear as the scale is reduced, the initial solution can be followed through scale space. Scale space tracking is not guaranteed to find the minimum of a non-convex energy function, but it can find a "significant" solution (i.e. a solution that first appears at large scale) [22].

The problem must first be expressed as the minimization of an energy functional, $E(\sigma, \mathbf{u})$, where σ is the scale and **u** is the solution collapsed into a vector. At each scale it is required that the energy function is at a minimum, so we have the condition:

$$\nabla E(\sigma, \mathbf{u}(\sigma)) = \mathbf{0}. \qquad (2)$$

To maintain this equilibrium, it is necessary to minimize $E(\sigma, \mathbf{u}(\sigma))$ at each new scale using a suitable optimization technique. The discrete transition from one scale to the next is most simply achieved by using the solution at the current scale as the initial condition at the next scale.

In the early work on scale space tracking (e.g. [20]) it was the input data (typically image data) that was Gaussian blurred to form the scale space. For those problems where the relationship between the input data and the solution is linear this approach leads to the solution space also exhibiting the required scale space behaviour, but in those cases where a non-linear relationship exists, the solution will not be correctly blurred. Just such a non-linear relationship occurs in shape-from-shading since the reflectance function is in general non-linear. One answer is to blur a transformed version of the image rather than the image itself [9], but this is not a complete cure. A better solution is to impose scale space behaviour onto the solution directly [22]. This is the approach we use in our algorithm.

7 Scale Space Shape-From-Shading

Whitten proposed a method for solving inverse problems by constructing solutions from an array of deformable curves, or "snakes" [22]. He realized that the concepts of smoothness and scale are essentially the same, so by controlling the smoothness (internal energy) of the snakes, scale space behaviour could be achieved. This equivalence between scale and smoothness is of importance in the following discussion.

In our algorithm, we drop the scale-dependent energy function of Whitten, $E(\sigma, \mathbf{u}(\sigma))$, in favour of a simpler objective function $F(\mathbf{u})$ that does not depend upon scale. Scale space behaviour is no longer enforced via the energy function, but by the tracking algorithm itself and by the choice of basis functions used to represent the solution.

Before describing the algorithm in detail, it is useful to look at scale space tracking in a slightly different way – as a dynamic system. First, consider a multi-dimensional space where every point represents a possible solution to the shape-from-shading problem (i.e. a *surface*). Associated with each point in this space is a pair ($F(\mathbf{u})$,$S(\mathbf{u})$), where $S(\mathbf{u})$ is a measure of the lack-of-smoothness of the solution. Now consider the point corresponding to an initial estimate for the solution. Any movement from this point involves some cost in smoothness. Scale space tracking can be performed by making moves which buy the greatest decrease in the objective function for the least cost in smoothness, i.e. always moving in the direction which maximizes the ratio:

$$-\frac{F(\mathbf{u}+\delta\mathbf{u})-F(\mathbf{u})}{S(\mathbf{u}+\delta\mathbf{u})-S(\mathbf{u})}=\frac{-\Delta F}{\Delta S} \qquad \text{for } (\Delta S \geq 0) \tag{3}$$

where $\delta\mathbf{u}$ is the change made to the solution. In this framework the scale space equilibrium condition (2) can be written as:

$$\nabla F(\mathbf{u}) \cdot \nabla S(\mathbf{u}) = \mathbf{0} \tag{4}$$

The trajectory which maximizes the ratio (3) is strongly attracted to the trajectory that maintains the equilibrium condition (2). To see why this should be so, consider a point where (4) does not hold. From such a point there must be a direction that reduces F for no cost in smoothness (since equation 4 is not satisfied) and this direction will maximize the ratio (3) (ΔS=0 so the ratio becomes $+\infty$). This direction is downhill in $F(\mathbf{u})$ and so making repeated moves eventually reaches an equilibrium position.

Tracking the solution using a method directly based upon (3) is likely to be difficult as the problem is very badly conditioned in this form. However, the conditioning can be greatly improved by using Gaussian basis functions to form the surface. The Gaussian is appropriate since it is the impulse response of the Gaussian blurring kernel – a near-optimal kernel for forming the scale space representation of a two-dimensional signal (optimal only in the continuous case – see [23]). The Gaussian needs to be of a broadness appropriate to express the solution at the particular scale being considered.

Rather than performing complete changes of basis function as the scale is reduced, it is simpler to use the Gaussian basis functions to express the changes (deformations) made to the solution. The solution surface at a particular scale can then be written as:

$$z_k = z_{k-1} + r_k \otimes G(\sigma_k) \qquad (5)$$

where $G(\sigma_k)=exp(-(x^2+y^2)/2\sigma_k^2)$ is the Gaussian convolution kernel, z_{k-1} is the solution before the change of basis, z_k is the solution after the change of basis, and r_k is an array of Gaussian amplitude coefficients. To track the solution through scale space, (5) is iterated from the initial condition $z_0=0$ until a full-resolution solution is reached.

During each iteration, the trajectory of the solution is advanced using the maximum downhill principle of (3), except that now it is the Gaussian coefficient array r_k that is variable rather than the solution vector **u**. After continuing the trajectory in this way for a while, progress becomes difficult as the conditioning of the problem worsens due to the current Gaussian basis functions becoming inappropriate at the reduced scale. The iteration is completed and a new one started using slightly narrower basis functions.

The use of Gaussian basis functions offers the further advantage that a reasonable approximation to scale space behaviour can be achieved even if the smoothness term is ignored in the calculation of the descent direction. This is because there is a tight bound on the loss of smoothness for a small step taken along *any* direction in r_k, given by $S(z_k) \le S(z_{k-1}) + C_k \sum_{i,j} r_{i,j}^2$, where C_k is a constant (see [5]). Since the bound is dependent on C_k, which increases as the Gaussians become narrower, a change of basis to narrower Gaussians should not be made until the rate of convergence using the current basis is so low as to make the change necessary.

Rather than using steepest descent optimization (which is often inefficient) to make the moves through scale space, we used *conjugate gradient descent* [23]. This method is suitable for large-scale, non-linear problems and we have found that it works well.

8 Blurring and Subsampling

As we track solutions through scale space we should use "blurred" versions of the observed images in calculating the brightness error term in the objective function (1); otherwise the solutions generated at at coarse scales will suffer from aliasing effects. Ideally we want to use the images that would have been obtained if a suitably blurred version of the *surface* had been obseved. An approximation to this can be obtained using Ron and Peleg's algorithm [9]. The method is to Gaussian blur the field $\sqrt{p^2+q^2}$ derived from each image using the relationship $\sqrt{p^2+q^2} = \overline{R}^{-1}(E)$, where $\overline{R}$ is a circularly symmetric approximation to the reflectance function. The blurred image is obtained using $E_b = \overline{R}(\sqrt{p^2+q^2})$. This works well if the light source direction is reasonably close to the viewing direction.

At larger scales the images and the surface contain only low frequency information, so the efficiency of the shape-from-shading algorithm can be improved by subsampling the images and reducing the number of grid points used to describe the surface.

9 Recovering the Illumination Direction

To recover the illumination direction, we made the direction vector a variable in the optimization process. However, we found that the approach did not work well until an initial estimate for the surface shape had been obtained. We therefore kept the illumination vector constant (vertical light) during the first few iterations of the algorithm. This worked well for both synthetic and real images, and in tests with synthetic images the value obtained for the illumination direction was always within 0.05 radians of the true direction.

9 Experimental Results

Results are shown for two real image pairs and for two synthetic image pairs (with and without added Gaussian noise). The synthetic images were 128 by 128 pixels and were shaded using the Lambertian model, $R = \hat{\mathbf{n}} \cdot \hat{\mathbf{l}}$. The real images were 256 by 256 pixels.

The shape factor (σ value) for the Gaussian basis functions was changed logarithmically, using $\sigma_k=0.9\sigma_{k-1}$. This schedule works well in practice and is not critical. After each change of basis, conjugate gradient descent was performed for up to 1000 iterations before the next change of basis. The algorithm was stopped when σ_k dropped below a value of 1 (in pixel units).

Figure 1 is a pair of synthetic images constructed from the surface shown in Figure 2 with a random surface texture superimposed. Figure 3 shows the reconstructed surfaces obtained using these and a similar pair of images with 10% additive noise.

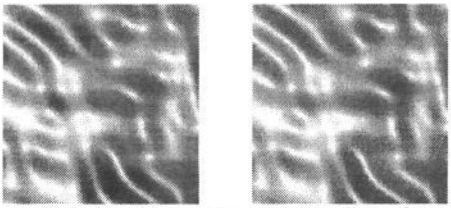


Figure 1: Left and right images of "wavy" surface, noiseless

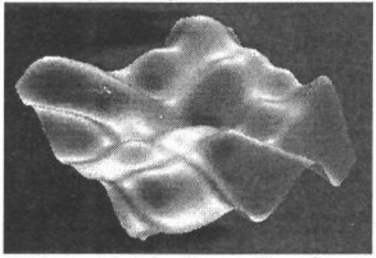
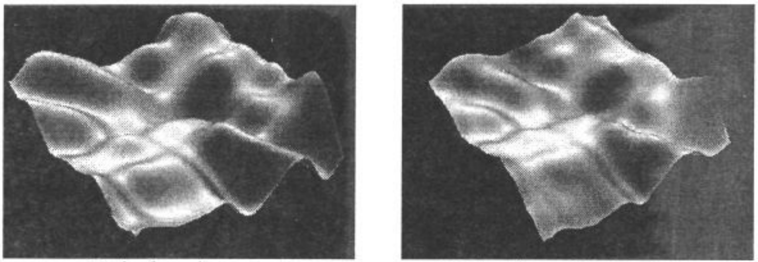


Figure 2: The "wavy" surface



Noiseless image

10% noise added to image

Figure 3: Surface recovered from images of "wavy" surface

Figure 4 shows two synthetic images of a hemisphere on a flat background lit from directly above. Figure 5 shows a view of the model used to create the images, and Figure 6 shows the reconstructed surfaces for noiseless and noisy image pairs.

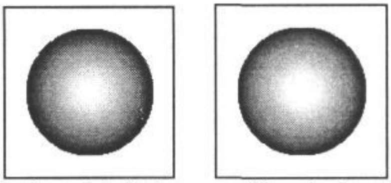


Figure 4: Left and right images of hemisphere (noiseless)

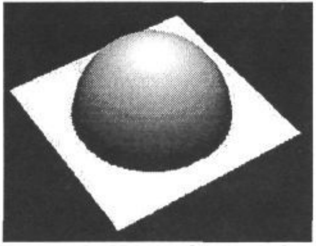
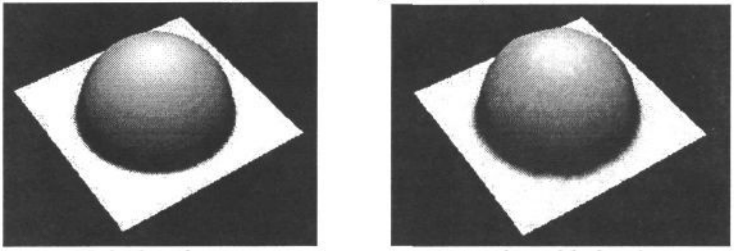


Figure 5: Hemisphere surface



Noiseless images 10% noise added to images

Figure 6: Reconstructed surface of hemisphere

Figure 7 is an SEM image pair of a cylindrical fibre taken by rotating the sample stage by an angle of approximately 0.1 radians; the reconstructed surface is shown in Figure 8. Figure 9 is a visible light image pair of a stone taken using a rotation of 0.1 radians. Figure 10 shows the reconstructed surface (assuming Lambertian shading).

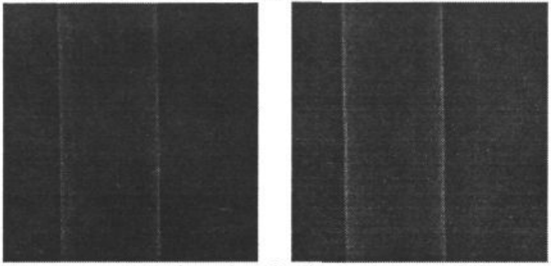


Figure 7: Left and right SEM images of fibre

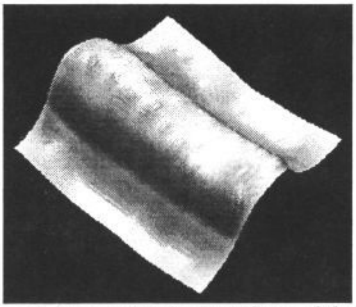


Figure 8: Reconstructed surface of the fibre

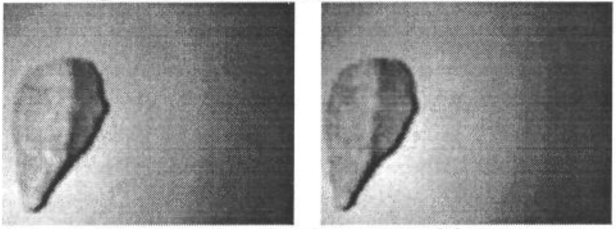


Figure 9: Left and right images of the stone

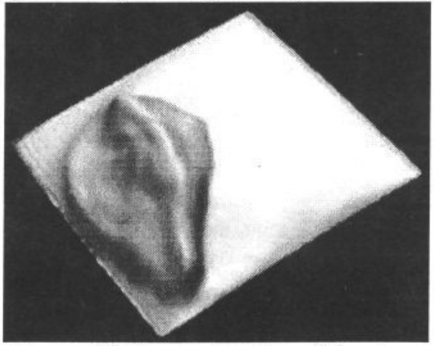


Figure 10: Reconstruction of the stone

The following table gives the brightness and depth errors of the reconstructions (depth error as a percentage of the maximum height of the surface used to create the images).

Surface	Brightness error, %		Depth error, % of max height	
	mean	s.d	mean	s.d
"wavy", noiseless	1.10	8.35	1.41	2.53
"wavy", 10% noise	1.87	15.23	4.27	5.65
Hemisphere, noiseless	1.08	6.53	0.83	0.72
Hemisphere, 10% noise	2.02	13.80	5.27	7.61
SEM image of fibre	4.83	11.55	unknown	unknown
Image of stone	6.02	21.25	unknown	unknown

10 Summary

We have described a novel algorithm for reconstructing a surface using both shape-from-shading and stereo evidence. The two types of evidence are combined within a Bayesian framework, and the surface reconstruction is achieved by minimizing a simple objective function. To avoid problems with local minima, the objective function is minimized using a scale space tracking algorithm. The scale space representation of the solution is constructed from Gaussian basis functions, a representation which incorporates the smoothness assumption and the integrability constraint in a natural way. Our experimental results show that the algorithm is extremely robust, giving good results even after considerable noise has been added to the images.

References

1 Horn, B K P "The Variational Approach to Shape from Shading", *Computer Vision, Graphics and Image Processing*, **33** (1986), pp174-208.

2 Dupuis, P and Oliensis, J "Direct Method for Reconstructing Shape from Shading", *IEEE Computer Vision and Pattern Recognition Conference*, Champaign IL (1992), pp453-458.

3 Pentland, A P "Local Shading Analysis", *IEEE PAMI*, **6** (1984), pp170-187.

4 Bruss, A R "The Eikonal Equation: Some Results Applicable to Computer Vision", *Journal Mathematical Physics*, **5** (1982), pp890-896.

5 Jones, A G and Taylor, C J "Robust Shape from Shading", *Image and Vision Computing*, **12** (1994), pp411-421.

6 Chellapa, R and Frankot, R T "A Method for Enforcing Integrability in Shape from Shading Algorithms", *First Int. Conf. on Computer Vision*, London (1987), pp118-127.

7 Horn, B K P "Height and Gradient from Shading", *AI memo No 1105*, MIT AI Laboratory (May 1989).

8 Szeliski, R "Fast Shape from Shading", *First Euro. Conf. Comp Vis*. Antibes, France (1990).

9 Peleg, S and Ron, G "Nonlinear Multiresolution: A Shape-from-Shading Example", *IEEE Transactions PAMI*, **12** (1990), pp1206-1210.

10 Shao, M, Simchony T and Chellapa R "New Algorithms for Reconstruction of a 3-D Depth Map from One or More Images", *Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Ann Arbor (1988), pp530-535.

11 Ellison, T P and Taylor, C J "Calculating the Surface Topography of Integrated Circuit Wafers from SEM Images", *Image and Vision Computing*, **9** (1991) pp3-9.

12 Ellison, T P. Personal correspondence with the author.

13 Lee, S and Brady, M "Integrating Stereo and Photometric Stereo to Monitor the Development of Glaucoma", *British Machine Vision Conference*, Oxford, UK (1990), pp193-197.

14 Hartt, K and Carlotto, M "A Method for Shape-From-Shading Using Multiple Images Acquired Under Different Viewing and Lighting Conditions", *IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, San Diego (1989), pp53-60.

15 Ikeuchi, K "Constructing a Depth Map from Images", *AI Memo No 744*, MIT AI Laboratory, Cambridge, MA, (1983).

16 Grimson, W E L "Binocular Shading and Visual Surface Reconstruction", *Computer Vision, Graphics and Image Processing*, **28** (1984), pp19-43.

17 Hougen, D R and Ahuja, N "Estimation of the Light Source Distribution and its Use in Integrated Shape Recovery from Stereo and Shading", *Fourth International Conf. Computer Vision*, Berlin (1993), pp148-155.

18 Goldstein, G I. *Scanning electron microscopy and X-ray Microanalysis*. Plenum Press. New York, (1981).

19 Jones AG. *Recovering 3D shape from 2D images*. PhD thesis, University of Manchester, 1995.

20 Witkin, A, Terzopoulos, D, and Kass, M "Signal Matching Through Scale Space", *Fifth Int. Conf. AI* (1986), pp714-719.

21 Witkin, A P "Scale Space Filtering", *Eighth International Joint Conf. AI*, (1983).

22 Whitten, G "Scale Space Tracking and Deformable Sheet Models for Computational Vision", *IEEE PAMI* **15** (1993), pp697-706.

23 Lindeburg, T "Scale Space for Discrete Signals", *IEEE Trans. PAMI*, **12** (1990), pp234-254.

24 Gill, P E and Murray, W *Practical Optimization*, Academic Press, New York (1981).