

Multiresolution Active Contour Models in Textured-Stereo Images.

A. E. Grace and D. Pycock,
School of Electronic and Electrical Engineering,
The University of Birmingham,
email: A.E.Grace@bham.ac.uk.

Abstract

This paper presents a method for generating sparse range data from textured surfaces which have structured light projected onto them. The work is motivated by the need to measure 3-D road defects rapidly and reliably. Traditional approaches to computing range from stereoscopic images have relied on either smooth or finely textured surfaces when using structured light. Conventional techniques that take advantage of the inherent texture in the images are not applicable. This is because corresponding stereoscopic road surface views are dissimilar due to the geometry of the cameras and the surface texture. The method described places initial edge points in a low resolution version of the intensity image. These points are used to initialise open active contour models or snakes which are propagated via a pyramid to a higher resolution. At this higher resolution, internal and external constraints are applied to the snake; the internal constraint being a smoothness functional and the external one being based on a maximum likelihood estimate of the edge strength across each light stripe. Computation is spatially localised at each stage and thus this algorithm could easily be parallelised.

1. Introduction

This paper describes a stereo multiresolution algorithm using active contour models, prior information and an edge model. The algorithm generates active contours that are refined over scale. The smoothness of the depth data is determined by a single regularisation parameter.

1.1. Motivation

The problems associated with generating depth data from a pair of stereo images are well known [1]. Arguably the most difficult task is establishing correspondence between features [2] that provides depth information. Motivation for this work stems from the need to improve on the poor quality data obtained in manually surveying road surfaces [3]. Inaccurate data from the surveys is believed to be a major cause of costly errors in road repair even when expert systems are used [4].

1.2. Related Work

Projecting light onto a 3-D surface to determine corresponding points in stereoscopic image pairs was first discussed by Will and Pennington [5,6] who located planar faces of polyhedral objects. Potmesil [7] described a heuristic approach in which the images included a set of calibration marks so that camera parameters could be estimated. In 1986 Hu, Jain and Stockman [8] projected a square grid onto a smooth surface and modelled the resulting squares as textures. The pattern elements were classified in terms of the surfaces they formed, such as: planar, non-planar, convex and non-convex, using a nearest neighbour technique. Both methods worked well but relied on the underlying surfaces being smooth.

Other approaches have included encoding the light pattern [9] and projecting three different patterns (generating three different images) [10]. The later approach involves recovering unique-corresponding points (for stereo images) from a combination of three images. Clearly this approach is not applicable if the target scene or imaging system is moving.

An algorithm that combines a passive illumination method with multiple resolutions was developed by Tate and Lai [11]. In this paper it was noted that occlusion and correspondence are the main problems. Laser range data combined with intensity at a coarse resolution was used to ensure correspondence across multiple scales. Others have described the use of projected light patterns at multiple scales [12], but not, to the best of our knowledge, the construction of a depth map using data from multiresolution images. The correspondence problem is addressed using the laser range data and where this does not exist, an interpolation method. Further, several ad-hoc thresholds are incorporated to ensure that the newly generated data is smooth. Related to this work by the use of multiple resolutions, is the work of Bajcsy and Kovacic [13] on multiresolution elastic matching. This work fits shape models to multiresolution data, the fit being performed in terms of best fit to a trained model.

1.3. Contributions of this work

- Projected light stripes combined with model based constraints to interpret stereo images without resorting to restrictive object models.
- A careful formulation of the model energy constraints. The internal energy constrains the snake to a straight line while the external energy is based on a maximum likelihood function of the edge that has been shown to be well behaved with respect to edge strength [14]. This forces the snake towards highly likely edges.
- The problem of initialising the snakes is performed using a matched filter and prior knowledge of stripe width and number of stripes. The matched filter and use of a pyramid data structure apply geometric constraints that adapt progressively as the interpretation is refined.
- No initial estimates of depth are required as in [11]. It is unlikely that a laser technique would work in this case due to the coarse structure of the road surface.
- The use of a single regularisation parameter to control depth data continuity. Post processing ([15] for example) is unnecessary since the depth data is smooth and discontinuities are preserved.

2. Algorithm Overview

This section gives a brief overview of the algorithm developed to establish edge points in images with projected light. The number of stripes, N_s , is known a priori as is the width of the stripe. Figure 1 shows an outline of the algorithm. On the left hand side of the figure the original image is sub-sampled to form lower resolution intensity images. At layer 2, candidate edge nodes are generated by convolving with a matched filter that is the width of the expected stripe. The strongest N_s responses are selected to form the candidate edge map. Each edge along a stripe forms an active contour model (snake), thus there are $2N_s$ snakes per image. Each snake is propagated to layer 1 and then optimised over a limited spatial domain constrained by internal and external energy functions together with a regularisation parameter. When the optimum snake has been found it is propagated to the next higher resolution (layer 0) and optimised again. The result of this stage is a set of corresponding edge points for each stereo pair.

Sampling rather than a Gaussian pyramid is used because: i) random noise is not a cause for concern and ii) to avoid the increased computational cost associated with the generating Gaussian data. Whilst this breaks with the concept of scale-space [16] it does not undermine the principle of the proposed algorithm. Here the main issue is not the preservation of a well ordered signal in scale-space but rather the initialisation of a boundary detection process. This approach is justified because the artifacts present in the image are not random signals but associated with the structure being analysed.

prior knowledge

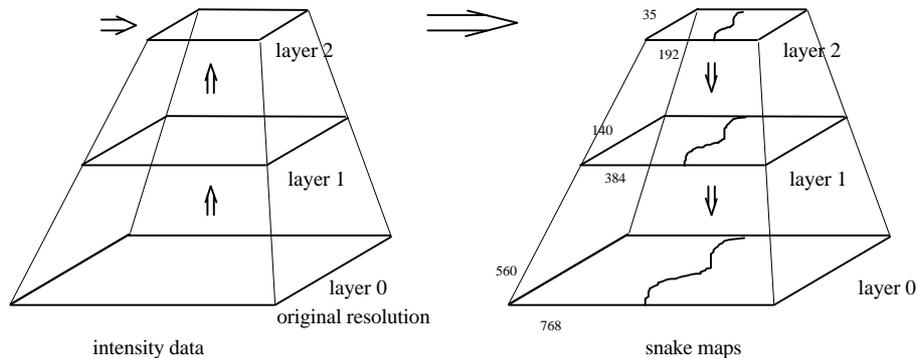


Figure 1. Overview of the algorithm, on the left hand side intensity data is sub sampled to provide lower resolutions of the original data. On the right hand side, snakes are formed from an initial convolution placement scheme.

3. Algorithm Detail

This section describes four parts of the algorithm: 1) the pyramid data structure, 2) the snake formulation, 3) initial edge placement (snake initialisation), 4) snake propagation and 5) snake optimisation and update strategy.

3.1. Pyramid structure

A three level intensity image pyramid is generated by sampling to reduce the vertical resolution by four and the horizontal resolution by two at each step. The lowest resolution image is 35 by 192.

3.2. The active contour model

The active contour model introduced by Kass et al. [17] is based on the minimisation of energy functions. Generally these functions are heuristic and of an arbitrary form and frequently formulated as energy functions. Active contour models do not require these heuristics to be energy terms but just to have a suitable form. In this work the internal and external heuristic functions are based on the cosine of an angle and the maximum likelihood probability of an edge. Given a set of nodes \mathbf{V} that form a snake, then the objective is to minimise

$$e(\mathbf{V}) = \sum_{i=1}^n \lambda_i E_{\text{int}}(v_i) + (1 - \lambda_i) E_{\text{ext}}(v_i) \quad \text{Eq. 1.}$$

where $\mathbf{V} = \{v_i = (x_i, y_i) : i = 1, 2, \dots, n\}$,

E_{int} is a smoothness constraint,

E_{ext} is the external energy, such as edge strength,

and λ are the regularisation parameters.

As noted by Lai and Chin [18], the regularisation parameters will greatly influence the result from being dominated by the external energy when $\lambda \ll (1 - \lambda)$ to being dominated by the internal energy or smoothness when $\lambda \gg (1 - \lambda)$ which in this case would force the snake to a straight line. Setting the regularisation parameter is a formidable task and despite efforts of many researchers (see reference [18] for details) the problem is unsolved for the case when the objective is to position the snake so that it reflects the strength of each energy at each node. The problem is not solved by the method discussed in [18] where a minimax approach is taken to the automatic-implicit selection of λ . Lai and Chin's method depends on the normalisation of the internal and external energies, therefore weak edges are weighted the same as strong edges.

The approach adopted in this work is based on the assumption that external energy terms should be adaptively weighted. Thus weak edges have a reduced affect on the snake. In this work the internal energy is defined as

$$E_{\text{int}} = (\cos(\theta) + 1) / 2$$

where θ is the angle defined in Figure 2 as the smallest angular difference between the lines subtended between the vector pairs (v_{i-1}, v_i) and (v_i, v_{i+1}) . Thus $E_{\text{int}} = 0$ when (v_{i-1}, v_i) and (v_i, v_{i+1}) form a straight line and tends to 1 otherwise.

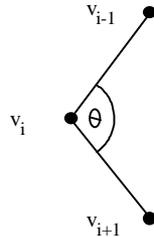


Figure 2. Definition of angle used in the smoothness constraint.

The external energy term is based on the maximum likelihood of the edge at v_i such that $E_{\text{ext}}(v_i) = 1 - L_n(v_i)$. The edge is assumed to be vertical and therefore the likelihood estimate is made in the horizontal plane. The normalised likelihood of the edge at node v_i is defined as

$$L_{\text{norm}}(v_i) = \frac{L(v_i)}{L_L(v_i)}$$

where $L_L(v_i) = \max[L(v_i) \forall i \in \mathbf{L}]$ where \mathbf{L} is a local support region defined as $\mathbf{L} = \{v_i - 2, v_i - 1, v_i, v_i + 1, v_i + 2\}$. Note that when $L_{\text{norm}}(v_i) = 1$ the edge likelihood is a maximum. The derivation of the maximum likelihood estimate, $L()$, is given in the Appendix.

3.3. Initialising the snake

The snakes are initialised by finding the maximum N_s responses to the convolution operation. Since stripes are projected vertically, the intensity image is convolved with a horizontal top-hat function \mathbf{F}_{w_2} of width w_2 , where w_2 is the expected width of the projected light stripe at the resolution level 2, in this case $w_2=5$. Let \mathbf{C}_p contain all snakes at level p . Thus the initial snake values at level 2, \mathbf{C}_2 , are based on the result of

$$\mathbf{C}_2 = \mathbf{G}_2 * \mathbf{F}_{w_2}$$

where $*$ indicates the convolution operation. The initial snake nodes are selected as the largest N_s responses in \mathbf{C}_2 . These responses represent the right hand edges of the stripes, the left hand edges are assumed to be at the position $\mathbf{C}_2(x_i, y_i - w_2)$ for a right hand snake node at $\mathbf{C}_2(x_i, y_i)$.

3.4. Propagating snakes in the Pyramid

Once the snakes have been initiated at \mathbf{C}_2 , they are propagated to the next highest resolution \mathbf{C}_1 since the resolution at layer 2 is such that any refinement is unlikely to produce significant improvement. A snake node at v_i in \mathbf{C}_2 represented $\mathbf{C}_2(v_i) = \mathbf{C}_2(x_i, y_i)$ propagates to position $\mathbf{C}_1(v_j) = \mathbf{C}_1(k_x x_i, k_y y_i)$, that is, the position x in the lower resolution projects to a new position in the higher resolution which is simply $x \cdot k$ where k is the scaling factor. In addition to this, parent nodes in \mathbf{C}_2 have a range of child nodes in \mathbf{C}_1 (similarly parent nodes in \mathbf{C}_1 have child nodes in \mathbf{C}_0) that are the possible sites that the snake node could move to. These are defined for node v_i as the set \mathbf{S} where $\mathbf{S} = \{(x_i, y_{i-1}), (x_i, y_i), (x_i, y_{i+1}), (x_i, y_{i+2})\}$. This is as defined in [19] and represents a 50% overlap in y compared with the intensity sampling. Note that the freedom in the snake is only in the y direction (horizontally) reflecting the expected orientation of the edge.

3.5. Optimisation and update strategy

For each snake \mathbf{V} there is an associated visit set \mathbf{T} . Initially the procedure to minimise Equation 1 starts at v_1 and visits each snake node. Since minimising each node only uses the local neighbourhood, minimising Equation 1 is not explicitly required. Thus when a snake node v_i is updated then only nodes v_{i-1} and v_{i+1} are affected. If v_i is not updated then the visit vector $t(i)$ is zeroed. If v_i is updated then $t(i-1)=t(i)=t(i+1)=1$. After the first complete iteration of the snake, a new site to be minimised is selected as

$$i = \arg \max_i (e(v_i) t_i).$$

Thus the snake node with the maximum $e(v_i)$ value is visited since this site represents the worst site in terms of fitting the snake model. Finally when $T=0$ the snake has converged to a minimum.

The update rule for each snake node is different from that described in [18]. Here a node is updated by performing a weighted-linear interpolation between the maxima of the internal and external energies. For a node v_i define v_{int} and v_{ext} as

$$v_{\text{int}} = \arg \min_i (\lambda E_{\text{int}}(v_i)) \quad \text{and} \quad v_{\text{ext}} = \arg \min_i ((1-\lambda) E_{\text{ext}}(v_i)) \quad \forall i \in \mathbf{S}.$$

Next update the node v_i as $v_i = \hat{v}_i + \Delta v_i$ such that

$$\hat{v}_i = \frac{v_{\text{int}} + v_{\text{ext}}}{2} \quad \text{and} \quad \Delta v_i = \frac{\Delta v_{\text{int}}(1 - E_{\text{int}}(v_{\text{int}})) + \Delta v_{\text{ext}}(1 - E_{\text{ext}}(v_{\text{ext}}))}{(1 - E_{\text{int}}(v_{\text{int}})) + (1 - E_{\text{ext}}(v_{\text{ext}}))}$$

where $\Delta v_{\text{int}} = \hat{v}_i - v_{\text{int}}$ and $\Delta v_{\text{ext}} = \hat{v}_i - v_{\text{ext}}$.

Note that if $E_{\text{int}}(v_{\text{int}}) = 1$, $E_{\text{ext}}(v_{\text{ext}}) = 0$, $\Rightarrow \Delta v_i = \Delta v_{\text{ext}}$

if $E_{\text{int}}(v_{\text{int}}) = 0$, $E_{\text{ext}}(v_{\text{ext}}) = 1$, $\Rightarrow \Delta v_i = \Delta v_{\text{int}}$

if $E_{\text{int}}(v_{\text{int}}) = E_{\text{ext}}(v_{\text{ext}}) \Rightarrow \Delta v_i = \frac{\Delta v_{\text{int}} + \Delta v_{\text{ext}}}{2} = 0$.

Figure 3 shows the complete algorithm.

```

Generate intensity images  $\mathbf{G}_1$  and  $\mathbf{G}_2$  from  $\mathbf{G}_0$ 
Estimate edge locations for snake initialisation at layer  $\mathbf{C}_2$ 
Step 1 propagate snakes to next higher resolution
  Do select each snake  $s$ 
    Do Minimise each node
      Generate visit set  $\mathbf{T}$ 
    Until all nodes in  $s$  have been visited
    Do Select a new node in snake  $s$ 
      Minimise the node
      Update  $\mathbf{T}$ 
    Until  $\mathbf{T}=0$ , that is, snake  $s$  has converged
  Until all snakes have converged
If not layer 0
  Goto step 1
end

```

Figure 3. The snake minimising algorithm.

4. Results

The algorithm has been tested on images of a laboratory road surface containing a pot-hole. In this work the baseline was at 2 metres, the camera separation distance was 1.3 metres and the area imaged was approximately 0.5m x 0.7m. Figure 4 shows the stereo pair and Figure 5 shows the depth map reconstruction from the sparse data set.

In this figure, the regularisation parameter was set to $\lambda = 0$, that is, the snake was optimised with only the external edge parameter. In Figure 6 the same view is shown except that the regularisation parameter has been set to $\lambda = 0.6$. Note that there is some degree of variation in the areas around the hole which initially may be expected to be a flat surface. However, the surface of the test pot-hole is not flat because of the coarse granularity of the surfacing material. Subjectively, the result obtained is an accurate interpretation; a detailed evaluation is planned. It can be seen in Figure 5 that there are a number of depths that are probably not correct. In Figure 6 the same view is shown except that the regularisation parameter has been set to $\lambda = 0.6$. Clearly the resulting reconstruction is smoother than that shown in Figure 5. In Figure 7 the dominant error in depth estimation is shown. The continuous plot is from Figure 5 and the dashed line is from Figure 6. The difference that the regularisation makes can be seen. The cause of this and similar errors is a node in the snake that has been incorrectly placed. Figures 8 and 10 both show the same section of one image from the stereo pair with the snake nodes superimposed. Figure 8 shows the node that causes the large error shown in Figure 5 within the boxed area and in Figure 9 it can be seen that the error has been corrected.

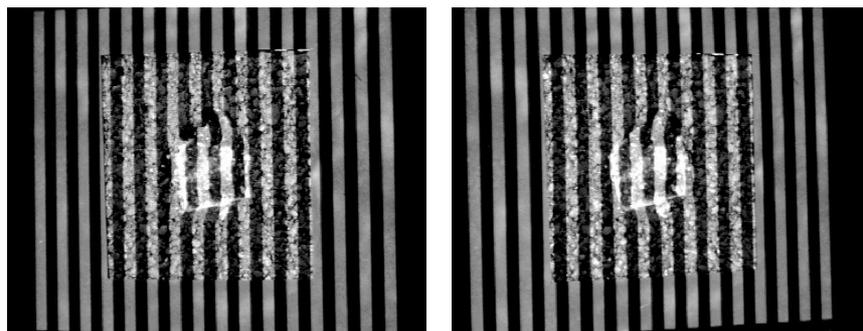


Figure 4. Stereo pair of full resolution images with projected light stripes.

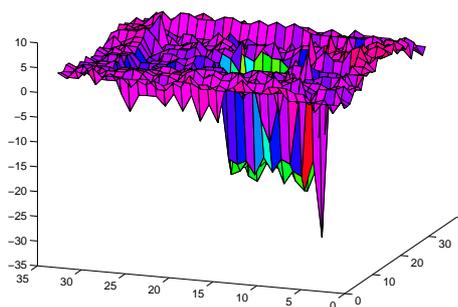


Figure 5. Reconstruction with $\lambda=0$.

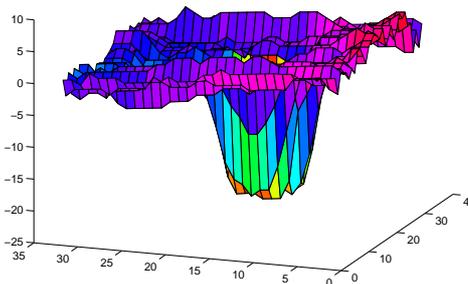


Figure 6. Reconstruction with $\lambda=0.6$

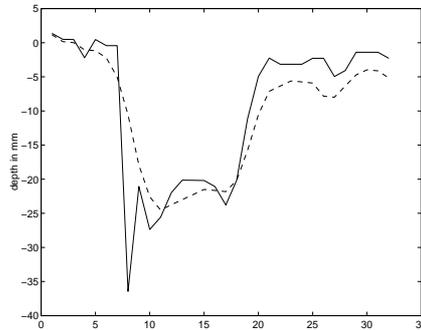


Figure 7. Plot showing the differences between constrained (dashed) and non-constrained (full line) depth information.

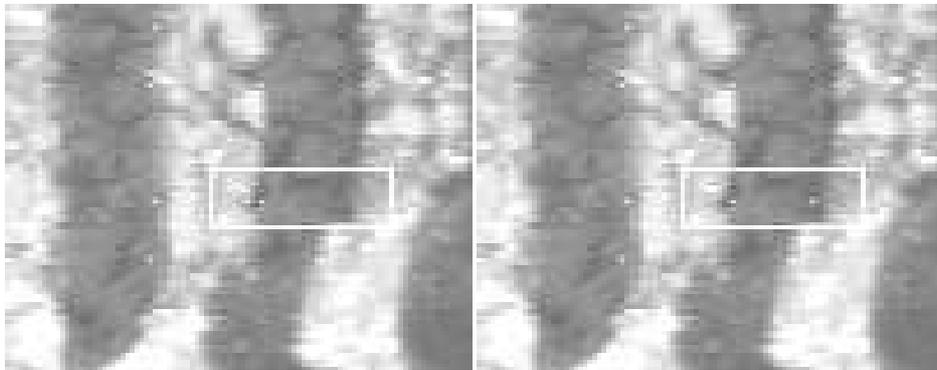


Figure 8. Node placement $\lambda=0$.

Figure 9. Node placement $\lambda=0.6$.

5. Conclusions

This paper has shown that a multiresolution snake can be incorporated into a 3-D reconstruction scheme that uses light stripes and prior knowledge regarding the number of stripes and their geometry. In this approach, knowledge of the projected pattern is used to constrain the search for corresponding points in a stereo image pair. The snake energy terms constrain how the position of initial points are altered in the higher resolution images. As a result, there is an implicit constraint on the corresponding points in each image which enforces a degree of continuity in the depth map. It has been shown that interpretations generated from a strategy that does not include a smoothness constraint is likely to contain errors due to local minima in the intensity images and that the inclusion of constraints provides a reliable 3-D reconstruction of the shape for an irregular object. Further work will include an initial node placement scheme that overcomes the problem of occluded stripes.

6. References

- [1] Jepson, A. D. and Jenkin, M. R. M., 'The fast computation of disparity from phase differences', in Proc CVPR '89: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 1989, pp. 398-403.

- [2] Barnard, S. T. and Thompson, W. B., 'Disparity analysis in images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2, 1980, pp. 333-340.
- [3] Hintz, R. J., Karakadas, C. and Kang, J., 'Analysis of pavement cracking and rutting using close range photography', *Photogrammetric Engineering and Remote Sensing*, **55**, 2, 1989, pp. 217-221.
- [4] The highway design and maintenance standards model, Volume 1, 'Description', 1987, John Hopkins University Press.
- [5] Will, P. M. and Pennington, K. S., 'Grid coding, a preprocessing technique for robot and machine vision', *AI*, **2**, (3/4), 1971, pp. 319-329.
- [6] Will, P. M. and Pennington, K. S., 'Grid coding, a novel technique for image processing', *Proc. IEEE*, **60**, (6), 1972, pp. 669-680.
- [7] Potmesil, M., 'Generation of 3-D surface descriptors from images of pattern illuminated objects', in *Proc. IEEE conf. Pattern Recognition and Image Processing*, Chicago, IL, August, 1979, pp. 553-559.
- [8] Hu, G., Jain, A. K. and Stockman, G., 'Shape from light stripe texture', in *Proc CVPR '86: IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami Beach, FL, USA, 22-26 June, 1986, pp. 412-414.
- [9] Vuylsteke, P. and Oosterlinck, A., 'Range image acquisition with a single binary-encoded light pattern', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 2, 1990, pp. 148-164.
- [10] McDonald, J. P., Lambert, R. and Fryer, R. J., '3-D measurement using stereo scene coding', *Colloquium on 3D Imaging and Analysis of Depth/Range Images*, IEE, Digest number 1994/054, 1994.
- [11] Tate, K. and Lai, Z., 'Depth map construction from range-guided multiresolution stereo matching', *IEEE Transactions on Systems, Man and Cybernetics*, **24**, 1, 1994, pp. 134-144.
- [12] McDonald, J. P., Siebert, J. P. and Fryer, R. J., 'A new approach to active illumination', *BMVC 1991*, pp. 210-216.
- [13] Bajcsy, R. and Kovacic, S., 'Multiresolution elastic matching', *Computer Vision, Graphics and Image Processing*, **46**, 1989, pp. 1-21.
- [14] Zhou, P. and Pycock, D., 'Robust Model-Based Boundary Cue Generation for Cell Image Interpretation', *BMVC 1995*, Birmingham, England, pp. 337-356.
- [15] Terzopoulos, D. and Metaxas, D., 'Dynamic 3D models with local and global deformations: deformable superquadrics', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 7, 1991, pp. 703-714.
- [16] Blom, J., Bart, M., Romney ter Haar, Bel, A. and Koenderink, J. J., 'Spatial derivatives and propagation of noise in Gaussian scale space', *Journal of Visual Communication and Image Representation*, **4**, 1, 1993, pp. 1-3.
- [17] Kass, M., Witkin, A. and Terzopoulos, D., 'Snakes: active contour models', *International Journal of Computer Vision*, 1988, pp. 321-331.

- [18] Lai, K. F. and Chin, R. T., 'On regularization, formulation and initialisation of the active contour models (snakes)', First Asian Conference on Computer Vision, Osalsa, 1993, pp. 542-545.
- [19] Spann, M. and Grace, A. E., 'Adaptive segmentation of noisy and textured images', **27**, 12, 1994, pp. 1717-1733.
- [20] Fukunaga, K., 'Introduction to statistical pattern recognition', Academic Press, London, England, 1972.

Appendix: Maximum Likelihood Edge Detection

The maximum likelihood principle [20] can be used to estimate the likelihood of a given data set being drawn from a particular distribution. In this case it has been applied as an edge detector. The set of grey levels W , is partitioned into two by m . The likelihood, L , is given by

$$L(W; m) = \prod_{j=1,2} \prod_{i \in R_j} p(w_i, m) = p(R_1)p(R_2)$$

where $R_1 = \{w_i : i = 1, i \leq m\}$, $R_2 = \{w_i : i > m, i \leq |W|\}$.

Taking logarithms

$$\log L(W; m) = \sum_{j=1,2} \sum_{i \in R_j} \log p(w_i, m)$$

Assume a Gaussian distribution for R_1 and R_2 gives

$$\log L(W; m) = \sum_{j=1,2} -\frac{|R_j|}{2} \log(2\pi\sigma_j^2) + \sum_{j=1,2} \sum_{i \in R_j} -\frac{1}{2} \left(\frac{\bar{w}_j - w_i}{\sigma_j} \right)^2$$

If σ is dependent on j then

$$\log L(W, m) = \sum_{j=1,2} -\frac{|R_j|}{2} \log(2\pi\sigma_j^2) - \sum_{j=1,2} \frac{1}{2\sigma_j^2} \sum_{i \in R_j} (\bar{w}_j - w_i)^2$$

$$\text{but } \sigma_j^2 = \frac{1}{|R_j|} \sum_{i \in R_j} (\bar{w}_j - w_i)^2.$$

$$\text{Therefore } \log L(W, m) = \sum_{j=1,2} -\frac{|R_j|}{2} (\log(2\pi\sigma_j^2) + 1).$$

For the complete test, the value of m is found that maximises the likelihood. To make the likelihood a continuous function, the standard deviation is incremented by one so that the range of the log likelihood function is between 0 and infinity. Thus given a candidate edge at position, m , the refined edge position \hat{m} is found as

$$\hat{m} = \arg \max_m \{\log L(W, m)\}$$

where

$$W = \{m - \Omega / 2, \dots, m + \Omega / 2\} \text{ and}$$

$$m \in L.$$

$$\text{In this work } \Omega \leq wp, \Omega_1 = 16 \text{ and } \Omega_0 = 32.$$

Note that W is generated as averages in the vertical plane, x , such that at levels 1 and 0, $x_1=5$ and $x_0=7$.
