NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Parameter estimation for a deformable template model

by

Merrilee A. Hurn, Ingelin Steinsland and Håvard Rue

PREPRINT
STATISTICS NO. 5/1999

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/1999/S5-1999.ps
Håvard Rue has homepage: http://www.math.ntnu.no/~hrue
E-mail: hrue@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491
Trondheim, Norway.

Parameter estimation for a deformable template model

Merrilee A. Hurn
Mathematical Sciences
University of Bath, U. K.

Ingelin Steinsland and Håvard Rue
Department of Mathematical Sciences
NTNU, Norway

February 12, 1999

Summary

High–level image analysis is an area which has seen considerable research interest in the past few years. In particular the task of object recognition has been successfully tackled using deformable template models in a Bayesian context. However these models are not being widely used in practise, and one of the reasons for this may be that they do require a number of rather non-intuitive parameters to be specified. We propose an approach which treats some parameters in a fully Bayesian way while estimating the remaining parameters from training data generated by the user outlining cells on a computer screen. This requires no expert statistical input. We illustrate the estimation of the parameters for a deformable template model using images generated by a confocal microscope.

Keywords: Bayesian image analysis; Confocal microscopy; Deformable templates; Object recognition; Markov chain Monte Carlo; Parameter estimation.

Address for Correspondence: Merrilee Hurn, Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, U. K.

E-mail: M.A.Hurn@maths.bath.ac.uk and havard.rue@math.ntnu.no

Acknowledgements: MAH and HR are grateful to the TMR Spatial and Computational Statistics Network (ERB-FMRX-CT96-0096). We thank Dr N. White and Dr R. Errington of the Plant Sciences Department of Oxford University for providing the confocal data.

1 INTRODUCTION

The use of high-level models in image analysis has increased in popularity during the last few years as the problem of interpretation and understanding of images has become more important. One specific task for which such models have been proposed is object recognition, locating objects in a scene and providing estimates of their type, position and outline. The confocal microscopy image in Figure 1 is one such application, showing an optical section through cartilage growth plate in which two types of cell are present; one stage of the analysis of these data is to identify and distinguish cells on the basis of their shape and size. Manual segmentation of each cell would be a tedious and time consuming task, especially as this image forms just one of a large number of similar images, and so there is a great need for some automated procedure. A number of different approaches, both statistical and non–statistical, exist for tackling the task. In this paper we concentrate on a high-level statistical model which represents each cell by a polygon outline model embedding these deformable templates into a marked point process to handle the unknown number of objects. A point process of stochastic geometrical objects is thus constructed which is conditioned on the observed data, i.e. the data image, and simulation techniques are then used to generate samples from this posterior distribution. These samples are used to estimate quantities such as the number of objects, their type, outline and position.

Although there has been much progress in understanding such models recently, they remain somewhat inaccessible in part due to the difficulty in specifying many of the parameters associated. Further work is needed before microscope users could routinely use these techniques for automatic extraction of information. A significant step towards this goal would be a way of automating parameter estimation, both for the parameters describing the stochastic variations in shape for the objects and for the parameters in the data model. The aim of this paper is to provide such a solution in a way that the only interaction between the microscope user and the estimation procedures is purely graphical; the end-user supplies computer–drawn templates of the cells and the training data for parameter estimation then take the form of polygons fitted on the screen to a display of image data.

The plan of the paper is as follows: In Section 2 we review the high–level deformable template model and we propose a likelihood model suitable for confocal microscopy applications. In Section 3 we briefly discuss inference from the model using reversible jump Markov chain Monte Carlo methods. We then discuss parameter estimation in Section 4, treating the image degradation parameters in a fully Bayesian way and using likelihood methods for the remaining parameters. We concentrate on deriving the maximum likelihood estimates of the deformation parameters under the assumption that the training data are recorded with measurement error due to the imperfect handtracking of individual cells. We conduct studies of this latter estimation problem in Section 5, using artificial data for which comparisons can be made with known values, and also using the cartilage data set.

2 A HIGH–LEVEL MODEL FOR OBJECT RECOGNITION

Adopting a Bayesian approach, we denote the scene by $\boldsymbol{x}$ and the recorded pixellated data by $\boldsymbol{y}$, the posterior of interest being given by $\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x})\,\pi(\boldsymbol{x})$. Since we are concerned with the components, i.e. the cells, of the scene, $\boldsymbol{x}$ will be specified as a "high–level" variable, composed of an unknown length list of the constituent components of the scene. This approach has been pioneered by Grenander and his co-authors, and recent sources for references can be found in Grenander & Miller (1994), Jain, Zhong & Lakshmanan (1996), Baddeley & Van Lieshout (1993) and Blake & Isard (1998).

2.1 THE OBJECT PRIOR $\pi(\boldsymbol{x})$

The prior model for a random configuration of cells is built around a stochastic template model for a single cell embedded in a mixture model of different cells types which is used as the mark distribution of a marked point process model.

2.1.1 THE TEMPLATE MODEL FOR A SINGLE CELL

The prototypical cell is described by an n–sided template defined by a set of vectors $\boldsymbol{g}_0, \boldsymbol{g}_1, \cdots, \boldsymbol{g}_{n-1}$ which give the edges of the polygon (Figure 2a). For example, if one type of cell is characteristically circular, then these edges describe a polygonal approximation to a circle. The closure of the polygonal template is equivalent to the condition $\sum_{i=0}^{n-1} \boldsymbol{g}_i = \boldsymbol{0}$. Described in this way, the template does not have any location information, and so we will consider its first vertex to be located at the origin, the second to be located at $\boldsymbol{g}_0$, the second at $\boldsymbol{g}_0 + \boldsymbol{g}_1$ and so on (Figure 2b). In order to accommodate scaling and rotational effects, the template may be globally scaled by a scalar R and rotated through an angle α (Figure 2c). To model natural shape variability occuring between cells of the same type, each edge $\boldsymbol{g}_i$ is subject to a stochastic deformation which incorporates the scaling R, the rotation α plus an additional edge–specific Gaussian deformation in length and direction (Figure 2d). This final edge-specific effect describes the change in length and direction between the undeformed $\boldsymbol{g}_i$ and the new edge prior to global scaling and rotation. Writing the deformed edge as $\boldsymbol{s}_i^{R\alpha}\boldsymbol{g}_i$ where $\boldsymbol{s}_i^{R\alpha}$ is the 2×2 matrix representing these changes, we thus have

$$\frac{1}{R}\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \boldsymbol{s}_i^{R\alpha}\boldsymbol{g}_i - \boldsymbol{g}_i = r_i \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix} \boldsymbol{g}_i.$$

Writing $t_i^{(0)} = Rr_i\cos(\alpha+\theta_i)$ and $t_i^{(1)} = Rr_i\sin(\alpha+\theta_i)$, determines that

$$\boldsymbol{s}_i^{R\alpha} = \begin{bmatrix} R\cos\alpha + t_i^{(0)} & R\sin\alpha + t_i^{(1)} \\ -R\sin\alpha - t_i^{(1)} & R\cos\alpha + t_i^{(0)} \end{bmatrix}.$$

Specifying the distribution of r_i and θ_i to be the angular and radial components of a bivariate Gaussian with zero correlation, then conditional on fixed values of R and α, $t_i^{(0)}$ and $t_i^{(1)}$ are independent

Gaussians with mean zero. Ignoring for a moment the constraint that the deformed template must be closed, i.e. $\sum_{i=0}^{n-1} \boldsymbol{s}_i^{R\alpha}\boldsymbol{g}_i = \boldsymbol{0}$, Grenander, Chow & Keenan (1991) suggest a first order cyclic Markov structure on the $\{t_i^{(0)}\}$ and the $\{t_i^{(1)}\}$ independently with each having an n–dimensional Gaussian distribution with mean $\boldsymbol{0}$ and circulant inverse covariance matrix incorporating the scaling

$$\boldsymbol{\Sigma}_R^{-1} = \frac{1}{R^2} \begin{bmatrix} \beta & \delta & & & & \delta \\ \delta & \beta & \delta & & & \\ & \delta & \beta & \delta & & \\ & & \ddots & \ddots & \ddots & \\ & & & \delta & \beta & \delta \\ \delta & & & & \delta & \beta \end{bmatrix}, \tag{1}$$

where all other entries are zero.

Define the length $2n$ vector $\boldsymbol{t} = (t_0^{(0)}, t_1^{(0)}, \ldots, t_{n-1}^{(0)}, t_0^{(1)}, \ldots, t_{n-1}^{(1)})^T$ then, still considering only the unconstrained non-closed polygon case

$$\boldsymbol{t} \sim N_{2n}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Sigma}_R & 0 \\ 0 & \boldsymbol{\Sigma}_R \end{bmatrix}\right). \tag{2}$$

Imposing the closure constraint on the deformed template will destroy the simple structure of (2); we will derive the distribution under the closure constraint in Section 4.2.2.

2.1.2 The marked point process for multiple objects

It is possible to allow objects of different types by forming a mixture model using different basic templates with different deformation parameters, where the weights in the mixture model represent the relative frequencies of occurrence; refer to Rue & Hurn (1999) for details. This mixture distribution is then used as the mark distribution of a marked point process model to model an unknown number of objects (Baddeley & Van Lieshout, 1993). Within this framework, each object x_i comprises a point which gives location (unspecified by the deformation model) and a set of marks which then specify its other attributes, in this case type and outline. The points lie in a window Λ related to the data observation coordinates, and the marks in the space M associated with the object types and shapes. A configuration of cells is described as a finite unordered set $\boldsymbol{x} = \{x_1, \ldots, x_k\}$ where $\boldsymbol{x}$ follows an object process, i.e. a marked point process on $\Lambda \times M$ with a Poisson object process as the basic reference process. Under the Poisson process, conditional on the number of objects k, the objects are uniformly and identically distributed. The joint density of $\{x_1, \ldots, x_k\}$ is defined by a density $f(\boldsymbol{x}, k)$ relative to the Poisson object process. For example, to model pairwise interactions between objects which are defined to be neighbours by some relation $\sim$, the function

$$f(\boldsymbol{x}, k) \propto \gamma^k \prod_{i \sim j} h(x_i, x_j), \quad \gamma > 0,$$

could be used; Van Lieshout (1995) discusses various interaction models. To model a situation where objects are not allowed to overlap, such as confocal microscopy where the images are optical sections, all objects are defined to be neighbours and the interaction function $h(x_i, x_j)$ is taken to be zero if objects x_i and x_j overlap and one otherwise. This model is known as the hard core object process. We take the point, denoted $\boldsymbol{c}$, to be the location of the first vertex for each object. The first mark is then a categorical variable defining the object type, and the remaining marks are the deformations $\boldsymbol{t}$ of the standard template for this type. We refer to Rue & Syversveen (1998) for more details.

2.2 A POISSON LIKELIHOOD MODEL $\pi(\boldsymbol{y}|\boldsymbol{x})$ FOR CONFOCAL MICROSCOPY

Confocal microscopy can produce relatively high–quality sectional images for certain specimen types. The technique works by scanning the focal point of a laser through the stained specimen, the fluorescence emitted from this focal point is then recorded, and the image is built up as a grid of fluorescent dye concentration measurements. The likelihood model for the data $\boldsymbol{y}$ approximates the microscope measuring fluorescence photon counts. We begin by assuming that cells all have a typical fluorescence level μ, while the background emits no fluorescence. The signal is measured on an 8 bit scale (i.e. on the integers 0 to 255) and in order to minimise clipping, the signal is operator adjusted at the experimental stage by a scale factor a and a shift b, so that the mean background value is now b, while the mean foreground value is $b + a\mu$ (and both are sufficiently far from the ends of the range of measurement). In addition, the recording process involves additive instrument white noise. We will approximate the record at pixel p by a Gaussian random variable with mean equal to the black–level b plus the expected scaled Poisson count for p, and with variance equal to the variance of the instrument noise σ^2 again plus the scaled expected Poisson count variance,

$$y_p \sim N(a\mu I_{[p\in\text{cell}]} + b, a^2\mu I_{[p\in\text{cell}]} + \sigma^2),$$

where $I_{[p\in\text{cell}]}$ is the indicator or whether the pixel is labelled as cell or background. The $\{y_p\}$ are assumed to be conditionally independent given the configuration $\boldsymbol{x}$.

3 INFERENCE USING MARKOV CHAIN MONTE CARLO

To do inference for the high-level object model, we have to resort to Markov chain Monte Carlo techniques (refer to Gilks, Richardson & Spiegelhalter (1996) for an introduction and list of references.) The key idea is to construct a Markov chain which has the posterior as the equilibrium density, and to estimate expected values by empirical average along the trajectory of a realization of the Markov chain. This makes it possible to access both estimated expected values and their uncertainty represented by the estimated variance of the estimated marginal density. To construct this Markov chain, we need a set of basic transition rules that will take us around in the state-space, hopefully moving efficiently around and thereby creating good estimates with low variance for the empirical averages. The basic transition types for our problem are as follows:

- Change a part of the template boundary of a randomly chosen object.
- Change the type of an object.
- Propose the birth of a new object.
- Propose the death of an object.

The details of the MCMC algorithm for these transitions are given in Rue & Husby (1998), Rue & Syversveen (1998), where the birth and death moves rely on applications of the Green (1995) algorithm for variable dimension distributions. In practice the basic transitions listed above do not ensure sufficient mixing of the Markov chain, the main problems being transitions between the various object types and difficulties when there is uncertainty over the whether one ”object” is actually two objects lying close together. Rue & Hurn (1999) approach this problem by extending the template model to allow it to have a variable number of sides, demonstrating that this allows smoother transitions between the object types, and easing the transitions which split and merge objects.

4 PARAMETER ESTIMATION FOR THE MODEL

The model described in the previous sections relies on the specification of the parameters controlling the marked point process prior and the degradation model. Some of these parameters, particularly those involved in the deformations, are hard to interpret in any intuitive way. In this section we propose a parameter specification approach which will hopefully allow models of this type to be more widely used in practice. The parameters of the degradation process are handled in a Bayesian way. The remaining parameters will be estimated using maximum likelihood methods.

4.1 BAYESIAN TREATMENT OF THE DEGRADATION PARAMETERS

The unknown parameters of the measurement process are μ the typical fluorescence level, a and b the operator controlled measurement adjustments, and σ^2 the recording noise. Since estimation of these parameters from a section of image data would be dependent on a good initial segmentation, we propose instead to treat them as hyperparameters. Our choice of prior distributions is intended mainly to impose the constraints that the variables are positive, with $0 \leq a\mu + b \leq 255$. The noise variance rather than the precision is given a $\chi^2(1)$ prior distribution since it is expected to be of the order of a few squared pixel units. The mean fluorescence is given a fairly diffuse $\chi^2(255)$ prior reflecting that it is positive variable expected to take a reasonably large value, although this will vary considerably between specimens. As the data carry considerable information regarding the parameters, it is intended that these priors will not exert much influence; refer to Hurn (1998) for more discussion in a similar setting,

$$\begin{array}{lll} \sigma^2 & \sim & \chi^2(1) \\ \mu & \propto & \chi^2(255) \\ a & \propto & U(0,255) \\ b & \propto & U(0,255) \end{array} \left.\vphantom{\begin{array}{l}\mu\\a\\b\end{array}}\right\} I[0 \le a\mu + b \le 255]$$

The joint posterior distribution of interest is then

$$\pi(\boldsymbol{x}, \sigma^2, \mu, a, b|\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x}, \sigma^2, \mu, a, b)\pi(\boldsymbol{x})\pi(\sigma^2)\pi(\mu, a, b).$$

Markov chain Monte Carlo methods are needed for the sampling which generally will be done one parameter at a time requiring the conditional posterior distributions

$$\begin{aligned} \pi(\sigma^2|\cdots) &\propto \frac{e^{-\sigma^2/2}}{\sqrt{\sigma^2}} \prod_{\text{pixels } p} \frac{\exp(-\frac{1}{2}(y_p - (a\mu I_{[p\in\text{cell}]} + b))^2)}{(a^2\mu I_{[p\in\text{cell}]} + \sigma^2)^{1/2}},\ \sigma^2 > 0 \\ \pi(\mu|\cdots) &\propto e^{-\mu/2}\mu^{253/2} \prod_{\text{pixels } p\in\text{cell}} \frac{\exp(-\frac{1}{2}(y_p - (a\mu + b))^2)}{(a^2\mu + \sigma^2)^{1/2}},\ 0 < \mu \le (255-b)/a \\ \pi(a|\cdots) &\propto \prod_{\text{pixels } p\in\text{cell}} \frac{\exp(-\frac{1}{2}(y_p - (a\mu + b))^2)}{(a^2\mu + \sigma^2)^{1/2}},\ 0 < a \le \min\{255, (255-b)/\mu\} \\ \pi(b|\cdots) &\propto \prod_{\text{pixels } p} \frac{\exp(-\frac{1}{2}(y_p - (a\mu I_{[p\in\text{cell}]} + b))^2)}{(a^2\mu I_{[p\in\text{cell}]} + \sigma^2)^{1/2}},\ 0 < b < 255 - a\mu, \end{aligned}$$

where "$\cdots$" denotes "all other variables".

4.2 Estimation of the prior model parameters

It is not computationally feasible to treat the remaining parameters in a similar Bayesian manner since it would require the intractable normalising constants which depend on these parameters. The alternative we suggest is to use likelihood–based methods which work with training data generated by the microscopist tracing cell outlines on a number of slices of image data using the computer mouse. This method has two benefits, the first is that the microscopist's expertise is being used in the data collection, and the second is that there is no further user–interaction in the estimation procedure.

4.2.1 The marked point process parameters p_i and γ

The use of a hard–core interaction model for the marked point process means that the normalising constant will be too complex for maximum likelihood estimation. However, previous experience in object recognition problems suggests that the results are not particularly sensitive to the specification of these parameters. We propose to use the naive estimators, that is the observed proportions of the different cell types in the training data for the $\{p_i\}$, and the observed frequency per unit area for γ.

These estimators will be least accurate in situations where there is significant packing observed in the training data. If this is the case, then an alternative is to simulate from the prior model (having first estimated the deformation parameters) searching over a range of values for $\{p_i\}$ and γ to match the observed and the simulated frequencies; this approach could be coded in advance. We also note some recent work by Baddeley & Turner (1998) who use maximum pseudo–likelihood methods to fit various point processes, although such an approach here would be considerably more complex.

4.2.2 THE DEFORMATION PARAMETERS β AND δ AND TRACKING ERRORS

We derive the likelihood function for β and δ using data which are a number of polygons the vertices of which are observed with error. Although the inclusion of the observation error complicates this derivation, it is a realistic assumption given the intended manner of collecting training data. We begin by transforming from the deformation model for the polygon edges to the model for the corresponding vertex locations. Recall that the first vertex defines the location of the entire polygon; if the first vertex is at location $\boldsymbol{c} = \boldsymbol{v}_0$, then the second vertex $\boldsymbol{v}_1$ is located at $\boldsymbol{v}_0 + \boldsymbol{s}_0^{R\alpha}\boldsymbol{g}_0$ and so on,

$$\boldsymbol{v}_j = \boldsymbol{v}_0 + \sum_{i=0}^{j} \boldsymbol{s}_i^{R\alpha}\boldsymbol{g}_i, \qquad j = 1, \dots, n.$$

Note that there are $n+1$ vertices in the non-closed polygon. Considering the x and y components separately, the vertices can be written

$$\begin{bmatrix} v_1^x \\ v_2^x \\ \vdots \\ v_n^x \\ v_1^y \\ v_2^y \\ \vdots \\ v_n^y \end{bmatrix} = \begin{bmatrix} g_0^x & & & & g_0^y & & & \\ g_0^x & g_1^x & & & g_0^y & g_1^y & & \\ \vdots & \vdots & \ddots & & \vdots & \vdots & \ddots & \\ g_0^x & g_1^x & \cdots & g_{n-1}^x & g_0^y & g_1^y & \cdots & g_{n-1}^y \\ g_0^y & & & & -g_0^x & & & \\ g_0^y & g_1^y & & & -g_0^x & -g_1^x & & \\ \vdots & \vdots & \ddots & & \vdots & \vdots & \ddots & \\ g_0^y & g_1^y & \cdots & g_{n-1}^y & -g_0^x & -g_1^x & \cdots & -g_{n-1}^x \end{bmatrix} \begin{bmatrix} t_0^{(0)} \\ t_1^{(0)} \\ \vdots \\ t_{n-1}^{(0)} \\ t_0^{(1)} \\ t_1^{(1)} \\ \vdots \\ t_{n-1}^{(1)} \end{bmatrix} + \mathcal{I}\boldsymbol{v}_0 + \boldsymbol{v}^{R\alpha} \qquad (3)$$

where $\mathcal{I}$ is the $2n \times 2$ matrix whose first column consists of n 1's followed by n 0's, and whose second column is n 0's followed by n 1's, and $\boldsymbol{v}^{R\alpha}$ is the vector of vertex x and y positions of the undeformed template with first vertex at zero rotated through α and scaled by R. We will write (3) in the form $\boldsymbol{v} = \boldsymbol{G}\boldsymbol{t} + \mathcal{I}\boldsymbol{v}_0 + \boldsymbol{v}^{R\alpha}$. We assume that we record the n vertices of the closed polygon with error $\tilde{\boldsymbol{v}}_0, \dots, \tilde{\boldsymbol{v}}_{n-1}$ where

$$\tilde{\boldsymbol{v}}_i = \boldsymbol{v}_i + \boldsymbol{\epsilon}_i, \qquad i = 0, \dots, n-1 \qquad (4)$$

and we further assume that these errors are normally distributed with

$$(\epsilon_0^x, \dots, \epsilon_{n-1}^x, \epsilon_0^y, \dots, \epsilon_{n-1}^y)^T \sim N_{2n}\left(\boldsymbol{0}, \boldsymbol{\Phi}\right).$$

Since the true location of the polygon is unobservable, we begin by finding the unconstrained distribution of $\boldsymbol{v}$ given $R, \alpha, \boldsymbol{\Phi}, \boldsymbol{\epsilon}_0$ and the observable $\tilde{\boldsymbol{v}}_0$, which by (3) and (4) is the distribution of

$$\boldsymbol{v} = \boldsymbol{G}\boldsymbol{t} + \mathcal{I}(\tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0) + \boldsymbol{v}^{R\alpha}.$$

Denoting the marginal variance of $(\epsilon_0^x, \epsilon_0^y)^T$ by $\boldsymbol{\Phi}_{22}$, the unconstrained distribution is

$$\boldsymbol{v}^T|(R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0) \sim N_{2n}\left(\mathcal{I}(\tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0) + \boldsymbol{v}^{R\alpha} \ , \ \boldsymbol{G}\begin{bmatrix} \boldsymbol{\Sigma}_R & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_R \end{bmatrix}\boldsymbol{G}^T\right). \tag{5}$$

To find the distribution of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T | \boldsymbol{v}_n = \boldsymbol{v}^*, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0$ for an arbitrary $\boldsymbol{v}^*$, we reorder the components of $\boldsymbol{v}$ from x then y components to the vertex pairs, rewriting (5) as

$$(\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_n)^T|(R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0) \sim N_{2n}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0 \end{bmatrix} , \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right), \tag{6}$$

where the partitioning of the mean and variance correspond to partitioning into the sets $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T$ and $\boldsymbol{v}_n$. Note that $\mathrm{E}(\boldsymbol{v}_n|(R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0)) = \tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0$ by closure of the undeformed template $\boldsymbol{v}^{R\alpha}$. Denote the partitioned inverse of the variance matrix

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^T & \boldsymbol{\Psi}_{22} \end{bmatrix}. \tag{7}$$

We will also rearrange the ordering of the $\{\boldsymbol{\epsilon}_i\}$ so that

$$(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{n-1}, \boldsymbol{\epsilon}_0)^T \sim N_{2n}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix} , \begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}^T & \boldsymbol{\Phi}_{22} \end{bmatrix}\right).$$

By manipulation of the conditional density $\pi(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}|\boldsymbol{v}_n = \boldsymbol{v}^*, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0)$, we show in the Appendix that

$$\begin{aligned}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T&|(\boldsymbol{v}_n = \boldsymbol{v}^*, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0) \\ &\sim N_{2n-2}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}(\tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0 - \boldsymbol{v}^*), \boldsymbol{\Psi}_{11}^{-1}\right).\end{aligned}$$

The particular $\boldsymbol{v}^*$ in which we are interested is $\boldsymbol{v}_0$, i.e. $\tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0$, in which case

$$(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T|(\boldsymbol{v}_n = \tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0) \sim N_{2n-2}\left(\boldsymbol{\mu}_1, \boldsymbol{\Psi}_{11}^{-1}\right). \tag{8}$$

To find the distribution of the vertices under the closure constraint, we use the known marginal distribution $\boldsymbol{\epsilon}_0 \sim N_2(\boldsymbol{0}, \boldsymbol{\Phi}_{22})$ together with (8), integrating out $\boldsymbol{\epsilon}_0$ to show in the Appendix that

$$(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T|(\text{closure}, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0) \sim N_{2n-2}\left(\boldsymbol{\mu}_2, \boldsymbol{\Psi}_{11}^{-1} + \bar{\mathcal{I}}\boldsymbol{\Phi}_{22}\bar{\mathcal{I}}^T\right), \tag{9}$$

where $\boldsymbol{\mu}_2$ is the first $2n-2$ components of the reordered $\mathcal{I}\tilde{\boldsymbol{v}}_0 + \boldsymbol{v}^{R\alpha}$ and $\bar{\mathcal{I}}$ is the $2n-2$ components of the reordered $\mathcal{I}$.

The distribution of $(\tilde{\boldsymbol{v}}_1, \dots, \tilde{\boldsymbol{v}}_{n-1})^T$ under the same conditioning uses (4). Since $(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{n-1})^T$ may not be independent of $\boldsymbol{\epsilon}_0$, the resulting distribution is

$$(\tilde{\boldsymbol{v}}_1, \dots, \tilde{\boldsymbol{v}}_{n-1})^T | (\text{closure}, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0) \\ \sim N_{2n-2}\left(\boldsymbol{\mu}_2, \boldsymbol{\Psi}_{11}^{-1} + \bar{\mathcal{I}}\boldsymbol{\Phi}_{22}\bar{\mathcal{I}}^T - \bar{\mathcal{I}}\boldsymbol{\Phi}_{12}^T - \boldsymbol{\Phi}_{12}\bar{\mathcal{I}}^T + \boldsymbol{\Phi}_{11}\right). \quad (10)$$

The distribution of $\tilde{\boldsymbol{v}}_0$ arises as the convolution of the distribution of $\boldsymbol{v_0}$ with the distribution of $\boldsymbol{\epsilon}_0$. Assuming that $\boldsymbol{v_0}$ is uniformly distributed in the observation window, and that the variance of $\boldsymbol{\epsilon}_0$ is small in comparison to the window size, $\tilde{\boldsymbol{v}}_0$ is taken to be also uniformly distributed in the window. Finally, under an assumption of independence of the polygon shapes, the likelihood for m cells will be the product of the individual likelihoods.

Finally, since in many applications it will not be reasonable to assume that the scale and rotation of each cell is known, the model allows each cell i to have an associated scaling R_i and rotation α_i. We will concentrate on the case where there is a common unknown R for all cells (as would be used in an example where size was a distinguishing characteristic in cell recognition). Treating scale and rotation as additional nuisance parameters adds parameters $R, \alpha_1, \dots, \alpha_m$, to the optimisation in order to obtain the maximum–likelihood estimates. However in the Appendix, we show that the rotation optimisation may be done exactly for any set of $\beta, \delta, R, \boldsymbol{\Phi}$, thus reducing the computational burden.

5 Examples of the estimation of the deformation parameters

5.1 An artificial example

We begin with an artificial example to demonstrate the properties of the estimation procedure since it allows us to consider the effect of various model misspecifications. The basic template for our experiments is an eight-sided regular approximation to a circle of radius 1, and the deformation model takes the parameters $\beta = 100$ and $\delta = -40$. Since it is analytically intractable to find the maximum likelihood estimates of β and δ, Matlab (The MathWorks Inc, 1994) has been used for the numerical optimisation; it can also be used for direct entry of vertex locations using the mouse.

5.1.1 The effect of model misspecification

In order to assess the importance of deriving the maximum likelihood estimates under the assumption of measurement error in recording the vertices, $m = 50$ simulated realisations from the deformed template model have been generated using independent measurement errors with variance ϕ_1^2, and estimates of β and δ have then been found from these data when misspecifying that $\phi_1^2 = 0$. Table 6 shows the mean and median estimates resulting from 250 repetitions of this procedure to assess the variability of the estimates. Not surprisingly, as the true variance increases, the estimates become severely biased;

$\hat{\beta}$ and $\hat{\delta}$ both decrease since the observed shapes demonstrate greater variability and less smoothness than expected.

Similarly, it is possible to fit a model with measurement error but for which the scale R and rotation α are assumed known but are incorrectly specified. The effect here is that the observed polygons do have the expected shapes but that in the attempt to fit the incorrectly specified template, ϕ_1^2 is overestimated, while the balance of $\hat{\beta}$ to $\hat{\delta}$ is perturbed to the point of degeneracy ($\beta = 2|\delta|$). This latter effect is due to the extremely high correlations required between deformations which are needed in order to mimic a rotation or a rescale.

5.1.2 Estimation under different error models

Having seen the importance of including measurement error in the modelling, we now investigate different possible models for the tracking errors. It seems unlikely that the errors will in practise be independent and so we consider the following three structures:

MODEL 1 Independent errors, $\boldsymbol{\Phi} = \phi_1 \boldsymbol{I}_{2n}$

MODEL 2 First-order Markov structure for the variance, $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_1 & 0 \\ 0 & \boldsymbol{\Phi}_1 \end{bmatrix}$ where $\boldsymbol{\Phi}_1$ has a circulant structure with ϕ_1 down the diagonal, and ϕ_2 on the second diagonal

MODEL 3 First-order Markov structure for the inverse variance, $\boldsymbol{\Phi}^{-1} = \begin{bmatrix} \boldsymbol{\Phi}_1 & 0 \\ 0 & \boldsymbol{\Phi}_1 \end{bmatrix}$ where $\boldsymbol{\Phi}_1$ is as above, although note that ϕ_1 and ϕ_2 have different interpretations under model 2 and 3.

Our training data are a set of m simulated cell outlines, where m takes the values 25, 50 or 100. In the simplest noise model, we have also found estimates when the scale and rotation are taken to be unknown (the nuisance parameter R takes the value 1). Table 2 summarises the results of the numerical experiments, while Figure 3 shows the pairwise scatterplots of the four estimated parameters for the case $m = 50$ under the independent noise model 1 with unknown scale and rotation. Both mean and median estimates have been quoted in all cases to highlight the skewed nature of the sampling distributions, a feature which can also be seen in Figure 3 (marginalising by eye). Figure 3 also demonstrates the high correlation structure within the estimates, particularly between β and δ and between the deformation parameters and the observation error variance. Such relationships might be explained by noting that in the unconstrained model whose inverse variance is given by (1) the variance of any particular $t_i^{(0)}$ given the remaining $\boldsymbol{t}_{-i}^{(0)}$ components is $1/\beta$ while the correlation of two consecutive $t_i^{(0)}$ given the remaining components is $-\delta/\beta$. The near non-identifiability due to the partitioning of the variability of the observed vertex locations into that due to the deformations and that due to the observational error is one of the reasons why this estimation problem is particularly difficult.

5.2 Estimation for imaging data

Figure 4 shows the two template types together with 100 cell outlines traced by one of the authors from a series of images of cell cartilage. For display purposes, the outlines have been relocated on a regular grid, the scalings and rotations have not been altered. Using these training data, the proposed estimation technique has been applied assuming unknown scale R and rotation and a common measurement error for both template types. The same three measurement noise models have been considered. Table 3 gives the resulting parameter estimates together with the quantity $-2\times$ the natural logarithm of the likelihood at those parameter estimates; this quantity will be used for determining which of the three models is most appropriate using the Akaike Information Criterion. Using this criterion, noise model 3 is chosen. Figure 5 shows some simulations from this fitted model; the top row shows 50 realisation from each of the two deformed template models, the bottom row shows the same realisations convolved with realisations from the fitted noise model. Note that the rotations here have been set to the value zero for displaying these outlines, which consequently appear less ordered than those in Figure 4 as a result of the natural structure of the cartilage.

6 Discussion

User–friendly techniques for parameter estimation in deformable template models are an important issue if such models are to find more wide-spread acceptance in applied work. Recent work by Kent, Dryden & Anderson (1998) has also considered the question of estimators for deformation parameters, although valid only for circular objects observed without measurement error. We have demonstrated that it is possible to treat all the parameters necessary for a Bayesian object recognition approach in either a fully Bayesian way or requiring only graphical intervention by the microscope user. We hope that approaches such as these will facilitate further work in the area of statistical image analysis in collaboration with the end–users of such techniques.

REFERENCES

BADDELEY, A. & TURNER, R. (1998). Practical maximum pseudolikelihood for spatial point patterns, *Technical report 16*, Department of Mathematics, University of Western Australia.

BADDELEY, A. J. & VAN LIESHOUT, M. N. M. (1993). Stochastic geometry models in high–level vision, *in* K. V. Mardia & G. K. Kanji (eds), *Statistics and Images*, Vol. 20, Abingdon: Carfax Publishing, chapter 11, pp. 235–256.

BLAKE, A. & ISARD, M. (1998). *Active Contours*, Berlin: Springer-Verlag.

GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.

GREEN, P. J. (1995). Reversible jump MCMC computation and Bayesian model determination, *Biometrika* 82(4): 711–732.

GRENANDER, U., CHOW, Y. & KEENAN, D. M. (1991). *Hands: a Pattern Theoretic Study of Biological Shapes*, Research Notes on Neural Computing, Springer, Berlin.

GRENANDER, U. & MILLER, M. I. (1994). Representations of knowledge in complex systems (with discussion), *Journal of the Royal Statistical Society, Series B* 56(4): 549–603.

HURN, M. A. (1998). Confocal fluorescence microscopy of leaf cells: An application of Bayesian image analysis, *Journal of the Royal Statistical Society, Series C* (47): 361–377.

JAIN, A. K., ZHONG, Y. & LAKSHMANAN, S. (1996). Object matching using deformable templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(3): 267–278.

KENT, J. T., DRYDEN, I. L. & ANDERSON, C. R. (1998). Using circulant symmetry to model featureless objects, *Techical report, Statistics 14*, Department of Statistics, University of Leeds, UK.

RUE, H. & HURN, M. A. (1999). Bayesian object identification, *Biometrika* XX(xx): xx–xx. (to appear).

RUE, H. & HUSBY, O. K. (1998). Identification of partly destroyed objects using deformable templates, *Statistics and Computing* 8(3): 221–228.

RUE, H. & SYVERSVEEN, A. R. (1998). Bayesian object recognition with Baddeley's delta loss, *Advances in Applied Probability (SGSA)* 30(1): 64–84.

THE MATHWORKS INC (1994). *MATLAB User's Guide, Version 4.2c*, Natick: The MathWorks Inc.

VAN LIESHOUT, M. N. M. (1995). Markov point processes and their applications in high-level imaging (with discussion), *Bulletin of the International Statistical Institute* LVI, BOOK 2: 559–576.

APPENDIX DERIVING THE LIKELIHOOD FUNCTION

A.1 IMPOSING THE CONSTRAINT $\boldsymbol{v}_n = \boldsymbol{v}^*$

Recalling the notation used in (6), denote $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1}$ by $\bar{\boldsymbol{v}}$, and $(R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0)$ by "$\cdots$" then

$$\pi(\bar{\boldsymbol{v}}, \boldsymbol{v}_n | \cdots) = \frac{(2\pi)^{-n}}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} \begin{bmatrix} \bar{\boldsymbol{v}} - \boldsymbol{\mu}_1 \\ \boldsymbol{v}_n - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0 \end{bmatrix}^T \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \bar{\boldsymbol{v}} - \boldsymbol{\mu}_1 \\ \boldsymbol{v}_n - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0 \end{bmatrix} \right\}$$

$$\pi(\boldsymbol{v}_n | \cdots) = \frac{(2\pi)^{-1}}{|\boldsymbol{\Sigma}_{22}|^{1/2}} \exp\left\{ -\frac{1}{2} [\boldsymbol{v}_n - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0]^T \boldsymbol{\Sigma}_{22}^{-1} [\boldsymbol{v}_n - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0] \right\}.$$

Using the same partitioning as in (7)

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{12}^T & \boldsymbol{\Psi}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{2n-2} & \boldsymbol{0}_{2n-2,2} \\ \boldsymbol{0}_{2n-2,2}^T & \boldsymbol{I}_2 \end{bmatrix}. \tag{11}$$

Then the conditional distribution of interest is

$$\frac{\pi(\bar{\boldsymbol{v}}, \boldsymbol{v}_n | \cdots)|_{\boldsymbol{v}_n = \boldsymbol{v}^*}}{\pi(\boldsymbol{v}_n | \cdots)|_{\boldsymbol{v}_n = \boldsymbol{v}^*}} = (2\pi)^{-(n-1)} \left(\frac{|\boldsymbol{\Sigma}|}{|\boldsymbol{\Sigma}_{22}|} \right)^{-1/2} \times$$
$$\frac{\exp(-1/2((\bar{\boldsymbol{v}}-\boldsymbol{\mu}_1)^T \boldsymbol{\Psi}_{11}(\bar{\boldsymbol{v}}-\boldsymbol{\mu}) + 2(\bar{\boldsymbol{v}}-\boldsymbol{\mu}_1)^T \boldsymbol{\Psi}_{12}(\boldsymbol{v}^* - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0) + (\boldsymbol{v}^* - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0)^T \boldsymbol{\Psi}_{22}(\boldsymbol{v}^* - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0)))}{\exp(-1/2((\boldsymbol{v}^* - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0)^T \boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{v}^* - \tilde{\boldsymbol{v}}_0 + \boldsymbol{\epsilon}_0)))}.$$

To see that this is the density of a $N_{2n-2}(\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{11}^{-1} \boldsymbol{\Psi}_{12}(\tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0 - \boldsymbol{v}^*), \boldsymbol{\Psi}_{11}^{-1})$, we need two identities: The first is that $|\boldsymbol{\Sigma}|/|\boldsymbol{\Sigma}_{22}| = |\boldsymbol{\Psi}_{11}^{-1}|$ and the second is that $\boldsymbol{\Psi}_{22} - \boldsymbol{\Sigma}_{22}^{-1} = \boldsymbol{\Psi}_{12}^T \boldsymbol{\Psi}_{11}^{-1} \boldsymbol{\Psi}_{12}$. We begin with the second, using two of the matrix equations of (11),

$$\begin{aligned} \boldsymbol{\Psi}_{12}^T \boldsymbol{\Psi}_{11}^{-1} \boldsymbol{\Psi}_{12} &= (-\boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{12}^T \boldsymbol{\Psi}_{11}) \boldsymbol{\Psi}_{11}^{-1} \boldsymbol{\Psi}_{12} \\ &= -\boldsymbol{\Sigma}_{22}^{-1}(I_2 - \boldsymbol{\Sigma}_{22} \boldsymbol{\Psi}_{22}) = \boldsymbol{\Psi}_{22} - \boldsymbol{\Sigma}_{22}^{-1} \end{aligned} \tag{12}$$

For the first, we apply Frobenius' theorem for the determinant of a partitioned matrix to $\boldsymbol{\Sigma}^{-1}$ and (12)

$$\begin{aligned} |\boldsymbol{\Sigma}^{-1}| &= |\boldsymbol{\Psi}_{11}| \, |\boldsymbol{\Psi}_{22} - \boldsymbol{\Psi}_{12}^T \boldsymbol{\Psi}_{11}^{-1} \boldsymbol{\Psi}_{12}| \\ &= |\boldsymbol{\Psi}_{11}| \, |\boldsymbol{\Sigma}_{22}^{-1}|. \end{aligned}$$

A.2 THE CLOSURE CONSTRAINT

Since $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T | (\boldsymbol{v}_n = \tilde{\boldsymbol{v}}_0 - \boldsymbol{\epsilon}_0, R, \alpha, \boldsymbol{\Phi}, \tilde{\boldsymbol{v}}_0, \boldsymbol{\epsilon}_0) \sim N_{2n-2}\left(\boldsymbol{\mu}_1, \boldsymbol{\Psi}_{11}^{-1}\right)$, to find the distribution of $\bar{\boldsymbol{v}}^T = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-1})^T$ under closure, we use the distribution of $\boldsymbol{\epsilon}_0$, integrating it out (i.e. considering all possible closures),

$$\pi(\bar{\boldsymbol{v}} | \cdots) \pi(\boldsymbol{\epsilon}_0) = \frac{|\boldsymbol{\Psi}_{11}^{-1}|^{-1/2} |\boldsymbol{\Phi}_{22}|^{-1/2}}{(2\pi)^n} \times$$
$$\exp\left\{ -\tfrac{1}{2}((\bar{\boldsymbol{v}} - \boldsymbol{\mu}_1)^T \boldsymbol{\Psi}_{11}(\bar{\boldsymbol{v}} - \boldsymbol{\mu}_1) + \boldsymbol{\epsilon}_0^T \boldsymbol{\Phi}_{22}^{-1} \boldsymbol{\epsilon}_0) \right\}.$$

Use the notation $\boldsymbol{\mu}_2$ to denote the first $2n-2$ rows of the reordered $\mathcal{I}\tilde{\boldsymbol{v}}_0 + \boldsymbol{v}^{R\alpha}$, then

$$\begin{aligned}
&(\bar{\boldsymbol{v}} - \boldsymbol{\mu}_1)^T \boldsymbol{\Psi}_{11} (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_1) + \boldsymbol{\epsilon}_0^T \boldsymbol{\Phi}_{22}^{-1} \boldsymbol{\epsilon}_0 \\
&= \boldsymbol{\epsilon}_0^T (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1}) \boldsymbol{\epsilon}_0 - 2\boldsymbol{\epsilon}_0^T \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2) + (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2)^T \boldsymbol{\Psi}_{11} (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2) \\
&= (\boldsymbol{\epsilon}_0 - (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1})^{-1} \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2))^T (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1}) \times \\
&\quad (\boldsymbol{\epsilon}_0 - (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1})^{-1} \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2)) \\
&\quad + (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2)^T (\boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1})^{-1} \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11}) (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2)
\end{aligned}$$

Since $\boldsymbol{\epsilon}_0$ can be seen to be Gaussian, integrating it out leaves the following terms

$$\begin{aligned}
&\frac{(2\pi)|\boldsymbol{\Psi}_{11}^{-1}|^{-1/2} |\boldsymbol{\Phi}_{22}|^{-1/2}}{(2\pi)^n |\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1}|^{1/2}} \times \\
&\quad \exp\left\{-\tfrac{1}{2}((\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2)^T (\boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1})^{-1} \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11}) (\bar{\boldsymbol{v}} - \boldsymbol{\mu}_2))\right\}
\end{aligned}$$

To see that this is the required density (9), first note by direct evaluation that

$$\boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} (\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1})^{-1} \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} = (\boldsymbol{\Psi}_{11}^{-1} + \bar{\mathcal{I}} \boldsymbol{\Phi}_{22} \bar{\mathcal{I}}^T)^{-1}.$$

It then remains to show that

$$|\boldsymbol{\Psi}_{11}^{-1} + \bar{\mathcal{I}} \boldsymbol{\Phi}_{22} \bar{\mathcal{I}}^T| \, |\bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}} + \boldsymbol{\Phi}_{22}^{-1}|^{-1} = |\boldsymbol{\Psi}_{11}^{-1}| \, |\boldsymbol{\Phi}_{22}|.$$

Consider the constructed partitioned matrix $A = \begin{bmatrix} \boldsymbol{\Psi}_{11}^{-1} & \bar{\mathcal{I}} \\ \bar{\mathcal{I}}^T & -\boldsymbol{\Phi}_{22}^{-1} \end{bmatrix}$ and apply Frobenius' formula for the determinant of a partitioned matrix twice, first treating the top left quadrant as the first factor, and then the bottom right quadrant as the first factor. This implies that

$$|\boldsymbol{\Psi}_{11}^{-1}| \, | - \boldsymbol{\Phi}_{22}^{-1} - \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}}| = |A| = | - \boldsymbol{\Phi}_{22}^{-1}| \, |\boldsymbol{\Psi}_{11}^{-1} + \bar{\mathcal{I}} \boldsymbol{\Phi}_{22} \bar{\mathcal{I}}^T|$$

or, as required

$$|\boldsymbol{\Psi}_{11}^{-1}| \, |\boldsymbol{\Phi}_{22}| = |\boldsymbol{\Psi}_{11}^{-1} + \bar{\mathcal{I}} \boldsymbol{\Phi}_{22} \bar{\mathcal{I}}^T| \, |\boldsymbol{\Phi}_{22}^{-1} + \bar{\mathcal{I}}^T \boldsymbol{\Psi}_{11} \bar{\mathcal{I}}|^{-1}.$$

A.3 The rotation parameter

Notice that in the log likelihood, the rotation of cell i occurs only through the $\boldsymbol{v}^{R\alpha}$ contribution to the mean of $(\tilde{\boldsymbol{v}}_1, \ldots, \tilde{\boldsymbol{v}}_{n-1})^T$ for that particular cell. Considering one particular cell denote the variance matrix in (10) by $\boldsymbol{V}$. Then the term to be minimised over α for a given β, δ, R and $\boldsymbol{\Phi}$ is $(\tilde{\boldsymbol{v}} - \boldsymbol{\mu}_2)^T \boldsymbol{V}^{-1} (\tilde{\boldsymbol{v}} - \boldsymbol{\mu}_2)$, where α occurs only through $\boldsymbol{\mu}_2$ (the first $2n-2$ components of the reordered $\mathcal{I}\tilde{\boldsymbol{v}}_0 + \boldsymbol{v}^{R\alpha}$). The scaled and rotated template $\boldsymbol{v}^{R\alpha}$ with the constrained vertex at zero can be written as a rotation applied to an already scaled template:

$$
\begin{aligned}
(\boldsymbol{v}^{R\alpha})_i &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} (\boldsymbol{v}^R)_i \\
&= \begin{bmatrix} (\boldsymbol{v}^R)_i^x & -(\boldsymbol{v}^R)_i^y \\ (\boldsymbol{v}^R)_i^y & (\boldsymbol{v}^R)_i^x \end{bmatrix} \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix} \\
(\boldsymbol{v}^{R\alpha})_{i=1,\dots,n-1} &= \begin{bmatrix} (\boldsymbol{v}^R)_1^x & -(\boldsymbol{v}^R)_1^y \\ (\boldsymbol{v}^R)_1^y & (\boldsymbol{v}^R)_1^x \\ \vdots & \vdots \\ (\boldsymbol{v}^R)_{n-1}^x & -(\boldsymbol{v}^R)_{n-1}^y \\ (\boldsymbol{v}^R)_{n-1}^y & (\boldsymbol{v}^R)_{n-1}^x \end{bmatrix} \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}.
\end{aligned}
$$

Writing this in the form $(\boldsymbol{v}^{R\alpha})_{j=1,\dots,n-1} = \boldsymbol{v}^R \boldsymbol{P}_\alpha$, the square to be minimised may be re-expressed as $(\tilde{\tilde{\boldsymbol{v}}} - \boldsymbol{v}^R\boldsymbol{P}_\alpha)^T \boldsymbol{V}^{-1} (\tilde{\tilde{\boldsymbol{v}}} - \boldsymbol{v}^R\boldsymbol{P}_\alpha)$, where $\tilde{\tilde{\boldsymbol{v}}}$ is the first $2n-2$ components of $\tilde{\boldsymbol{v}} - \mathcal{I}\tilde{\boldsymbol{v}}_0$. Introducing the further notation $\boldsymbol{Q} = [Q_1\ Q_2]$ for $-2\tilde{\tilde{\boldsymbol{v}}}^T \boldsymbol{V}^{-1}\boldsymbol{v}^R$ and $\boldsymbol{T} = \begin{bmatrix} T_1 & T_2 \\ T_2 & T_3 \end{bmatrix}$ for $(\boldsymbol{v}^R)^T \boldsymbol{V}^{-1}\boldsymbol{v}^R$, then α is chosen to minimise

$$
\begin{aligned}
S &= \boldsymbol{Q}\boldsymbol{P}_\alpha + \boldsymbol{P}_\alpha^T \boldsymbol{T} \boldsymbol{P}_\alpha \\
&= Q_1\cos\alpha + Q_2\sin\alpha + T_1\cos^2\alpha + T_3\sin^2\alpha + 2T_2\cos\alpha\sin\alpha.
\end{aligned}
$$

Differentiating with respect to α, and using the $\tan$–transform $t = \tan(\alpha/2)$, leads to a quartic in t

$$
(2T_2 - Q_2)t^4 + (4(T_1 - T_3) - 2Q_1)t^3 - 12T_2t^2 + (4(T_3 - T_1) - 2Q_1)t + (Q_2 + 2T_2) = 0
$$

which may be readily solved in Matlab to find the minimising α.

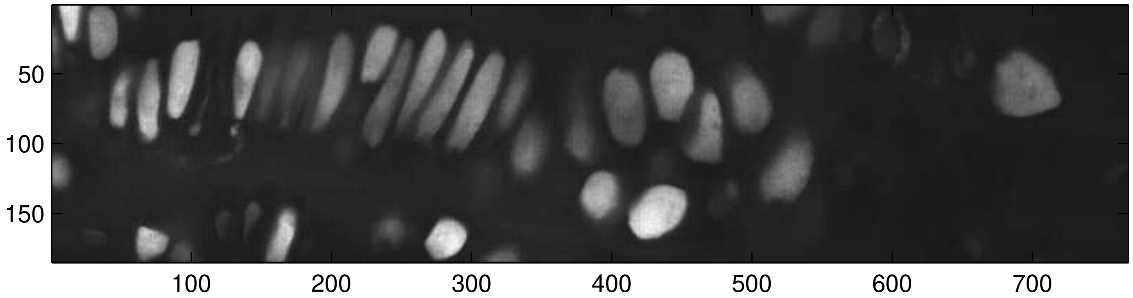

FIGURE 1: An optical section through cartilage growth plate.

True noise value	Mean estimates	Median estimates
$\phi_1^2 = (0.0)^2$	$\beta = 99.974$ $\delta = -39.856$	$\beta = 99.411$ $\delta = -39.904$
$\phi_1^2 = (0.01)^2$	$\beta = 94.623$ $\delta = -37.003$	$\beta = 94.140$ $\delta = -36.923$
$\phi_1^2 = (0.05)^2$	$\beta = 42.253$ $\delta = -9.668$	$\beta = 42.347$ $\delta = -9.583$
$\phi^2 = (0.1)^2$	$\beta = 20.446$ $\delta = -2.008$	$\beta = 18.459$ $\delta = 0.287$

TABLE 1: Parameter estimates constraining $\phi_1^2 = 0$ for $m = 50$, 250 repetitions with known R and α, true $\beta = 100$ and $\delta = -40$ under different true values of ϕ_1^2.

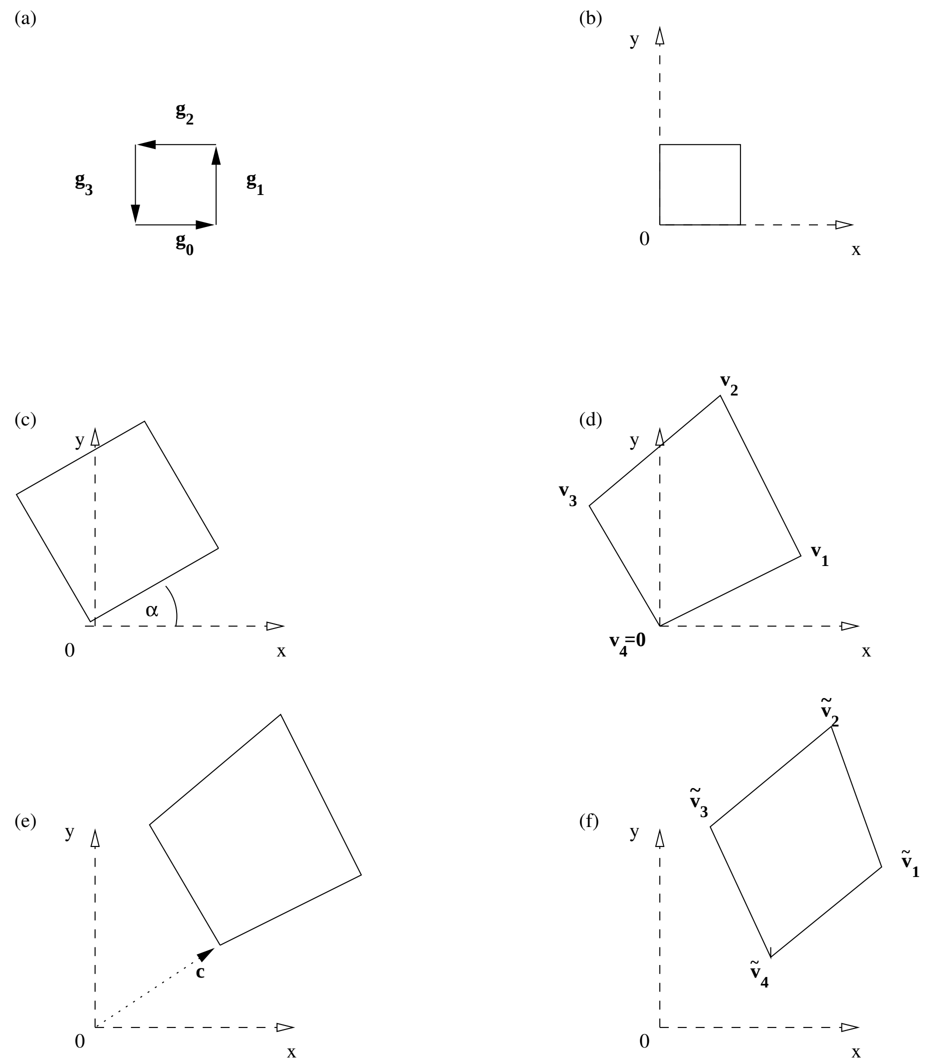

FIGURE 2: The stages of data formation.

	Noise model 1 $\phi_1 = 0.0025$ Known R and α	Noise model 1 $\phi_1 = 0.0025$ Unknown R and α	Noise model 2 $\phi_1 = 0.0025$ $\phi_2 = 0.0005$	Noise model 3 $\phi_1 = 500$ $\phi_2 = -100$
$m = 25$				
Mean β	104.028	103.159	119.363	157.173
Median β	98.915	98.940	98.242	94.870
Mean δ	-41.318	-30.520	-49.168	-68.168
Median δ	-39.042	-30.042	-39.638	-37.088
Mean ϕ_1	0.00248	0.00242	0.00249	1279.08
Median ϕ_1	0.00245	0.00246	0.00239	542.011
Mean ϕ_2	–	–	0.000510	265.256
Median ϕ_2	–	–	0.000473	-68.513
Mean R	–	1.0036	–	–
Median R	–	1.0043	–	–
$m = 50$				
Mean β	103.755	98.510	104.926	116.685
Median β	101.173	97.065	96.726	101.414
Mean δ	-41.180	-28.797	-42.282	-47.685
Median δ	-40.806	-27.740	-38.360	-40.567
Mean ϕ_1	0.00252	0.00240	0.00246	695.541
Median ϕ_1	0.00250	0.00239	0.00245	523.072
Mean ϕ_2	–	–	0.000488	-5.1085
Median ϕ_2	–	–	0.000504	-88.9432
Mean R	–	1.0029	–	–
Median R	–	1.0037	–	–
$m = 100$				
Mean β	101.602	96.371	103.048	111.526
Median β	99.121	96.344	101.340	102.223
Mean δ	-40.758	-28.281	-41.630	-45.361
Median δ	-39.783	-28.038	-40.511	-40.444
Mean ϕ_1	0.00248	0.00239	0.00247	561.077
Median ϕ_1	0.00246	0.00240	0.00247	501.423
Mean ϕ_2	–	–	0.000476	-61.844
Median ϕ_2	–	–	0.000462	-90.032
Mean R	–	1.0025	–	–
Median R	–	1.0030	–	–

TABLE 2: The simulation results for various scenarios.

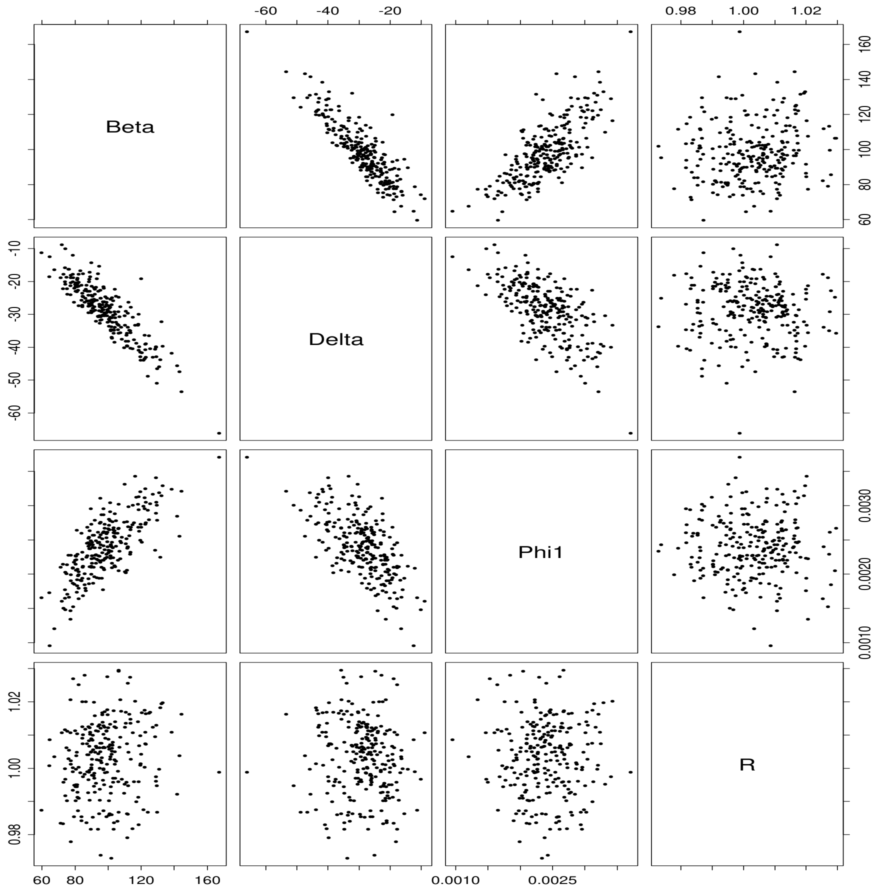

FIGURE 3: The $m = 50$ case under noise model (1) with unknown scale and rotation.

		Noise model 1	Noise model 2	Noise model 3
Circular template	$\hat{\beta}$	12.641	12.957	16.891
	$\hat{\delta}$	-3.914	-2.633	0.167
	$\hat{R}$	35.503	35.161	35.586
Oblong template	$\hat{\beta}$	12.660	17.349	16.003
	$\hat{\delta}$	-2.329	3.917	0.082
	$\hat{R}$	59.995	59.155	59.048
Noise parameters	$\hat{\phi}_1$	2.612	6.760	0.215
	$\hat{\phi}_2$	–	3.202	-0.108
$-2\ln(\text{max likelihood})$		5188.2	5178.7	5206.6

TABLE 3: The parameter estimates under different noise models.

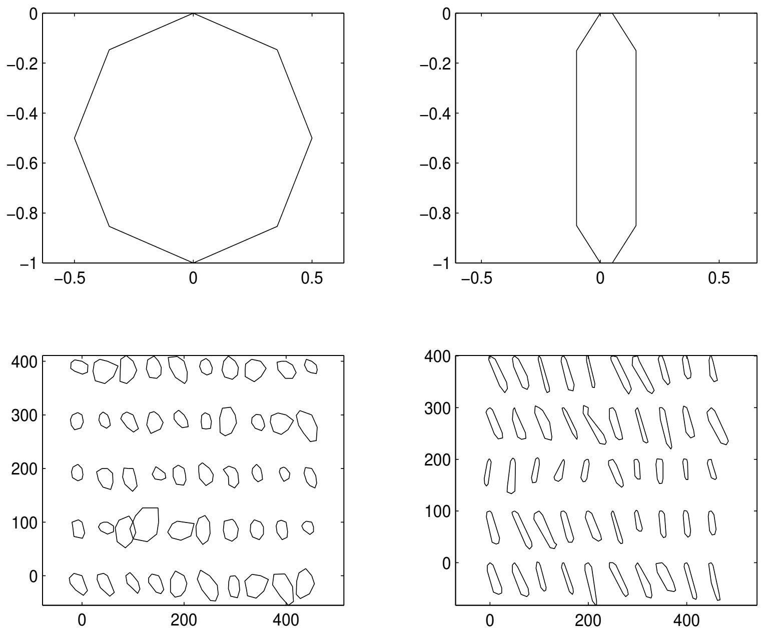

FIGURE 4: The two template types and the training data

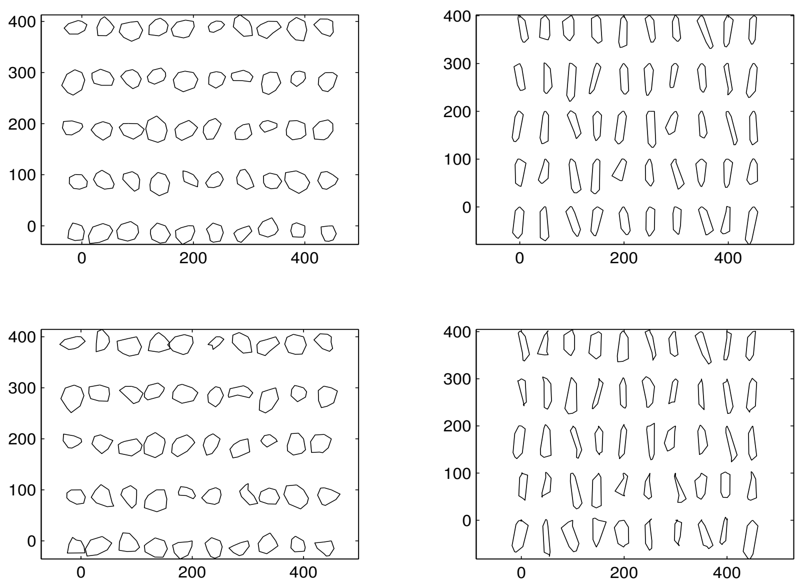

FIGURE 5: Simulations using the estimated parameter values (top row without measurement error, bottom row the same realisations convolved with measurement error).