2001, Journal of the Royal Statistical Society, Series B, 63, 465-492 (discussion omitted)

A penalised likelihood approach to image warping

C.A. Glasbey
Biomathematics and Statistics Scotland
JCMB, King's Buildings, Edinburgh EH9 3JZ

and K.V. Mardia
Department of Statistics
University of Leeds, Leeds LS2 9JT

November 16, 2000

Abstract

A warping is a function that deforms images by mapping between image domains. The choice of function is formulated statistically as maximum penalised likelihood, where the likelihood measures the similarity between images after warping and the penalty is a measure of distortion of a warping. The paper addresses two issues simultaneously, of how to choose the warping function and how to assess the alignment. A new, Fourier-von Mises image model is identified, with phase differences between Fourier-transformed images having von Mises distributions. Also, new, null-set distortion criteria are proposed, with each criterion uniquely minimised by a particular set of polynomial functions. A conjugate gradient algorithm is used to estimate the warping function, which is numerically approximated by a piecewise bilinear function. The method is motivated by, and used to solve, three applied problems: to register a remotely-sensed image with a map, to align microscope images obtained using different optics, and to discriminate between species of fish from photographic images.

Key words: bijective transformation, conjugate gradients, cross-covariance, digital microscopy, distortion criteria, Fast Fourier transform, fish species discrimination, phase correlation, polynomial transformation, registration, SAR, similarity transformation, thin-plate splines, von Mises distribution.

1 Introduction

Image analysis, the extraction of information from pictures, is a broad, interdisciplinary field with many challenging problems to which statistical methods are applicable (for overviews,

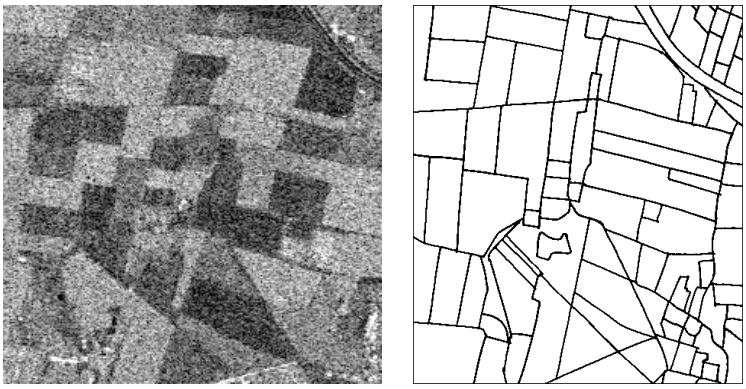
see Mardia, 1994; Glasbey and Horgan, 1995). One such topic is *image warping*, a function that deforms images by mapping between image domains. Warping is a fundamental stage in many applications of image analysis, whether to register an image with a map or template, or to align multiple images. It dates back over a century, to Galton (1878), who used analogue methods to construct average faces of criminals and mental patients from photographs. Since then, the subject has had a large and diverse literature. The images to be aligned may be different specimens to be compared in order to characterise population variation, or the same specimen at different times to be interpolated between ('morphed'), or complementary sources of information to be fused. Alternatively, they may be either successive two-dimensional sections or stereoscopic pairs, from which a three-dimensional scene is to be reconstructed. In some applications, different types of deformation or even discontinuities may be permissible in parts of images. The transformation may be constrained to be one-to-one, i.e. bijective, or folding may be acceptable. Also, it may or may not be appropriate for the boundaries of one image domain to map to the boundaries of the other domain. The accuracy required of the alignment is another issue: if a radiologist is to make a visual assessment of two medical images then precise alignment may be unnecessary, whereas in remote sensing where quantitative use is to be made of images, subpixel registration may be critical.

To illustrate, we consider three applied problems, the data for which are obtainable from `ftp://ftp.bioss.sari.ac.uk/pub/chris/warping/`

1. Fig 1(a) shows a remotely-sensed synthetic aperture radar (SAR) image of an area near Feltwell, England. SAR is an active remote-sensing system: microwave radiation is beamed down to the earth's surface from a plane or satellite, a sensor detects the reflected signal, and from this an image is constructed. Before any practical use can be made of such an image, it needs to be registered with a map, such as the digitised map of field boundaries in Fig 1(b). Registration of remotely-sensed images, including SAR, is often performed manually (see, for example, Vornberger and Bindschadler, 1992; Dobson et al., 1996). Li et al. (1995) reviewed automatic methods, distinguishing between area- and feature-based ones. To locate features, Caves et al. (1992) used linear filters, whereas Kher and Mitra (1993) used morphological methods. Registration of SAR can also simplify the task of segmenting the images into homogeneous regions (Glasbey, 1997).

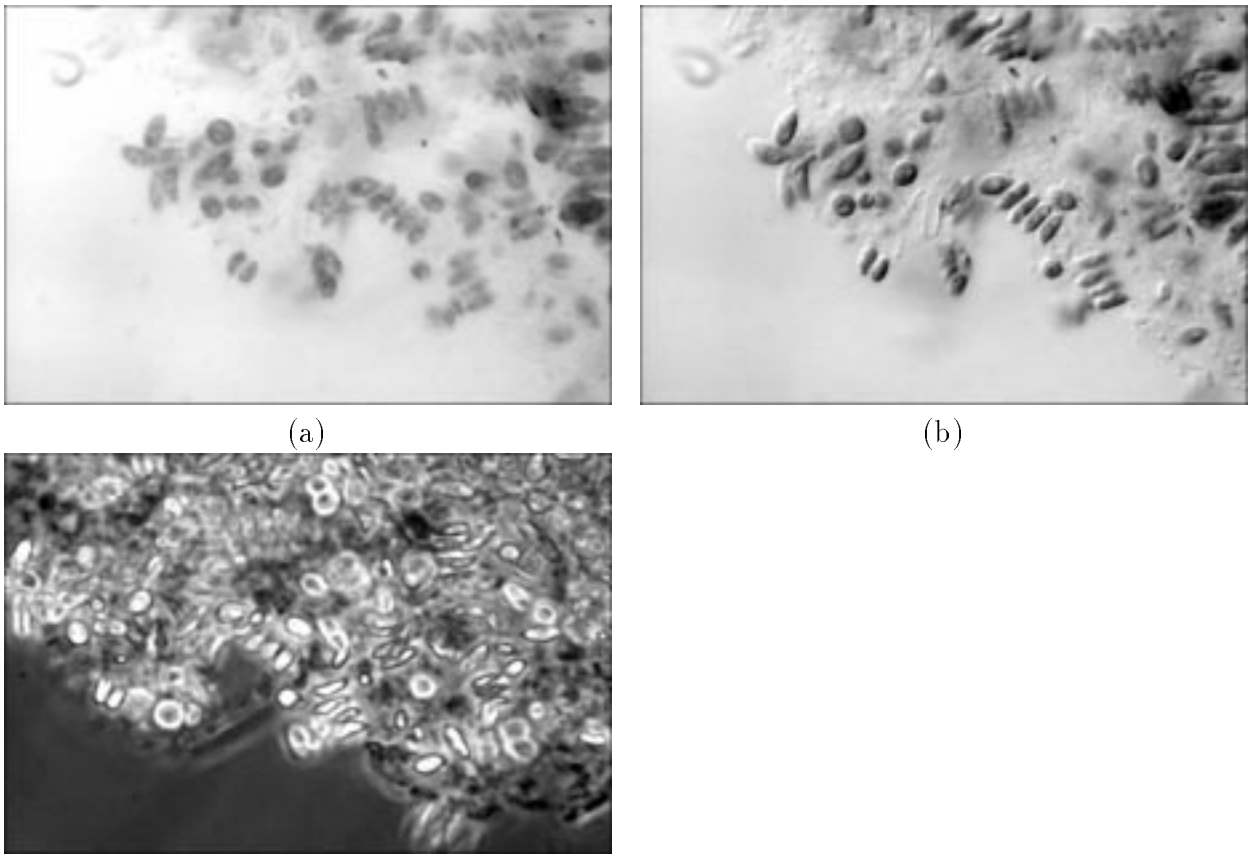
2. Fig 2 shows a sample of algae imaged using three light-microscope modalities: brightfield, differential interference contrast (DIC) and phase contrast. Brightfield microscopy reveals the optical attenuation of the specimen, whereas DIC microscopy responds to the refractive properties of the specimen and phase contrast microscopy shows diffractive properties. By fusing the images, these sources of complementary information can be combined (Modrusan et al., 1994; Ried et al., 1992). However, this requires a translation to be applied to the images to compensate for changes in image alignment resulting from imperfect centration of the different lens systems. Galbraith and Farkas (1993) described two methods for aligning images, involving either the imaging of a rectangular grid or the manual identification of control points.

3. Fig 3 shows photographic images, obtained under controlled conditions, of two species of fish (haddock and whiting) that we wish to discriminate. These are part of a larger data set consisting of images of ten haddock and ten whiting. Strachan et al. (1990) analysed



(a) (b)

Figure 1: *Area to the north of the village of Feltwell in East Anglia:* **(a)** *an aerial SAR image,* 250×250 *pixels in size (3km* $\times$ *3km),* **(b)** *digital line drawing of field, road and other boundaries for approximately the same region.*



(a) (b)

(c)

Figure 2: *A sample of algae imaged using three light-microscope modalities:* **(a)** *brightfield,* **(b)** *differential interference contrast,* **(c)** *phase contrast. (Images are 512* $\times$ *768 pixels in size.)*

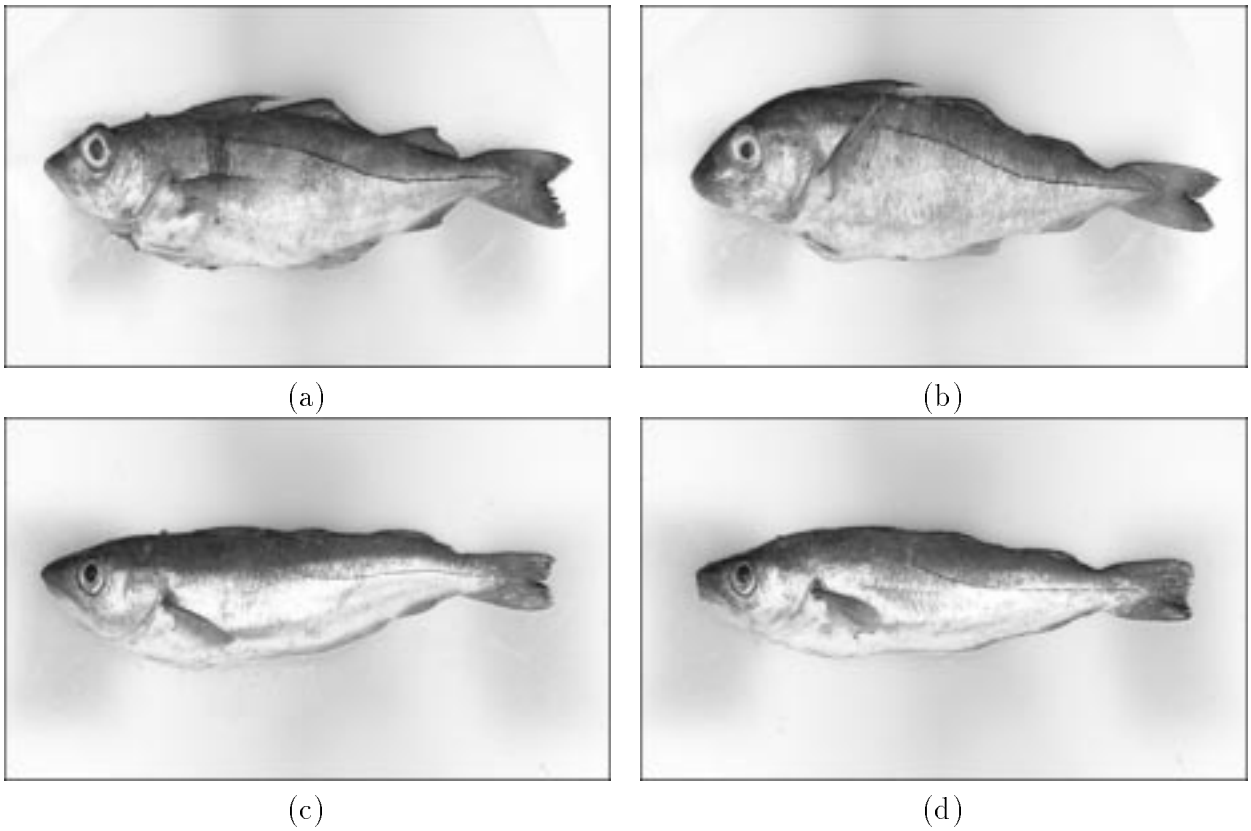


Figure 3: *Images of two species of fish, photographed on a light-table:* **(a)** *haddock 1,* **(b)** *haddock 2,* **(c)** *whiting 1,* **(d)** *whiting 2. (Images are 300 × 500 pixels in size.)*

images of seven species and found these two species to be the hardest to distinguish. One way of comparing images is by warping them to align with each other. These images have already been aligned globally, and our concern is with local alignment. The study of fish shape is a subject with a long history. Comparisons have typically been restricted to the fish outlines and a few other features, but simple measures such as length to width ratios are not sufficient in this application. Thompson (1917) used a mapping which superimposed an outline of one fish on another, as a way of comparing shapes. Bookstein (1991) developed this further, whereas Strachan et al. (1990) used summary statistics derived from outlines to discriminate among seven species of fish, and Mokhtarian (1995) used curvature scale space to recognise marine animals.

Our proposal is a statistical formulation of image warping, using a penalised likelihood approach. However, we first summarise a large number of alternative approaches, predominantly in the computer vision and engineering literatures. There are recent reviews of image warping in general (Glasbey and Mardia, 1998; Goshtasby and Le Moigne, 1999), of medical applications and computational anatomy (Grenander and Miller, 1998; Maintz and Viergever, 1998; Singh et al., 1998), and brain imaging specifically (Toga, 1999; Cao and Worsley, 1999), of comparison of faces (Hallinan et al., 1999), and of templates and shape analysis (McInerney and Terzopoulos, 1996; Dryden and Mardia, 1998; Loncaric, 1998). For the special case of one-dimensional

curve registration, see Ramsay and Li (1998). Measures of similarity to assess the quality of image alignment have included mean square differences and correlation between pixels in images (see, for example, Rosenfeld and Kak, 1982, §9.4), phase correlation (Kuglin and Hines, 1975), coincidence of landmark points (Cross and Hancock, 1998; Hill et al., 2000) or edges in images (Bajcsy and Kovacic, 1989; Moshfeghi, 1991), mutual information (Meyer et al., 1996; Viola and Wells III, 1997; Rangarajan et al., 1999; Studholme et al., 1999) and distance metrics (Baddeley and Molchanov, 1998; Kaijser, 1998). Measures used to ensure that the warping is not too severe have been motivated by thin-plate splines (Bookstein, 1991), elastic deformations (Burr, 1981; Younes, 1999), optical or fluid flow (Barron et al., 1994; Christensen et al., 1996; Joshi and Miller, 2000), diffusion (Amit et al., 1991), numerical regularisers (Thompson et al., 1991), Hopfield neural networks (Cote and Tatnall, 1997), and Bayesian prior distributions (Carstensen, 1996; Gee, 1999).

Our proposal builds on much of this earlier work, but is distinctive. It is also our intention to give image warping more exposure to a statistical audience, which we think it needs. In §2 we formulate image warping as a penalised likelihood problem, incorporating new classes of both similarity measures and distortion penalties. We restrict attention to two-dimensional images, though the theory extends in a straightforward manner to three and higher dimensions. Then, in §3 we apply the method to solve the three problems above. Finally, in §4 we discuss the results.

2 Method

Suppose that we have a single image, Y, that we wish to align with another, given image, μ, sometimes referred to in the computer vision literature as a greyscale template. We propose to do so by estimating the warping function, $f : \Re^2 \mapsto \Re^2$, to maximise a *penalised likelihood* functional, P, consisting of two components:

$$P(Y|\mu, f, \xi, \mathcal{C}, \lambda) \quad = \quad L(Y|\mu, f, \xi) \quad - \quad \lambda D(f, \mathcal{C}). \tag{1}$$

Here, L is the log-likelihood for Y, which depends on the warping function f and parameters ξ. The log-likelihood operates as a measure of similarity between the warped version of Y and μ. The second component, D, is a non-negative *measure of distortion* of f, chosen to be zero if and only if $f \in \mathcal{C}$, a *null set* of functions, and λ is a non-negative constant that determines the relative weighting between L and D. We use the term 'distortion' in preference to the commonly-used term 'roughness', because we sometimes wish to penalise warpings that would not be considered rough in the general sense of that word. As f is infinite dimensional, in the absence of a measure of distortion, the problem would be ill-conditioned. Penalised likelihoods have appeared in the statistical literature in many other contexts, and may be justified in several ways, including as regularisers and in Bayesian formulations (see, for example, Green and Silverman, 1994; Green, 1999). We could place a probabilistic interpretation on f, via $D(f, \mathcal{C})$, (see, for example, Grenander and Miller, 1998) but prefer to leave it ambiguous.

If we have two images, then it may be natural to use one of them as the greyscale template, μ. For example, in Problem 1 we take the SAR image to be Y, which we align with the digital

map, taken to be μ. However, if we have two images that we wish to treat interchangeably, or K (> 2) images, $Y^{(1)}, \ldots, Y^{(K)}$, then μ takes on the role of a consensus image that we also need to estimate. We generalise (1) to

$$P^{(K)}(Y^{(1)}, \ldots, Y^{(K)} \mid \mu, f^{(1)}, \ldots, f^{(K)}, \xi^{(1)}, \ldots, \xi^{(K)}, \mathcal{C}, \lambda) = \sum_k P(Y^{(k)} | \mu, f^{(k)}, \xi^{(k)}, \mathcal{C}, \lambda) \quad (2)$$

where, $f^{(k)}$ denotes the warping function from $Y^{(k)}$ to μ, and we maximise $P^{(K)}$ also with respect to μ, which is an array of location parameters. Note, however, that it is not always possible to simultaneously estimate μ and ξ, a topic to which we will return in §3.2.

We consider specific forms for L in §2.1 and for D in §2.2, then describe an algorithm to estimate f in §2.3. There are many ways to choose λ (see, for example, Thompson et al., 1991). We illustrate some specific strategies in the applications, namely: cross-validation (§3.1), prior knowledge (§3.2) and discriminatory power (§3.3).

2.1 Fourier-von Mises image model

We consider two log-likelihoods, the first one is based on a Gaussian model for Y after warping, but our main, novel proposal is for a *Fourier-von Mises image model*. First we need further notation.

Image μ is a real function, either on a discrete domain, $X = \{1, \cdots, n_1\} \times \{1, \cdots, n_2\}$, so that $\mu : X \mapsto \Re$, or on a continuous domain, so that $\mu : (0, n_1) \times (0, n_2) \mapsto \Re$, according as which is more convenient (although values of μ are typically only known or estimated on the discrete domain). In this section the domain is taken to be discrete. We use μ_x to denote the pixel value at location $x = (x_1, x_2)$. Image Y is similarly specified, and has a possibly different size, $n' = (n'_1, n'_2)$. We define Y_f to be the warped version of Y under f. It is an array of size $n = (n_1, n_2)$ specified by

$$(Y_f)_x \equiv Y_{f(x)} \qquad \forall x \in X,$$

and so its pixel value at x is defined to be the value of Y at location $f(x)$. Typically, $(f_1(x), f_2(x))$ are not integers, so $Y_{f(x)}$ is obtained by interpolation, and, if $f(x)$ lies outside the domain of Y, $Y_{f(x)}$ is defined to be a constant: either zero or a mean pixel value. We use bilinear interpolation, though it would be possible to use alternatives such as splines or kernels.

Consider a simple, Gaussian model for Y, conditional on f, of the form:

$$Y_{f(x)} \sim \mathrm{N}(\mu_x, \sigma^2) \qquad \forall x \in X, \quad (3)$$

with Y_x for other values of x specified deterministically, by bilinear interpolation, for example. We regard Y as a single entity rather than as an array of individual observations, as in the philosophy in functional data analysis of Ramsay and Silverman (1997, pp. 37-38). The log-likelihood of Y, to within additive and scaling constants, is

$$L^*(Y|\mu, f) \quad = \quad -\sum_x \left(Y_{f(x)} - \mu_x\right)^2, \quad (4)$$

where, throughout the paper, the x-summation is over X. This Gaussian model may be reasonable for Problem 3, the fish images in §1, but not for all applications. For example, in Problem 2, algal cells appear dark in Fig 2(a), whereas in Fig 2(b), one side of each cell appears dark whereas the other side appears light, and in Fig 2(c) cells look different again. Therefore, it would be more appropriate to model the relationship between the edges of cells in the two images, rather than the image intensities directly. By constructing an image model in the Fourier domain, we can be flexible in allowing either intensities or edges to be related, provided that the edges can be extracted using linear filters, as we show below.

The Fourier representation of Y_f is

$$Y_{f(x)} = \frac{1}{\sqrt{n_1 n_2}} \sum_{\omega} A_{\omega}^{(Y_f)} \cos\left(\theta_{\omega}^{(Y_f)} + 2\pi\ \omega^T x\right) \qquad \forall x \in X, \tag{5}$$

where $A^{(Y_f)}$ and $\theta^{(Y_f)}$ are, respectively, the arrays of amplitudes and phases of the Fourier transform of Y_f. Throughout the paper, the ω-summation is over Ω, the set of frequencies $\omega = (j_1/n_1, j_2/n_2)$ for $j_i = -\frac{1}{2}n_i, (-\frac{1}{2}n_i + 1), \ldots, -1, 0, 1, \ldots, (\frac{1}{2}n_i - 2), (\frac{1}{2}n_i - 1)$ if n_i is even, or $j_i = -\frac{1}{2}(n_i - 1), \ldots, \frac{1}{2}(n_i - 1)$ if n_i is odd. Similarly, we define $A^{(\mu)}$ and $\theta^{(\mu)}$ to be the arrays of amplitudes and phases of the Fourier transform of μ. Note, the arrays have a rotational symmetry, as $A_{\omega}^{(Y_f)} = A_{-\omega}^{(Y_f)}$ and $\theta_{\omega}^{(Y_f)} = -\theta_{-\omega}^{(Y_f)}$. Also, any linear filter applied to μ can be interpreted and computed simply as a rescaling of each element in $A^{(\mu)}$ and the addition of a constant to $\theta^{(\mu)}$. The arrays can be computed efficiently using Fast Fourier Transforms, and we taper the image boundaries using a cosine bell, to remove artificial image discontinuities produced by wrap-round of the image domain. For introductory background to Fourier analysis of images, see, for example, Glasbey and Horgan (1995, Ch.3, pp60-70).

We now specify our Fourier-von Mises image model for Y, conditional on f. The Fourier phases, $\theta^{(Y_f)}$, are independently von Mises distributed, conditional on $A^{(Y_f)}$, as follows:

$$(\theta_{\omega}^{(Y_f)} | A^{(Y_f)}) \quad \sim \quad \mathrm{M}(\theta_{\omega}^{(\mu)}, \kappa_{\omega}(\xi)) \qquad \forall \omega \in \Omega, \tag{6}$$

were the concentration, $\kappa(> 0)$, is given by

$$\kappa_{\omega}(\xi) = \exp\left[\xi_0 + \xi_1|\omega| + \xi_2|\omega|^2 + \xi_3 \log A_{\omega}^{(\mu)} + \xi_4 \log A_{\omega}^{(Y_f)}\right]. \tag{7}$$

Here, $|\omega|$ denotes the modulus of ω, a non-direction frequency, and κ is a log-linear function of $|\omega|$, $|\omega|^2$, $A^{(\mu)}$ and $A^{(Y_f)}$, with parameters ξ. The Fourier amplitudes, $A^{(Y_f)}$, are regarded as an array of constants rather than random variables, and we use (5) to define Y_x for all values of x, not just for $f(x) \in X$. Therefore, the log-likelihood for Y, to within an additive constant including terms in $A^{(Y_f)}$ for the Jacobian of the transformation, is

$$L(Y|\mu, f, \xi) = \sum_{\omega} \kappa_{\omega}(\xi) \cos\left(\theta_{\omega}^{(Y_f)} - \theta_{\omega}^{(\mu)}\right) \ - \ \sum_{\omega} \log I_0(\kappa_{\omega}(\xi)), \tag{8}$$

where, I_0 is the normalising term for the von Mises distribution, a modified Bessel function of the first kind and of order zero (see, for example, Mardia and Jupp, 1999). We have the following four theoretical and empirical motivations for choosing this model.

Firstly, for particular choices of ξ, the log-likelihood simplifies to commonly-used measures of similarity between images. If $\xi = (\xi_0, 0, 0, 1, 1)$, then $\kappa = A^{(\mu)}A^{(Y_f)}$ and L can be re-expressed as

$$L(Y|\mu, f, (\xi_0, 0, 0, 1, 1)) = \sum_x \mu_x \; Y_{f(x)} \; - \; \sum_\omega \log I_0 \left(A_\omega^{(\mu)} A_\omega^{(Y_f)} \right). \tag{9}$$

The first term is the cross-covariance or cross-product between μ and Y, which is closely related to L^*, given by (4). If $\xi = (\xi_0, 0, 0, 0, 0)$, then κ is a constant and we obtain the phase correlation measure (Kuglin and Hines, 1975), the cross-covariance between the images after the application of a high-pass filter which results in the filtered images having flat spectra (Glasbey and Horgan, 1995, Fig 3.6b, p.65). In general, L can be re-expressed, and interpreted, as the cross-covariance between Y_f and a filtered version of μ, which we denote by $\mu^{<\xi>}$,

$$L(Y|\mu, f, \xi) = \sum_x \mu_x^{<\xi>} \; Y_{f(x)} \; - \; \sum_\omega \log I_0(\kappa_\omega(\xi)), \tag{10}$$

where

$$\mu_x^{<\xi>} = \frac{1}{\sqrt{n_1 n_2}} \sum_\omega \frac{\kappa_\omega(\xi)}{A_\omega^{(Y_f)}} \cos\left(\theta_\omega^{(\mu)} + 2\pi \; \omega^T x\right) \qquad \forall x \in X. \tag{11}$$

Array $\mu^{<\xi>}$ is a filtered version of μ, obtained by modifying the amplitudes in the Fourier transform and then back-transforming. It is also a function of f but we suppress this dependence for reasons discussed in the optimisation algorithm in §2.3. Alternatively, we could have applied the filter to Y_f, or shared its effects between both μ and Y_f, but we will make use of (10) in §2.3. Typically the effect of the filter will be to enhance edges in images. Thus, we have combined intensity matching and edge matching in one measure, unlike, for example, Hallinan et al. (1999), who treated them separately. Note, our approach is different from those of others, who have used local Fourier methods, such as Gabor filters (Lades et al., 1993) and frequency-varying chirp-like filters (Bonmassar and Schwartz, 1997; Tabernero et al., 1999) to extract landmarks from images.

Secondly, we specify $A^{(Y_f)}$ to be an array of constants because most information about the warping, f, is contained in $\theta^{(Y_f)}$ rather than in $A^{(Y_f)}$, and also because in general it is difficult to construct a realistic stochastic model for $A^{(Y_f)}$. In particular, when f is simply a translation function, as is appropriate for Problem 2, the microscopy images in §1, all the information is in $\theta^{(Y_f)}$. This follows because there is a simple relationship between the Fourier transforms of Y and Y_f, given by

$$A_\omega^{(Y_f)} = A_\omega^{(Y)}, \quad \theta_\omega^{(Y_f)} = \theta_\omega^{(Y)} + 2\pi\omega^T\alpha \quad \forall \omega \in \Omega, \qquad \text{where } f_i = \alpha_i + x_i \mod n_i, \qquad i = 1, 2, \tag{12}$$

for constants α_1 and α_2, provided we allow modulo n wrap-round in the translation.

Thirdly, a von Mises distribution is a natural choice for an angular variable such as $\theta^{(Y_f)}$, as it is, in many ways, the circular equivalent of the Gaussian distribution. It can also be derived from the Gaussian model (3), because then

$$A_\omega^{(Y_f)} \cos\theta_\omega^{(Y_f)} \; \sim \; N(A_\omega^{(\mu)} \cos\theta_\omega^{(\mu)} \; , \; \sigma^2), \qquad A_\omega^{(Y_f)} \sin\theta_\omega^{(Y_f)} \; \sim \; N(A_\omega^{(\mu)} \sin\theta_\omega^{(\mu)} \; , \; \sigma^2), \qquad \forall \omega \in \Omega.$$

These are all independently distributed terms, except for the symmetry constraints already mentioned, which we will ignore as they simply introduce a scaling factor of one-half into the

final log-likelihood. The joint probability density over all frequencies, ω, including the Jacobian of the transformation to $\left(A^{(Y_f)}, \theta^{(Y_f)}\right)$, is

$$p\left(A^{(Y_f)}, \theta^{(Y_f)}\right) \propto \prod_{\omega} A_{\omega}^{(Y_f)} \exp\left[\frac{-1}{2\sigma^2}\left\{\left(A_{\omega}^{(Y_f)}\right)^2 + \left(A_{\omega}^{(\mu)}\right)^2 - 2A_{\omega}^{(\mu)}A_{\omega}^{(Y_f)}\cos\left(\theta_{\omega}^{(Y_f)} - \theta_{\omega}^{(\mu)}\right)\right\}\right].$$

We see that the distribution for $\theta^{(Y_f)}$, conditional on $A^{(Y_f)}$, is independent von Mises with concentration, $\kappa = A^{(\mu)}A^{(Y_f)}/\sigma^2$.

Fourthly, rather than restricting κ to this form, we generalise to the log-linear model given in (7). Fisher and Lee (1992) also modelled the concentration of circular data using log-linear models. This choice ensures positivity, is isotropic, and some experimentation indicated that a quadratic function in $|\omega|$ is sufficiently flexible to model the observed patterns of κ in our examples. Terms in $A^{(\mu)}$ and $A^{(Y_f)}$ are included, both because we would expect phases to be less susceptible to sampling variability when amplitudes are large, and because it leads to some standard models as special cases, as already discussed. In applications where there is near colinearity in the explanatory variables there will be some redundancy in this model and lack of identifiability in ξ, but this should not affect estimation of f. Other forms of κ have been considered, particularly in the one-dimensional case of signal processing. Hamon and Hannan (1974) showed that the optimal choice is $\kappa_\omega = c_\omega^2/(1 - c_\omega^2)$, where c^2 is the coherence between two series, which they estimated nonparametrically. See also Hannan and Thomson (1988), and on the subject to subpixel alignment, Berman et al. (1994), to which we will return in §3.2.

2.2 Null-set distortion criteria

We formulate distortion criteria, D, that are uniquely minimised by particular null sets of functions, $\mathcal{C}$. To motivate this approach, consider Problem 3 introduced in §1, the discrimination of fish species. Fish are not rigid bodies, so shape comparisons should allow for small distortions. Fig 4 shows an abstracted version of this problem, with the object in Fig 4(b) having the same shape as the triangle in Fig 4(a) except for a small non-linear deformation. Figs 4(c) and (d) show a second, differently shaped triangle, and a non-linearly deformed version. (Gaussian white noise has been added to all four images, for reasons that will become apparent in §3.3.) We need a distortion criterion that penalises warpings that align Figs 4(a) and (c) more than those that align Figs 4(a) and (b). As far as we are aware, no existing distortion criterion is tailored to this problem. For example, the thin-plate-spline distortion criterion (see D_{B_2} in (16), below), does not penalise affine transformations such as the one that would warp Fig 4(a) to align exactly with Fig 4(c). In order to penalise such an affine transformation, we construct $D(f, \mathcal{C})$, taking for $\mathcal{C}$ the set of Euclidean similarity transformations (see (20), below). By making λ arbitrarily large in (1), the function that maximises F will be a similarity transformation, and as λ is reduced, warpings are obtained which are nonparametric departures of increasing magnitude from this transformation.

Let $D_B(f)$ be a functional, called the *base distortion criterion*, such that

$$D_B(f) \geq 0, \quad D_B(0) = 0. \tag{13}$$

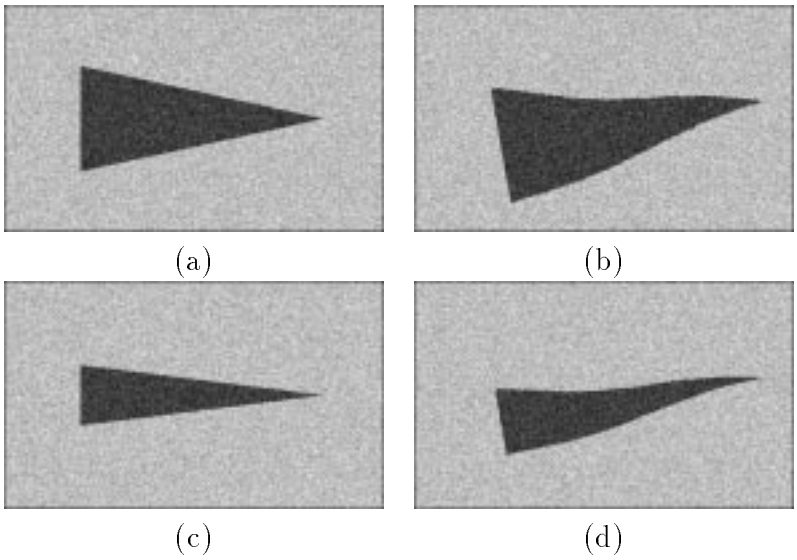


Figure 4: *Test images:* **(a)** *and* **(b)** *are triangles of the same shape, except for a smooth deformation of (a) into (b), and similarly* **(c)** *and* **(d)**.

We define the *null-set distortion criterion* to be

$$D(f,\mathcal{C}) = \min_{g\in\mathcal{C}} D_B(f-g). \tag{14}$$

Therefore, for $f \in \mathcal{C}$, $D(f,\mathcal{C}) = 0$, and by an appropriate choice of D_B we seek to ensure that

$$D(f,\mathcal{C}) > 0 \quad \text{for } f \notin \mathcal{C}.$$

If, as will usually be the case, $0 \in \mathcal{C}$, then $D(f,\mathcal{C}) \le D_B(f)$, and a necessary condition for D_B is that $\{f : D_B(f) = 0\} \subseteq \mathcal{C}$. Silverman (1982, pp. 116-117) was the first to look for criteria that were zero if and only if f was in a specified set of functions: for imaging applications, see Arad et al. (1994) and Hallinan et al. (1999, ch. 4). Our idea is qualitatively similar, but different in that we construct the functional which annihilates a specific set of functions. This enables us to construct many criteria, each appropriate to a specific set of imaging applications.

For the null set, $\mathcal{C}$, we consider subsets of pth-order polynomial transformations:

$$g_i = \alpha_i + \sum_{j_1=1}^{2} \alpha_{ij_1} x_{j_1} + \sum_{j_1,j_2} \alpha_{ij_1j_2} x_{j_1} x_{j_2} + \cdots + \sum_{j_1,\ldots,j_p} \alpha_{ij_1\ldots j_p} x_{j_1} \ldots x_{j_p} \qquad \text{for } i = 1,2. \tag{15}$$

Low-order polynomials occur repeatedly in image warping applications, including third and higher orders in registration of remotely sensed images (see Glasbey and Mardia, 1998, for specific references). For the base distortion criterion, D_B, we use functionals of partial derivatives, such as the following first and second partial derivatives:

$$D_{B_1}(f) = \sum_{i=1}^{2}\sum_{j=1}^{2} \int_{\square} \left(\frac{\partial f_i}{\partial x_j}\right)^2 dx, \qquad D_{B_2}(f) = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} \int_{\square} \left(\frac{\partial^2 f_i}{\partial x_j \partial x_k}\right)^2 dx, \tag{16}$$

integrated over domain $\Box = (0, n_1) \times (0, n_2)$, where dx denotes $dx_1\ dx_2$. (In this section it is convenient to treat image domains as continuous.) Both these functionals satisfy condition (13) and have been proposed many times: D_{B_1} is referred to as the Gaussian prior (Hallinan et al., 1999, p. 93) and D_{B_2} is the bending energy of a pair of thin-plate splines in a finite window (see Green and Silverman, 1994, pp.150-155). If the domains of integration were $\Re^2$, both functionals would be translationally and rotationally invariant, particular cases of the functionals considered by Wahba (1990). By specifying distortion using first partial derivatives, warpings are produced which are similar to the deformations of elastic membranes, and can have discontinuous second derivatives. For a detailed treatment of such penalties from a general viewpoint see Blake and Zisserman (1987). As with snakes, which are linear templates that deform smoothly to align with features in images (Kass et al., 1988), first-order derivatives can be regarded as tension constraints and second-order derivatives as rigidity constraints.

By combining (14), (15) and (16), many null-set distortion criteria are produced, most of which are new, and add to the range of first- and second-derivative functionals used by others. For example, if we choose for $\mathcal{C}$ the set of bilinear transformations

$$\mathcal{B} = \{g : g_i = \alpha_i + \alpha_{i1}x_1 + \alpha_{i2}x_2 + \alpha_{i12}x_1x_2, \quad i = 1, 2\}.$$

and use D_{B_2}, then

$$D(f, \mathcal{B}) = \min_{\alpha_{112}, \alpha_{212}} \sum_{i,j,k} \int_{\Box} \left(\frac{\partial^2 f_i}{\partial x_j \partial x_k} - \alpha_{ijk} \right)^2 dx,$$

where $\alpha_{i11} = \alpha_{i22} = 0$. It can be seen that $D(f, \mathcal{B}) = 0$ if and only if $f \in \mathcal{B}$. The minimising values of α_{i12} are

$$\tilde{\alpha}_{i12} = \frac{1}{n_1 n_2} \int_{\Box} \frac{\partial^2 f_i}{\partial x_1 \partial x_2}\ dx \qquad i = 1, 2,$$

producing

$$D(f, \mathcal{B}) = D_{B_2}(f) - 2n_1 n_2(\tilde{\alpha}_{112}^2 + \tilde{\alpha}_{212}^2). \tag{17}$$

If instead we were to choose the set of affine transformations

$$\mathcal{A} = \{g : g_i = \alpha_i + \alpha_{i1}x_1 + \alpha_{i2}x_2, \quad i = 1, 2\}, \tag{18}$$

then, as is well known, $\{f : D_{B_2}(f) = 0\} = \mathcal{A}$, and therefore

$$D(f, \mathcal{A}) \quad = \quad \min_{g \in A} D_{B_2}(f - g) \quad = \quad D_{B_2}(f). \tag{19}$$

If, for the null set, $\mathcal{C}$, we wish to consider a subset of $\mathcal{A}$, such as a translation, a translation in combination with either a rotation or scaling, or the Euclidean similarity transformation

$$\mathcal{S} = \{g : g_1 = \alpha_1 + \alpha_{11}x_1 + \alpha_{12}x_2\ , \quad g_2 = \alpha_2 - \alpha_{12}x_1 + \alpha_{11}x_2\ \}, \tag{20}$$

then we cannot use D_{B_2} alone as the base distortion criterion, because $\{f : D_{B_2}(f) = 0\} \not\subseteq \mathcal{S}$. Instead we use D_{B_1}. Set $\mathcal{S}$ is important since shapes are defined to be invariant under these transformations (see, for example, Dryden and Mardia, 1998). We have

$$D(f, \mathcal{S}) = \min_{\alpha_{11}, \alpha_{12}} \sum_{i,j} \int_{\Box} \left(\frac{\partial f_i}{\partial x_j} - \alpha_{ij} \right)^2 dx,$$

where $\alpha_{21} = -\alpha_{12}$ and $\alpha_{22} = \alpha_{11}$. Again, $D(f, \mathcal{S}) = 0$ if and only if $f \in \mathcal{S}$, so we have an appropriate distortion criterion. The minimising values of α_{11} and α_{12} are

$$\tilde{\alpha}_{11} = \frac{1}{2n_1 n_2} \int_{\square} \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} \right) \, dx, \qquad \tilde{\alpha}_{12} = \frac{1}{2n_1 n_2} \int_{\square} \left(\frac{\partial f_1}{\partial x_2} - \frac{\partial f_2}{\partial x_1} \right) \, dx,$$

producing

$$D(f, \mathcal{S}) = D_{B_1}(f) - 2n_1 n_2 (\tilde{\alpha}_{11}^2 + \tilde{\alpha}_{12}^2). \tag{21}$$

To illustrate, if $g \in \mathcal{A}$, given by (18), then

$$D_{B_1}(g) = n_1 n_2 (\alpha_{11}^2 + \alpha_{12}^2 + \alpha_{21}^2 + \alpha_{22}^2), \qquad \tilde{\alpha}_{11} = \frac{1}{2}(\alpha_{11} + \alpha_{22}) \qquad \tilde{\alpha}_{12} = \frac{1}{2}(\alpha_{12} - \alpha_{21}),$$

and

$$D(g, \mathcal{S}) = \frac{n_1 n_2}{2} \left[(\alpha_{11} - \alpha_{22})^2 + (\alpha_{12} + \alpha_{21})^2 \right],$$

which is zero if and only if $\alpha_{11} = \alpha_{22}$ and $\alpha_{12} = -\alpha_{21}$, the constraints for $g \in \mathcal{S}$, given by (20).

For other subsets of $\mathcal{A}$, we can similarly derive $D(f, \mathcal{C})$ based on (14) using D_{B_1}. We could also add a term involving D_{B_2} or higher-order derivatives to $D(f, \mathcal{C})$, in order to constrain f to have a continuous first derivative, while still retaining the property that $D(f, \mathcal{C})$ is zero if and only if $f \in \mathcal{C}$. Therefore, it is important to note that, although our null-set distortion criterion, $D(f, \mathcal{C})$, is uniquely minimised by $f \in \mathcal{C}$, $D(f, \mathcal{C})$ is not itself unique: there are many alternative distortion criteria with the same property.

2.3 Optimisation algorithm

We first consider maximising P, given by (1), then the multi-image problem of maximising $P^{(K)}$, given by (2). The maximisation of P, with respect to f and parameters ξ, has, in general, no known analytic solution for the functionals we have considered in §2.1 and §2.2. Therefore, we have to resort to numerical methods.

Numerically, we approximate f by specifying its values at a $(q_1 + 1) \times (q_2 + 1)$ lattice of points:

$$f\left(\frac{k_1 n_1}{q_1}, \frac{k_2 n_2}{q_2} \right) = \beta_k \qquad k_1 = 0, \ldots, q_1, \quad k_2 = 0, \ldots, q_2, \tag{22}$$

involving an array of parameters, β, and interpolate $f(x)$ elsewhere using a piecewise bilinear transformation:

$$f(x) = \beta_k + \beta_k^{+0} \left(\frac{x_1 q_1}{n_1} - k_1 \right) + \beta_k^{0+} \left(\frac{x_2 q_2}{n_2} - k_2 \right) + \beta_k^{++} \left(\frac{x_1 q_1}{n_1} - k_1 \right) \left(\frac{x_2 q_2}{n_2} - k_2 \right). \tag{23}$$

Here

$$k_i = \text{int}\left[\frac{x_i q_i}{n_i} \right] \qquad i = 1, 2,$$

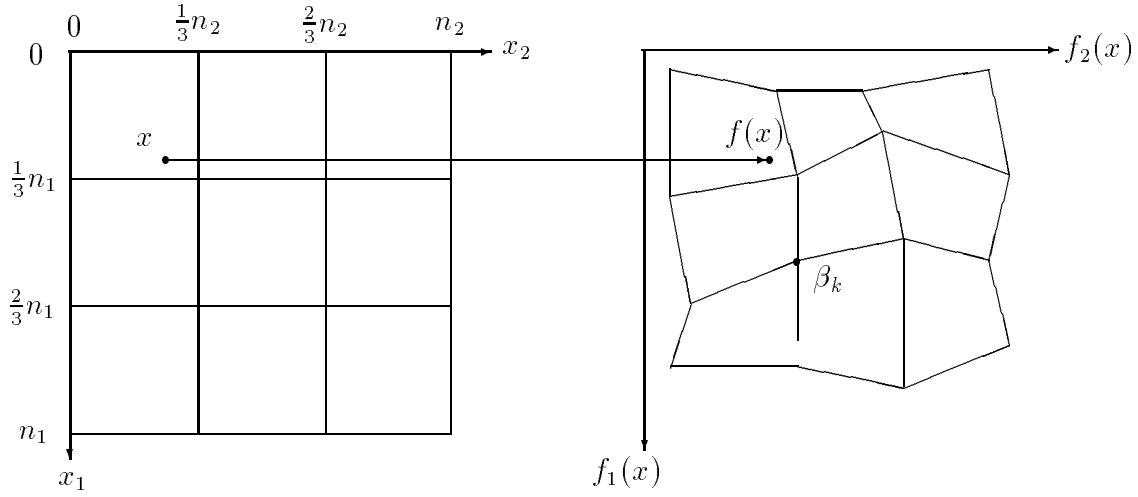


Figure 5: *Illustration of a piecewise bilinear approximation to f for a 3×3 grid.*

with 'int$[z]$' used to denote the integer part of z, and

$$\beta_k^{+0} = \beta_{k+(1,0)} - \beta_k, \qquad \beta_k^{0+} = \beta_{k+(0,1)} - \beta_k, \qquad \beta_k^{++} = \beta_{k+(1,1)} - \beta_{k+(1,0)} - \beta_{k+(0,1)} + \beta_k.$$

Alternatively we could have interpolated using B-splines (Rueckert et al., 1999). Fig 5 illustrates the case when $q_1 = q_2 = 3$.

For a piecewise bilinear transformation, it is straightforward to evaluate first derivative terms in our null-set distortion criteria. For example,

$$\begin{aligned} \int_{\square} \frac{\partial f}{\partial x_1}\, dx &= \frac{n_2}{q_2} \sum_{k_1=0}^{q_1-1} \sum_{k_2=0}^{q_2-1} \left(\beta_k^{+0} + \frac{1}{2}\beta_k^{++} \right), \\ \int_{\square} \left(\frac{\partial f}{\partial x_1} \right)^2 dx &= \frac{n_2}{q_2} \sum_{k_1=0}^{q_1-1} \sum_{k_2=0}^{q_2-1} \left\{ \left(\beta_k^{+0}\right)^2 + \beta_k^{+0}\beta_k^{++} + \frac{1}{3}\left(\beta_k^{++}\right)^2 \right\}, \end{aligned}$$

and other terms in $D(f, \mathcal{S})$, given by (21), can be similarly computed. If D involves derivatives of higher order, these can only be approximated. For example, $D(f, \mathcal{A})$ given by (19), can be approximated as:

$$D(f, \mathcal{A}) \approx \frac{q_1^3 n_2}{n_1^3 q_2} \sum_{k=(1,0)}^{q-(1,0)} (\beta_k^{+0} - \beta_{k-(1,0)}^{+0})^2 + 2\frac{q_1 q_2}{n_1 n_2} \sum_{k=(0,0)}^{q-(1,1)} (\beta_k^{++})^2 + \frac{n_1 q_2^3}{q_1 n_2^3} \sum_{k=(0,1)}^{q-(0,1)} (\beta_k^{0+} - \beta_{k-(0,1)}^{0+})^2,$$

using an abbreviated summation notation. Incidentally, this particular distortion criterion can be interpreted as a negative log-density of a Gaussian Markov random field on the $(q_1 + 1) \times (q_2 + 1)$ lattice. In all cases, $D(f, \mathcal{C})$ is a positive definite quadratic form in β.

We use a conjugate gradient method to maximise F (see, for example, Press, 1994). This is a general optimisation algorithm which requires only first partial derivatives. At each iteration, the search direction is that of steepest ascent, modified by the previous search direction, in such a way that, if the function were an n-dimensional quadratic, it would be optimised in n

steps. This algorithm is well suited to our problem, as first partial derivatives can be obtained relatively simply, as:

$$\frac{\partial P}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\sum_x \mu_x^{<\xi>}\, Y_{f(x)} - \sum_\omega \log I_0(\kappa_\omega(\xi)) - \lambda D(f,\mathcal{C})\right) \approx \sum_x \mu_x^{<\xi>} \frac{\partial Y_{f(x)}}{\partial \beta} - \lambda \frac{\partial D(f,\mathcal{C})}{\partial \beta}, \tag{24}$$

using the formulation of the Fourier-von Mises log-likelihood, L, given by (10), and

$$\begin{aligned}\frac{\partial P}{\partial \xi} &= \frac{\partial}{\partial \xi}\left(\sum_\omega \kappa_\omega(\xi)\cos\left(\theta_\omega^{(Y_f)} - \theta_\omega^{(\mu)}\right) - \sum_\omega \log I_0(\kappa_\omega(\xi))\right) \\ &= \sum_\omega \frac{\partial \kappa_\omega(\xi)}{\partial \xi}\cos\left(\theta_\omega^{(Y_f)} - \theta_\omega^{(\mu)}\right) - \sum_\omega \frac{\partial \log I_0(\kappa_\omega(\xi))}{\partial \xi},\end{aligned} \tag{25}$$

using the formulation of L given by (8). Derivatives are computed using difference methods, taking advantage of changes in β_k only affecting a subset of terms in Y_f and D, and we achieve substantial gains in speed by ignoring the second-order dependence of $\mu^{<\xi>}$ and $\kappa(\xi)$ on β. Various strategies can be adopted to guard against becoming trapped in local sub-optima. These include a multiresolution approach, where q is increased as iterations proceed, and permitting greater distortion by decreasing λ as iterations proceed. For the examples in §3, the algorithm typically took thirty minutes of CPU time on a single processor of a SUN Enterprise 450 using Fortran77. However, parallelisation would considerably reduce this time.

In applications where it is important, bijectivity can be ensured. Necessary and sufficient conditions for the piecewise bilinear transformation to be bijective are that the transformed boundary does not self-intersect, and that each quadrilateral, specified by the ordered set of four vertices β_k, $\beta_{k+(1,0)}$, $\beta_{k+(1,1)}$, $\beta_{k+(0,1)}$, is convex, with the vertices ordered anticlockwise. Convexity ensures that the bilinear interpolant is bijective within quadrilaterals, (a result best seen geometrically by plotting the deformations of lines parallel to the axes), and the anti-clockwise constraint prevents the mapping from folding at the divisions between quadrilaterals. Computationally, convexity can be ensured by checking that the two diagonals intersect inside the quadrilateral.

To maximise $P^{(K)}$, given by (2), we alternate between estimating the consensus image, μ, and aligning each individual image with it, in a way analogous to Generalised Procrustes Analysis. Fluete and Lavallée (1998) used a similar method to align shape outlines and Ramsay and Li (1998) to align curves. If we use the log-likelihood, L^*, given by (4), then

$$P^{(K)} = -\sum_{k=1}^{K}\left\{\sum_x \left(Y^{(k)}_{f^{(k)}(x)} - \mu_x\right)^2 + \lambda D(f^{(k)},\mathcal{C})\right\} \tag{26}$$

is maximised with respect to μ simply by averaging at each pixel:

$$\hat{\mu}_x = \frac{1}{K}\sum_k Y^{(k)}_{f^{(k)}(x)}. \tag{27}$$

Alternatively, if we use the Fourier-von Mises log-likelihood, L, given by (8), then we can only estimate the phases of the Fourier transform of μ, denoted ζ, by

$$\hat{\zeta}_\omega = \tan^{-1}\left[\frac{\sum_k \kappa_\omega(\xi^{(k)})\sin\theta_\omega^{(Y_f^{(k)})}}{\sum_k \kappa_\omega(\xi^{(k)})\cos\theta_\omega^{(Y_f^{(k)})}}\right]. \tag{28}$$

In either case, given $\hat{\mu}$, each warping $f^{(k)}$ and $\xi^{(k)}$ can be estimated by maximising the component functional, $P^{(k)}$, using the conjugate gradients algorithm already discussed. Because $P^{(K)}$ increases at each iteration and is bounded above, the algorithm is guaranteed to converge. Note, however, $P^{(K)}$ need not have a unique maximum, and the solution can depend on the initial choice of an average image. In general, it will not be adequate to start by averaging all the unwarped images. We consider one solution for Problem 3, in §3.3.

3 Applications

We now apply the methodology developed in §2 to the three practical problems introduced in §1.

3.1 Problem 1: SAR registration

For Problem 1, we wish to align the SAR image, Fig 1(a), with the digital map, Fig 1(b), which it is natural to take as μ. The appropriate transformation is a projection,

$$\begin{aligned} f_1 &= \alpha_1 + \gamma\{x_1(-\cos\phi_1 \sin\phi_2 \sin\phi_3 + \cos\phi_2 \cos\phi_3) \\ &\qquad + x_2(-\cos\phi_1 \cos\phi_2 \sin\phi_3 - \sin\phi_2 \cos\phi_3) + h(x)\sin\phi_1 \sin\phi_3\} \\ f_2 &= \alpha_2 + \gamma\{x_1(+\cos\phi_1 \sin\phi_2 \cos\phi_3 + \cos\phi_2 \sin\phi_3) \\ &\qquad + x_2(+\cos\phi_1 \cos\phi_2 \cos\phi_3 - \sin\phi_2 \sin\phi_3) - h(x)\sin\phi_1 \cos\phi_3\}, \end{aligned} \tag{29}$$

as shown in Fig 6. In addition to the elevation function, $h : \Re^2 \mapsto \Re$, there are six unknown parameters, the translation and scale parameters (α, γ) and the Euler angles, ϕ. The first Euler angle, ϕ_1, is shown in Fig 6, and the other two relate to the orientations of the two sets of axes. If the ground were planar the transformation would be affine, as given by (18), but with a different parametrisation. We penalise non-linear functions, h, using the thin-plate spline distortion criterion $D(h, \mathcal{A})$, a one-dimensional version of that in (19).

We use the Fourier-von Mises image model (6). However, for Y, we use an edge-filtered version of the SAR image, as shown in Fig 7(a). Details are given in Glasbey (1997). It is not possible to subsume this filter in the Fourier-von Mises image model, because linear filters are incapable of transforming Fig 1(a) to an image that looks like Fig 1(b). However, our model does allow the fine-tuning, by linear filters, of this empirically-chosen edge-filter, to optimise its performance for the warping task.

The algorithm of §2.3 was used to maximise P with respect to f, given by (29), and parameters ξ. For simplicity, in this application we set $\xi_3 = \xi_4$. Experimentation with different values of q showed $q_1 = q_2 = 16$ to be sufficiently large to approximate f. The warping transformation was constrained to be bijective, although this need not be the case in this type of application: the presence of hills could lead to occlusions which would need to be taken into account also in the image model. A range of values of λ was used, as is common practice (see, for example, Silverman, 1986). Table 1 summarises the results. It can be seen that, as λ decreases, P, L

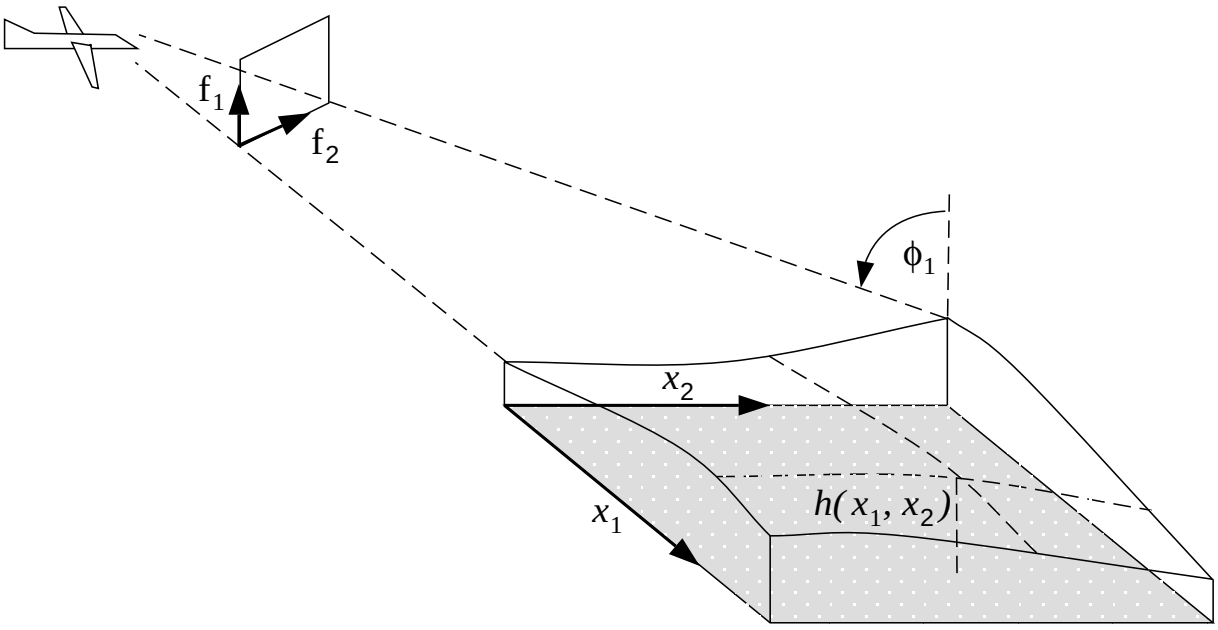
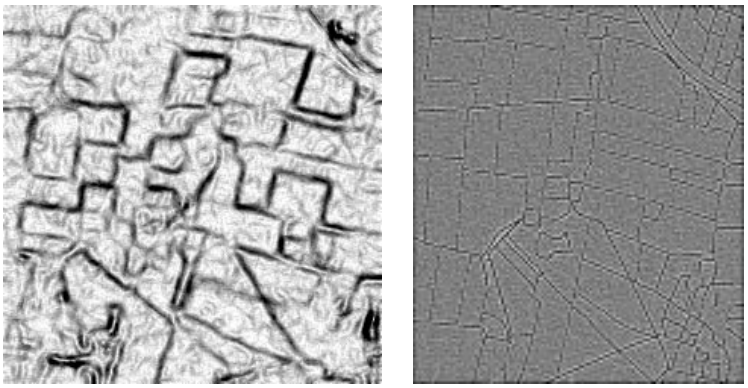


Figure 6: *Illustration of the projective transform from the map to the SAR image in Fig. 1.*



(a) (b)

Figure 7: *Filtered SAR image and map, with larger values displayed as darker shades of grey:* **(a)** *edge-filtered SAR image,* **(b)** *filtered map, given by* $\mu^{<\xi>}$.

and $\lambda D(h, \mathcal{A})$ all increase, except for some tailing off in $\lambda D(h, \mathcal{A})$ for the smallest values of λ. We used a cross-validatory approach to choose λ, by estimating f with a 50×50 block of pixels in the 250×250 array, Y, set to a constant mean value, then evaluating the covariance between the complete images. This was repeated for each of the 9 blocks with pixel locations x_1 and x_2 in the range 51 to 100, 101 to 150, or 151 to 200. We chose this size of block, rather than individual pixels, because adjacent pixels are likely to be correlated, and to reduce the computational effort. The final column of Table 1 gives the results, from which it can be seen that $\lambda = 100$ appears to be best.

For $\lambda = 100$, we obtained $\hat{\xi} = (-2.9, \quad 15, \quad -61, \quad 0.45, \quad 0.45)$. The sample mean-resultant-

λ	P	L	$\lambda D(h, \mathcal{A})$	cross-validated covariance
100 000	285	286	1	34.5
30 000	290	293	4	34.9
10 000	302	312	10	35.6
3 000	321	335	14	36.2
1 000	343	363	20	37.4
300	367	406	38	37.2
100	415	513	97	**40.4**
30	513	683	170	34.0
10	599	817	216	31.6
3	869	1084	215	21.9
1	1086	1244	158	31.1

Table 1: *The effect of varying λ on criteria for aligning SAR image with map.*

length was calculated for a range of values of the non-directional frequency, $|\omega|$,

$$\frac{1}{N(\Lambda_{|\omega|})} \sum_{\nu \in \Lambda_{|\omega|}} \cos\left(\theta_\nu^{(Y_f)} - \theta_\nu^{(\mu)}\right), \qquad \text{where } \Lambda_{|\omega|} = \{\nu \ : \ | \ |\nu| - |\omega| \ | < 0.05\} \tag{30}$$

and $N(\Lambda_{|\omega|})$ denotes the number of elements in set $\Lambda_{|\omega|}$. According to the von Mises model, the mean resultant length has expectation

$$\frac{1}{N(\Lambda_{|\omega|})} \sum_{\nu \in \Lambda_{|\omega|}} \frac{I_1(\kappa_\nu)}{I_0(\kappa_\nu)}, \tag{31}$$

where I_0 and I_1 are Bessel functions. Fig 8 shows these sample and expected values plotted against $|\omega|$, from which we see that the agreement is excellent. For comparison, the expected values for the best-fitting model of the form, $\xi = (\xi_0, 0, 0, 1, 1)$, with $\hat{\xi}_0 = -4.6$, is also shown. The value of $\hat{\xi}_3 = \hat{\xi}_4 = 0.45$ in the full model suggests that the best choice of model leads to a measure of similarity which is a half-way compromise between covariance and phase correlation, i.e., a *band-pass filter*. In contrast, Koch and Snowdon (1994) advocated the use of a low-pass filter in an application involving alignment of X-ray images. The filtered map, denoted $\mu^{<\xi>}$, as defined in (11), is displayed in Fig 7(b).

Fig 9 shows the SAR image registered with the digital map, obtained by applying the estimated warping to the original SAR image. The alignment can be seen to be very good, and automatically yields an almost complete segmentation of the image into homogeneous regions.

3.2 Problem 2: Multimodal microscopy

For Problem 2, we know *a priori* that a translation,

$$f = \alpha + x, \tag{32}$$

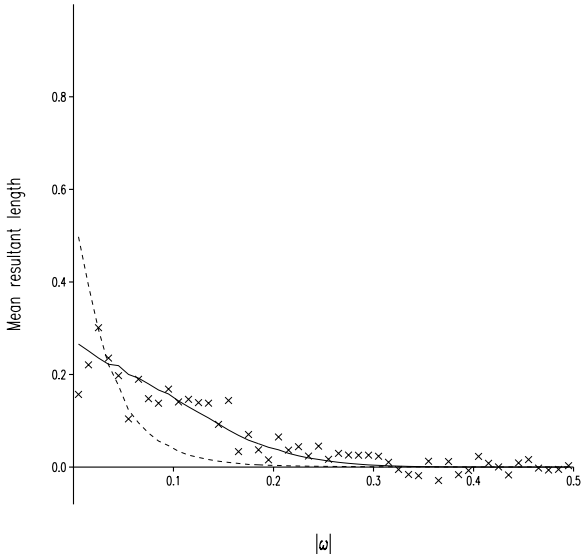


Figure 8: *Mean resultant length for SAR edge-filtered image and map, averaged over all orientations, plotted against* $|\omega|$*:* $\times$ *denotes sample values, obtained using (30),* — *denotes expected values from the full model, obtained using (31), and - - - shows expected values when* $\xi = (\xi_0, 0, 0, 1, 1)$.

Figure 9: *Superposition of aligned SAR image with map.*

is sufficient to align any pair of microscope images. Therefore, formally, we choose the null-set distortion criterion D to be uniquely minimised by translations, using the method of §2.2, but also take $\lambda \to \infty$. In practice, we simply use the parametric transformation.

By combining (12) and (8), the Fourier-von Mises log-likelihood can be re-expressed as

$$L(Y|\mu, f, \xi) = \sum_{\omega} \kappa_\omega(\xi) \cos\left(\theta_\omega^{(Y)} - \theta_\omega^{(\mu)} + 2\pi\omega^T\alpha\right) - \sum_{\omega} \log I_0(\kappa_\omega(\xi)), \tag{33}$$

provided we allow modulo n wrap-round in the translation. L can be evaluated simultaneously for all integer values of α by a single Fast Fourier Transform. For Problem 2, alignment to the nearest pixel is sufficiently accurate and is probably all that is achievable. The model is a two-dimensional variant of the 'barber's pole' proposed by Gould (1969). In some applications it is possible to estimate α to subpixel accuracy, especially if adjustments to take account of aliasing, proposed by Berman et al. (1994), are also included.

We wish to simultaneously align all three microscopy images under translation. However, it turns out that we cannot maximise $P^{(K)}$ with respect to parameters in both the concentration function and the consensus image. This is similar to the Neyman-Scott problem (see, for example, Stuart et al., 1999, pp. 80-81). So, instead we propose to maximise a pseudo-log-likelihood:

$$P^* = \sum_{k<l} L(Y^{(l)}|Y^{(k)}, f^{(k,l)}, \xi^{(k,l)}), \tag{34}$$

with respect to $\alpha^{(k,l)}$, which specifies $f^{(k,l)}$ as given in (32), and $\xi^{(k,l)}$, subject to constraints

$$f^{(k,m)} = f^{(k,l)} \circ f^{(l,m)} \qquad \forall k < l < m, \tag{35}$$

where '$\circ$' denotes a composite of functions. This construction eliminates the consensus image, μ. In general, these constraints are difficult to enforce, but for parametric transformations they take simple forms, In particular, for translations

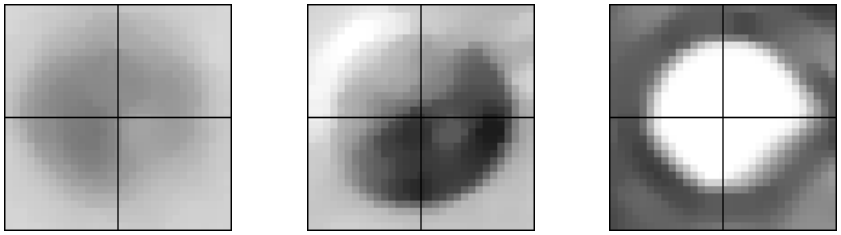
$$\alpha_i^{(1,3)} = \alpha_i^{(1,2)} + \alpha_i^{(2,3)} \mod n_i, \qquad i = 1, 2. \tag{36}$$

Also, when f is a translation, Y and μ are interchangeable in L, given by (33), so we need only consider all unordered pairs in P^*.

We use the conjugate gradient method described in §2.3, to maximise P^* with respect to $\xi^{(1,2)}$, $\xi^{(1,3)}$ and $\xi^{(2,3)}$, but for each value of ξ we conduct a grid search to estimate α. An exhaustive search would have to consider $n_1^2 n_2^2$ possibilities. Therefore, we approximate by a local optimum, by only searching values around $\text{argmax}_{\alpha^{(1,2)}}(L^{(1,2)})$ and $\text{argmax}_{\alpha^{(2,3)}}(L^{(2,3)})$. Similarly, we consider values around each of the other two pairs of maxima. Table 2 gives the results, which agree with those reported in Glasbey and Martin (1996), using an *ad hoc* similarity criterion. So, for example, we estimate that Fig 2(b) needs to be shifted down by 3 rows and shifted right by 6 columns to align with Fig 2(a). Fig 10 shows a single algal cell, after alignment, in the three microscope modalities. A cross-wire has been superimposed to aid comparison in alignments, which can be seen to be very good. The individual pixels can be discerned at this magnification, and it can be appreciated that even a shift as small as 3 rows and 6 columns has a marked effect.

k	l	$\hat{\alpha}_1^{(k,l)}$	$\hat{\alpha}_2^{(k,l)}$	$\hat{\xi}_0^{(k,l)}$	$\hat{\xi}_1^{(k,l)}$	$\hat{\xi}_2^{(k,l)}$	$\hat{\xi}_3^{(k,l)}$	$\hat{\xi}_4^{(k,l)}$
1	2	3	6	−6	43	−118	0.91	0.28
1	3	28	170	−20	616	−14000	0.72	1.53
2	3	25	164	−17	502	−11900	0.65	1.33

Table 2: *Parameter estimates to align microscope images.*



(a) (b) (c)

Figure 10: *Single algal cell, after alignment of three microscope images, and with a cross-wire superimposed:* **(a)** *brightfield,* **(b)** *DIC,* **(c)** *phase contrast.*

We compare our method of alignment with two alternatives, the covariance and phase-correlation criteria. Each of 70 sub-images of Fig 2(a) (128 × 192 pixels in size) was aligned with a sub-image of Fig 2(b) *shifted* by 20 rows and 20 columns. For the new criterion and covariance and phase-correlation criteria, the means and standard deviations of the 70 estimates of α were evaluated, as given in Table 3. We see that L produces by far the most consistent results, with standard deviations of less than one pixel. Also, in spite of the sub-images containing far less information than the full images, the estimated translation agrees well with the earlier results, which with the additional translation of (20,20) should now be (23,26).

3.3 Problem 3: Fish species discrimination

For Problem 3, it is natural to use the null-set distortion criterion based on the Euclidean similarity transformation, $D(f, \mathcal{S})$, given by (21), since this is then shape invariant. To assess our procedure, we first apply the method to the synthetic example of triangles in Fig 4.

Similarity criterion	mean		s.d.	
	α_1	α_2	α_1	α_2
Covariance	17.5	22.7	8.0	4.3
Phase-correlation	21.2	23.6	11.4	23.4
Fourier-von Mises log-likelihood, L	22.8	25.8	0.7	0.7

Table 3: *Summary of results for estimating translation parameters to align 70 different 128 × 192 sub-images of Fig 2(a) with sub-images of Fig 2(b) shifted by 20 rows and 20 columns, using three similarity criteria.*

	$P = L^* - \lambda D(f, \mathcal{S})$		
	mean (s.d.)		
λ	within-shape	between-shape	Studentised difference in means
1	−2090 (15.6)	−2662 (159.0)	3.6
0.3	−1972 (5.7)	−2397 (64.6)	**6.6**
0.1	−1865 (9.1)	−2142 (55.4)	4.9
0.03	−1743 (8.8)	−1869 (26.4)	4.5
0.01	−1633 (8.9)	−1694 (10.6)	4.4
0.003	−1546 (3.2)	−1567 (7.1)	2.7
0.001	−1498 (3.2)	−1510 (3.9)	2.3

Table 4: *The effect of varying λ on discrimination between the two shapes of triangle in Fig 4.*

We propose to use the Gaussian image model with log-likelihood, L^*, given by (4). The algorithm of §2.3 was used to align each image in Fig 4 with every other image, twelve ordered pairs in total, for each of a range of values of λ. Experimentation with different values of q showed $q_1 = q_2 = 64$ to be sufficiently large to approximate f and, again, the bijective constraint was used. Results are summarised in Table 4, by the average values of the criterion, P, for within- and between-shape comparisons. Within-shape comparisons are defined to be those between Figs 4(a) and (b), and between Figs 4(c) and (d), of which there are four. The remaining eight ordered pairs are regarded as between-shape comparisons. In both cases, P increases with λ, because the warping is progressively less constrained to be smooth and can therefore achieve greater agreement in pixel values between images. The criterion is smaller when two images of different shape are aligned than with two of the same shape, for all values of λ. Standard deviations of values of P are also given in Table 4, from which the Studentised difference between the means of the two groups can be computed, as the difference in the means divided by the square-root of the sum of the two variances. However, note that it is inappropriate to perform t-tests as samples are not independently distributed. The distances show that the two shapes are well discriminated, with the best choice shown emboldened for $\lambda = 0.3$. Thus we conclude that our method provides satisfactory answers for this simplified problem.

Table 5 compares the results we obtained using $D(f, \mathcal{S})$ with what we would have obtained if we had instead used the thin-plate-spline distortion criterion, D_{B_2}, having selected an appropriate value of $\lambda = 10^6$. In both cases, the largest values are shown emboldened. We see that, using D_{B_2}, Figs 4(a) and (c) are assessed as being most similar, which is as we would expect, as an affine transformation is sufficient to transform one triangle to the other and this is not penalised by D_{B_2}. It is clear that this distortion criterion will not enable us to discriminate between the two shapes of triangle. Similar, inadequate results will be produced using any distortion criterion other than $D(f, \mathcal{S})$.

The same algorithm was then used to align all pairs of images of fish in Fig 3 for each of a range of values of λ. Optimised values of P are given in Table 6, summarised as before. The two species are well discriminated, with marginally the best choice of λ being 0.01. Fig 11 illustrates the warping for this optimal choice of λ, for alignment of a haddock with another

	$P = L^* - 0.3D(f,\mathcal{S})$				$P = L^* - 10^6 D_{B_2}$			
	(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
(a)	*	**−1975**	−2309	−2350	*	−1906	**−1854**	−1908
(b)	**−1979**	*	−2348	−2357	−1903	*	−1912	−1910
(c)	−2475	−2470	*	**−1967**	**−1853**	−1894	*	−1893
(d)	−2409	−2458	**−1968**	*	−1911	−1911	−1890	*

Table 5: *Maximised values of penalised likelihood for pairwise comparison of images in Fig 4, using two distortion criteria.*

	$P = L^* - \lambda D(f,\mathcal{S})$		
	mean (s.d.)		
λ	within-species	between-species	Studentised difference in means
1	−643 (152)	−1064 (175)	1.8
0.3	−538 (163)	−952 (114)	2.1
0.1	−383 (138)	−784 (49)	2.7
0.03	−255 (92)	−586 (22)	3.5
0.01	−182 (60)	−433 (15)	**4.1**
0.003	−130 (40)	−298 (15)	3.9
0.001	−99 (29)	−218 (16)	3.6

Table 6: *The effect of varying λ on discrimination between haddock and whiting, using the four images in Fig 3.*

haddock, and with a whiting. Figs 11(a) and (b) show grids of the two estimated warps. The deformations in Fig 11(a) are less severe than in Fig 11(b), and the distortion is less when the two haddock are aligned than when a haddock and whiting are aligned. Fig 11(c) shows how haddock 1 (Fig 3(a)) is warped to look like haddock 2 (Fig 3(b)), and Fig 11(e) shows the pixel-by-pixel difference between the two images after alignment. In comparison, Fig 11(d) shows how haddock 1 is warped to look like whiting 1 (Fig 3(c)), and Fig 11(f) shows the pixel-by-pixel difference between the two images. The sum of squared differences is greater than for the within-species comparison.

We now consider the analysis of the larger data set, consisting of images of ten haddock and ten whiting. For each species, we used eight images to characterise the population average and variation, by maximising $P^{(K)}$, given by (26). Two images of each species, chosen at random, were then available to validate the method.

An *ad hoc* procedure to overcome some of the numerical problems in estimating the template, $\hat{\mu}$, for each species is as follows. To obtain an initial estimate of $\hat{\mu}$, we warped image 2 to image 1, and formed a composite image:

$$\hat{\mu}^{(1,2)}_{(x+f(x))/2} = \frac{Y^{(1)}_x + Y^{(2)}_{f(x)}}{2} \qquad \forall x \in X. \tag{37}$$

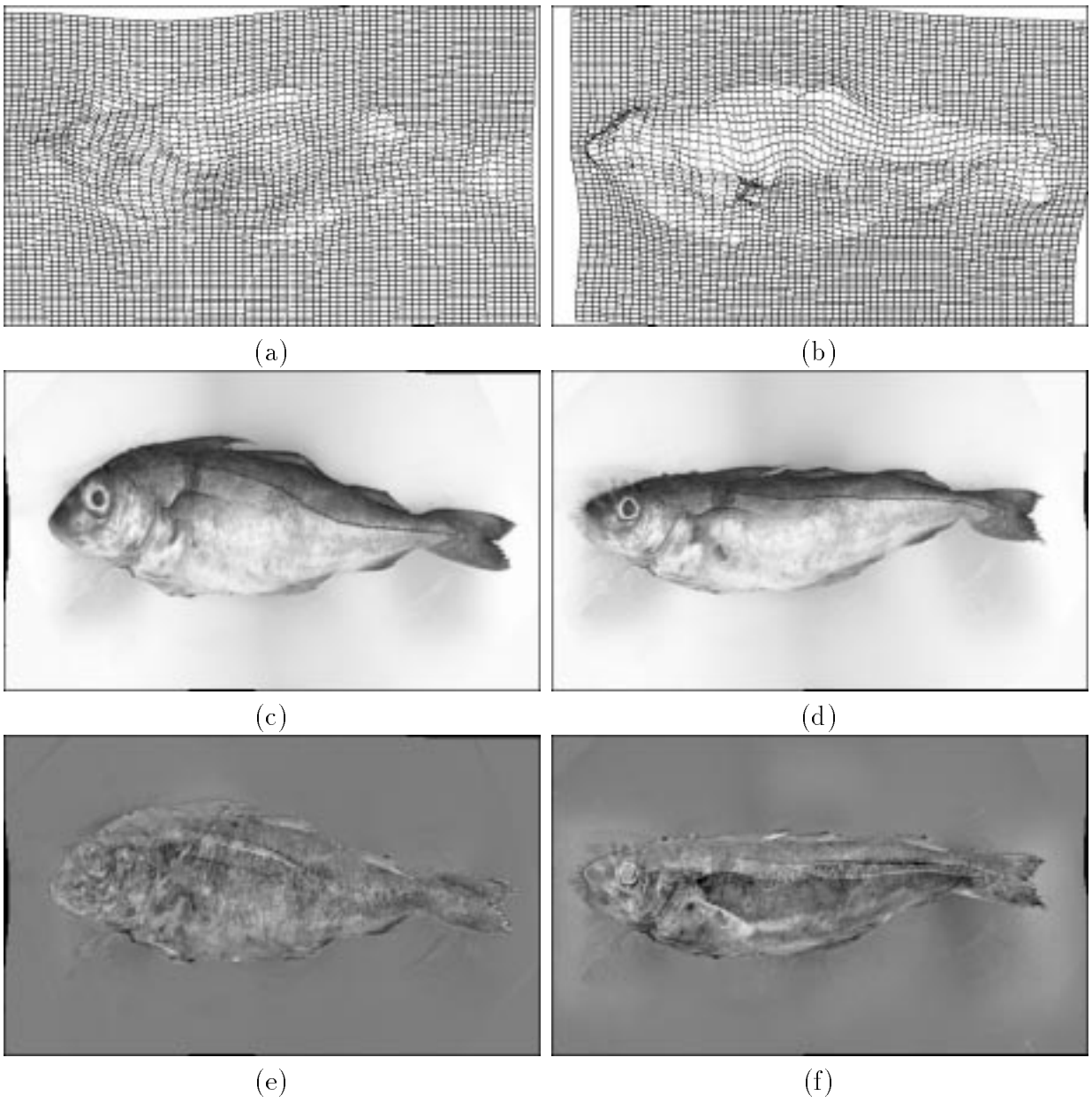
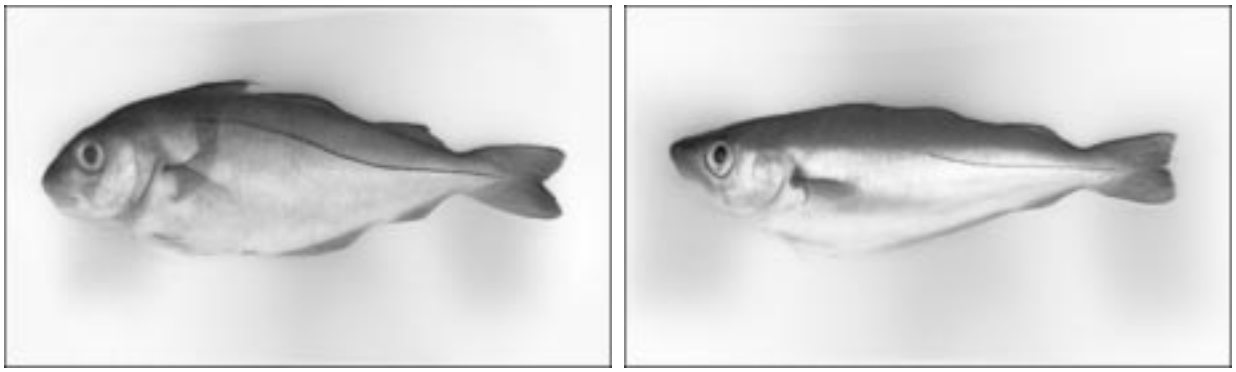


Figure 11: *Result of warping haddock 1, with* $\lambda = 0.01$. *Grid of deformations for alignment with:* **(a)** *haddock 2,* **(b)** *whiting 1. Warped image of haddock 1, to align with:* **(c)** *haddock 2,* **(d)** *whiting 1. Pixel-by-pixel difference (zero values displayed as mid-grey) between warped image of haddock 1 and:* **(e)** *haddock 2,* **(f)** *whiting 1.*



(a) (b)

Figure 12: *Average images of two species of fish, obtained by averaging eight images:* **(a)** *haddock,* **(b)** *whiting.*

Here, we have taken the average of the pixel at location x in image 1 and the pixel at location $f(x)$ in image 2, and assigned it to the pixel at location $(x + f(x))/2$ in the composite image. Unassigned pixels in $\hat{\mu}^{(1,2)}$ were given the same value as their nearest neighbour. We similarly formed the average of 3 and 4 and then averaged (1,2) with (3,4) to obtain ((1,2),(3,4)), and so on until finally,

$$\hat{\mu} = \hat{\mu}^{(\ ((1,2),(3,4)),\ ((5,6),(7,8))\ )}.$$

We then warped the eight original images to $\hat{\mu}$ to maximise $P^{(K)}$, re-estimated μ using (27), and repeated until convergence. This procedure treats all eight images equivalently, and could be modified to handle other sizes of training set. For both haddock and whiting, values stabilised within a couple of iterations. Fig 12 shows the two average fish.

Finally, each of the twenty images in turn was warped to the average haddock, by maximising P. Fig 13 shows the maximised values of P plotted against the corresponding values when the images were instead warped to the average whiting. The maximised values of the penalised likelihoods are measures of similarity of individual images from the two species. We can see two clusters of points, and the two species are clearly distinguishable. The ten haddock, including the two not used previously, are far more similar to the average haddock than to the average whiting, and a similar pattern occurs with the whiting. However, we see that the circled points lie on the extremes of the two clusters, nearer to the other cluster, indicating some slight overfitting in the model.

By making full use of the grey-level information (including texture) in the photographic images of fish, we have improved on the discriminating power of Strachan et al. (1990) and Glasbey et al. (1995). We could develop the model further, and apply principal components analysis, both to the grids of warpings, β given by (22), and to the differences between pixel values in the aligned images, as Lanitis et al. (1995) did with images of faces. These could be used to replace L^* by differences between principal component scores, thereby weighting pixel values according to how variable they are in different parts of an image (Cootes et al., 1998). Further, Moghaddam et al. (1996) modelled variation in facial expression, both within and between individuals, and Duta et al. (1999) formed clusters of similar shapes and modelled intra-cluster

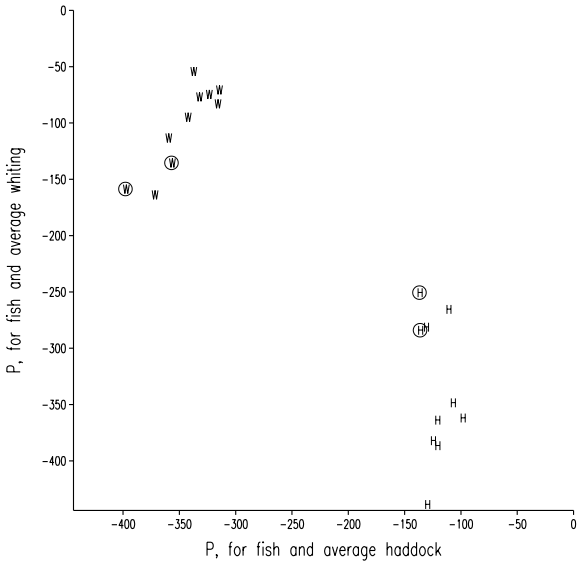


Figure 13: *Plot of P between individual images and the average haddock and the average whiting, with* `H` *denoting haddock,* `W` *denoting whiting, and the circled letters showing the four images which were not used in obtaining the averages.*

variation. Rao (2000, pp. 580-583) has discussed an unbiased and a consistent estimator of the template under certain conditions when the errors are coloured and the warps are elements of a similarity group.

4 Discussion

The subject of warping has been reviewed comprehensively in Glasbey and Mardia (1998) and elsewhere, as cited in §1. Therefore, here we will focus on the techniques developed in this paper. We have shown that image warping can be formulated statistically, as maximum penalised likelihood, and this has allowed us to understand and generalise existing methods. Our approach has produced good results for the three applied problems. We do not claim that ours are the only methods capable of producing such results, but there would seem to be little opportunity for alternative methods to improve on them. Also, it is clear from Table 3 that the Fourier-von Mises log-likelihood outperforms simpler measures of image similarity, and from Table 5 that the thin-plate-spline distortion criterion is inappropriate for making similarity shape comparisons, though also see point 2, below. Also, although we have focussed on point estimation, as we have formulated image warping in a statistical framework we could also obtain measures of precision of estimators. For example, we could base inference on multiple samples drawn from the Fourier-von Mises image model.

We now deal with the three main ingredients of the paper, the Fourier-von Mises image model, the null-set distortion criterion and the algorithm, respectively.

1. The Fourier-von Mises image model offers a flexible approach to modelling the relationship between images, which will work for general lighting conditions due to its Fourier basis. We have shown the model to be plausible and used it in both Problems 1 and 2. However, the general form is inappropriate to solving Problem 3, where we are concerned with discrimination between fish species, and a simple Gaussian model is what is required. Other grey level metrics, such as Kantorovich distance (Kaijser, 1998) are more computationally expensive. Our image model is elegant in that it combines intensity matching and edge matching in one measure. In the machine vision literature, the two terms for intensity matching and edge matching are treated separately (see, for example, Hallinan et al., 1999, p.79). We believe our procedure has some advantages since it removes the necessity of estimating the weights required to combine the two terms.

2. The null-set distortion criteria furnish us with a rich class. In Problem 1 we have taken the roughness penalty from thin-plate splines as our distortion criterion, whereas in Problem 2 the distortion criterion is used only implicitly to constrain the warp to be in the null set of translations ($\lambda \to \infty$). In Problem 3 the distortion criterion is a shape invariant criterion. Thus, our formulation allows us to select a criterion appropriate to the application within our general null class. Our shape invariant criterion uses only first derivatives, so there can be degenerate solutions if only a limited number of points such as landmarks are used (for example, Green and Silverman, 1994, p.159). However in our case there is always a unique solution for finite q since D is quadratic in β. Note that for landmark-based methods, there is explicit expression for kriging warps including thin-plate splines (see, for example, Mardia and Hainsworth, 1993; Kent and Mardia, 1994). However, we need to add an extra penalty term to the thin-plate spline criterion if we want to penalise affine transformations that are not shape-preserving.

3. Our algorithm has some similarities with finite difference methods, though our use of a piecewise bilinear transformation eliminates numerical integration in calculating distortion criteria only involving first derivatives. The conditions we have imposed on the piecewise bilinear transformation leads to local as well as global bijectivity. Bijectivity is important if we wish to warp a standard coordinate system to the image. However, in the SAR example, the projective transformation may not be bijective or continuous due to occlusion. The method extends in a straightforward manner to three and higher dimensions, although the computational cost will be high. Also, stochastic methods such as simulated annealing and MCMC could be implemented, which would also give more information on the posterior distributions of $\hat{\xi}$ etc., but again the method may prove expensive. The issue of global registration has not arisen because, for example, in Problem 3, the fish were placed in a pre-specified orientation with fixed camera position. If this were not so, we could resort to any of a number of global registration methods, such as the matching of low-order moments (Wong and Hall, 1978; Yang and Cohen, 1999). Alternatively, an additional penalty term could be added to P (see, for example, Mardia et al., 1997). We have taken different q for Problems 1–3. Its selection depends on the size of the images, and the overall accuracy required. A wavelet based distortion criterion in turn is another approach (see, for example, Downie et al., 1996). Also, whether one should use compositional warps at different resolutions or additive warps is another issue.

There remain many challenging problems in image analysis, to which statistical methods are

applicable, both in general and in particular in image warping. This Read Paper follows earlier, ground-breaking papers on image analysis by Besag (1986) and Grenander and Miller (1994). We hope that our paper similarly stimulates further work in this area.

Acknowledgements

Our profound thanks to John Kent for his generous comments on earlier drafts and, in particular, the suggestion to use (14) to formalise the method of construction of distortion criteria. Our thanks are also due to Ulf Grenander, Anil Jain, the RSS Research Section committee and anonymous referees for their comments, and to Kevin de Souza for help with preparation of the manuscript. We also thank Nick Martin and Norval Strachan for permission to use the microscopy and fish images respectively. The first author's work was supported by funds from the Scottish Executive Rural Affairs Department, and the second author was in receipt of an EPSRC grant.

References

Amit, Y., Grenander, U., and Piccioni, M. (1991). Structural image restoration through deformable templates. *Journal of the American Statistical Association*, 86:376–387.

Arad, N., Dyn, N., Reisfeld, D., and Yeshurun, Y. (1994). Image warping by radial basis functions: applications to facial expressions. *CVGIP: Graphical Models and Image Processing*, 56:161–172.

Baddeley, A. and Molchanov, I. (1998). Averaging of random sets based on their distance functions. *J. Maths. Imaging and Vision*, 8:79–92.

Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46:1–21.

Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77.

Berman, M., Bischof, L. M., Davies, S. J., Green, A. A., and Craig, M. (1994). Estimating band-to-band misregistrations in aliased imagery. *CVGIP: Graphical Models and Image Processing*, 56:479–493.

Besag, J. (1986). Statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B*, 48:259–302.

Blake, A. and Zisserman, A. (1987). *Visual Reconstruction*. MIT Press, Mass.

Bonmassar, G. and Schwartz, E. L. (1997). Space-variant Fourier analysis: The exponential chirp transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:1080–1089.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data*. Cambridge University Press, Cambridge.

Burr, D. J. (1981). A dynamic model for image registration. *Computer Graphics and Image Processing*, 15:102–112.

Cao, J. and Worsley, K. J. (1999). The detection of local shape changes via the geometry of Hotelling's T^2 fields. *Annals of Statistics*, 27:925–942.

Carstensen, J. M. (1996). An active lattice model in a Bayesian framework. *Computer Vision and Image Understanding*, 63:380–387.

Caves, R. G., Harley, P. J., and Quegan, S. (1992). Matching map features to synthetic aperture radar (SAR) images using template matching. *IEEE Transactions on Geoscience and Remote Sensing*, 30:680–685.

Christensen, G. E., Rabbitt, R. D., and Miller, M. I. (1996). Deformable templates using large deformation kinetics. *IEEE Transactions on Image Processing*, 5:1435–1447.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (1998). Active appearance models. In Burkhardt, H. and Neumann, B., editors, *Proceedings of the European Conference on Computer Vision*, volume 2, pages 484 – 498, Berlin. Springer-Verlag.

Cote, S. and Tatnall, A. R. L. (1997). The Hopfield neural network as a tool for feature tracking and recognition from satellite sensor images. *International Journal of Remote Sensing*, 18:871–885.

Cross, A. D. J. and Hancock, E. R. (1998). Graph matching with a dual-step EM algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1236–1253.

Dobson, M. C., Pierce, L. E., and Ulaby, F. T. (1996). Knowledge-based land-cover classification using ERS-1/JERS-1 SAR composites. *IEEE Transactions on Geoscience and Remote Sensing*, 34:83–99.

Downie, T. R., Shepstone, L., and Silverman, B. W. (1996). A wavelet based approach to deformable templates. In Mardia, K. V., Gill, C. A., and Dryden, I. L., editors, *Proceedings of the Leeds Annual Statistics Research Workshop*, pages 163–169. Leeds University Press.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester.

Duta, N., Jain, A. K., and Dubuisson-Jolly, M.-P. (1999). Learning 2D shape models. In *IEEE Computer Science Conference on Computer Vision and Pattern Recognition*, volume 2, pages 8–14.

Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response. *Biometrics*, 48:665–677.

Fluete, M. and Lavallée, S. (1998). Building a complete surface model from sparse data using statistical shape models: applications to computer assisted knee surgery. *Medical Image Computing and Computer-Assisted Intervention*, 22:878–887.

Galbraith, W. and Farkas, D. L. (1993). Remapping disparate images for coincidence. *Journal of Microscopy*, 172:163–176.

Galton, F. (1878). Composite portraits. *Journal of the Anthropological Institute of Great Britain and Ireland*, 8:132–142.

Gee, J. C. (1999). On matching brain volumes. *Pattern Recognition*, 32:99–111.

Glasbey, C. A. (1997). SAR image registration and segmentation using an estimated DEM. In Pelillo, M. and Hancock, E. R., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, number 1223 in Lecture Notes in Computer Science, pages 507–520. Springer, Berlin.

Glasbey, C. A. and Horgan, G. W. (1995). *Image Analysis for the Biological Sciences*. Wiley, Chichester.

Glasbey, C. A., Horgan, G. W., Gibson, G. J., and Hitchcock, D. (1995). Fish shape analysis using landmarks. *Biometrical Journal*, 37:481–495.

Glasbey, C. A. and Mardia, K. V. (1998). A review of image warping methods. *Journal of Applied Statistics*, 25:155–171.

Glasbey, C. A. and Martin, N. J. (1996). Multimodality microscopy by digital image processing. *Journal of Microscopy*, 181:225–237.

Goshtasby, A. A. and Le Moigne, J. (1999). Image Registration: Special Issue. *Pattern Recognition*, 32:1–149.

Gould, A. L. (1969). A regression model for angular variates. *Biometrics*, 25:683–670.

Green, P. J. (1999). Penalized likelihood. In *Encyclopedia of Statistical Sciences*, volume update 3, pages 578–586. Wiley.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London.

Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 46:549–603.

Grenander, U. and Miller, M. I. (1998). Computational anatomy: an emerging discipline. *Quarterly of Applied Mathematics*, 56:617–694.

Hallinan, P. W., Gordon, G. G., Yuille, A. L., Giblin, P., and Mumford, D. (1999). *Two- and Three-Dimensional Patterns of the Face*. A. K. Peters, Mass.

Hamon, B. V. and Hannan, E. J. (1974). Spectral estimation of time delay for dispersive and nondispersive systems. *Applied Statistics*, 23:134–142.

Hannan, E. J. and Thomson, P. J. (1988). Time delay estimation. *Journal of Time Series Analysis*, 9:21–33.

Hill, A., Taylor, C. J., and Brett, A. D. (2000). A framework for automatic landmark identification using a new method of nonrigid correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:241–251.

Joshi, S. C. and Miller, M. I. (2000). Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:1357–1370.

Kaijser, T. (1998). Computing the Kantorovich distance for images. *Journal of Mathematical Imaging and Vision*, 9:173–191.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331.

Kent, J. T. and Mardia, K. V. (1994). The link between kriging and thin plate splines. In Kelly, F. P., editor, *Probability, Statistics and Optimization: a tribute to Peter Whittle*, pages 325–339. Wiley, New York.

Kher, A. and Mitra, S. (1993). Registration of noisy SAR imagery using morphological feature extractor and 2-D cepstrum. *Applications of Digital Image Processing*, XV:281–291.

Koch, I. and Snowdon, S. (1994). Image registration by smoothed Fourier methods: An application to medical X-ray images. Technical Report 94-4, Department of Statistics, University of Newcastle, NSW 2308, Australia.

Kuglin, C. D. and Hines, D. C. (1975). The phase correlation image alignment method. In *Proceedings of the IEEE 1975 International Conference on Cybernetics and Society*, pages 163–165.

Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., Vandermalsburg, C., Wurtz, R. P., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42:300–311.

Lanitis, A., Taylor, C. J., and Cootes, T. F. (1995). Automatic face identification system using flexible appearance models. *Image and Vision Computing*, 13:393–401.

Li, H., Manjunath, B. S., and Mitra, S. K. (1995). A contour-based approach to multisensor image registration. *IEEE Transactions on Image Processing*, 4:320–334.

Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern Recognition*, 31:983–1001.

Maintz, J. B. A. and Viergever, M. A. (1998). A survey of medical image registration. *Medical Image Analysis*, 2:1–36.

Mardia, K. V., editor (1994). *Statistics and Images*, volume 2. Carfax Publications, Abingdon.

Mardia, K. V. and Hainsworth, T. J. (1993). Image warping and Bayesian reconstruction with grey-level templates. *Journal of Applied Statistics*, 20:257–280. special issue.

Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. Wiley, Chichester.

Mardia, K. V., McCulloch, C., Dryden, I. L., and Johnson, V. (1997). Automatic scale-space method of landmark detection. In Mardia, K. V., Gill, C. A., and Aykroyd, R. G., editors, *Proceedings of the Leeds Annual Statistics Research Workshop*. Leeds University Press.

McInerney, T. and Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1:91–108.

Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., and Wahl, R. L. (1996). Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1:195–206.

Modrusan, Z., Reiser, R., Feldmann, R. A., Fischer, R. L., and Haughn, G. W. (1994). Homeotic transformation of ovules into carpel-like structures in *Arabidopsis*. *Plant Cell*, 6:333 –349.

Moghaddam, B., Nastar, C., and Pentland, A. (1996). A Bayesian similarity measure for direct image matching. In *International Conference on Pattern Recognition*, Vienna, Austria.

Mokhtarian, F. (1995). Silhouette-based isolated object recognition through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:539–544.

Moshfeghi, M. (1991). Elastic matching of multimodality medical images. *Computer Vision, Graphics and Image Processing*, 53:271–282.

Press, W. H., editor (1994). *Numerical Recipes in Fortran: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2nd edition.

Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B*, 60:351–363.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.

Rangarajan, A., Chui, H., and Duncan, J. S. (1999). Rigid point feature registration using mutual information. *Medical Image Analysis*, 3:425–440.

Rao, M. M. (2000). *Stochastic Processes: Inference Theory*. Kluwer Academic Publishers, Dordrecht.

Ried, T., Baldini, A., Rand, T. C., and Ward, D. C. (1992). Simultaneous visualisation of seven different DNA probes by *in situ* hybridisation using combinatorial fluorescence and digital imaging microscopy. *Proceedings of the National Academy of Sciences, USA*, 89:1388–1392.

Rosenfeld, A. and Kak, A. C. (1982). *Digital Picture Processing*. Academic Press, San Diego, 2nd edition.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: Applications to breast MR images. *IEEE Transactions on Medical Imaging*, 18:712–721.

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Annals of Statistics*, 10:795–810.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Singh, A., Goldgof, D., and Terzopoulos, D., editors (1998). *Deformable Models in Medical Image Analysis.* IEEE Computing Society, Los Alamitos, California.

Strachan, N. J., Nesvadba, P. P., and Allen, A. R. (1990). Fish species recognition by shape analysis of images. *Pattern Recognition*, 23:539–544.

Stuart, A., Ord, J. K., and Arnold, S. (1999). *Kendall's Advanced Theory of Statistics, Volume 2A: Classical Inference and the Linear Model.* Arnold, London.

Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86.

Tabernero, A., Portilla, J., and Navarro, R. (1999). Duality of log-polar image representations in the space and spatial-frequency domains. *IEEE Transactions on Signal Processing*, 47:2469–2479.

Thompson, A. M., Brown, J. C., Kay, J. W., and Titterington, D. M. (1991). A study of methods of choosing the smoothing parameter in image restoration by regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:326–339.

Thompson, D. A. W. (1917). *On Growth and Form.* Cambridge University Press, Cambridge.

Toga, A. W., editor (1999). *Brain Warping.* Academic Press, San Diego.

Viola, P. and Wells III, W. M. (1997). Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24:137–154.

Vornberger, P. L. and Bindschadler, R. A. (1992). Multispectral analysis of ice sheets using co-registered SAR and TM imagery. *International Journal of Remote Sensing*, 13:637–645.

Wahba, G. (1990). *Spline Models for Observational Data.* Society for Industrial and Applied Mathematics, Philadelphia.

Wong, R. Y. and Hall, E. L. (1978). Scene matching with invariant moments. *Computer Graphics and Image Processing*, 8:16–24.

Yang, Z. W. and Cohen, F. S. (1999). Cross-weighted moments and affine invariants for image registration and matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:804–814.

Younes, L. (1999). Optimal matching between shapes via elastic deformations. *Image and Vision Computing*, 17:381–389.