Journal of Applied Statistics, Vol. 25, No. 2, 1998, 155–171

CARFAX

A review of image-warping methods

C. A. GLASBEY[1] **& K. V. MARDIA**[2], [1]*Biomathematics and Statistics Scotland, King's Buildings, Edinburgh, UK and* [2]*Department of Statistics, University of Leeds, UK*

SUMMARY *Image warping is a transformation which maps all positions in one image plane to positions in a second plane. It arises in many image analysis problems, whether in order to remove optical distortions introduced by a camera or a particular viewing perspective, to register an image with a map or template, or to align two or more images. The choice of warp is a compromise between a smooth distortion and one which achieves a good match. Smoothness can be ensured by assuming a parametric form for the warp or by constraining it using differential equations. Matching can be specified by points to be brought into alignment, by local measures of correlation between images, or by the coincidence of edges. Parametric and non-parametric approaches to warping, and matching criteria, are reviewed.*

1 Introduction

Warping of images is an important stage in many applications of image analysis. It may be needed to remove optical distortions introduced by a camera or viewing perspective (Heikkila & Silven, 1997; Tang & Suen, 1993), to register an image with a reference grid such as a map, or to align two or more images (Brown, 1992). For example, matching is important in reconstructing three-dimensional shape from either a series of two-dimensional sections or stereoscopic pairs of images. Much effort has been expended in developing algorithms for registering satellite images with both geographic information systems and with other forms of remote sensing system, such as optical sensors and synthetic aperture radar (see, for example, Richards, 1986, Chapter 2). Recently, there has been considerable interest in registering images produced by medical sensing systems with body atlas information (Colchester & Hawkes, 1991, Section 3). Combined images produced using different imaging modalities also have great potential. For example, X-ray

Correspondence: K. V. Mardia, Department of Statistics, University of Leeds, Leeds LS2 9JT, UK. Tel: 0113 233 5101.

0266-4763/98/020155-17 $7.00

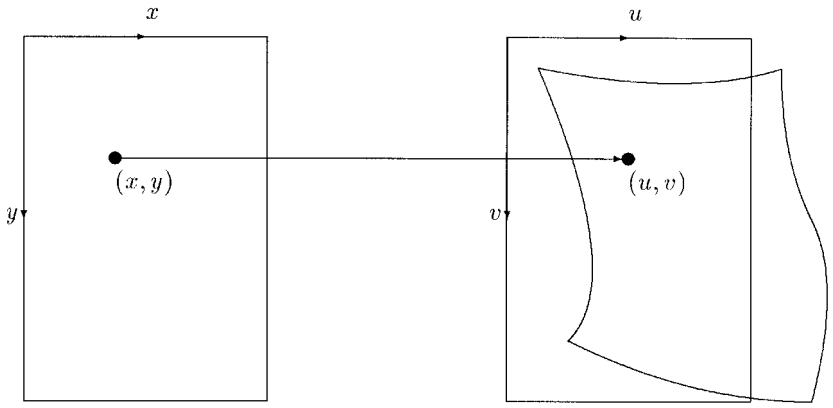

FIG. 1. Notation for image warping.

images reveal structure, whereas magnetic resonance images reveal functionality, so their synthesis generates more informative images (for example, see Hurn *et al.*, 1996; Mardia & Little, 1994).

A related problem is the warping of one-dimensional signals to bring them into alignment. This is sometimes referred to as dynamic time warping. Dynamic programming methods have been applied to speech processing (Sakoe & Chiba, 1978), handwriting analysis (Burr, 1983), alignment of boundaries of ice floes (McConnell *et al.*, 1991) and of tracks in electrophoresis gels (Skovgaard *et al.*, 1995). Wang and Gasser (1997) considered theoretical issues. Where features on two curves are already matched, the problem simplifies to one of monotone regression (see, for example, Ramsay, 1988).

A warping is a pair of two-dimensional functions, $u(x,y)$ and $v(x,y)$, which map a position (x,y) in one image, where x denotes column number and y denotes row number, to position (u,v) in another image (see Fig. 1). There have been many approaches to finding an appropriate warp, but a common theme is the compromise between insisting the distortion is smooth and achieving a good match. In some recently published cases the warp seems unnecessarily rough (Conradsen & Pedersen, 1992, Fig. 8(b); Grenander & Miller, 1994, Fig. 7(f)). Smoothness can be ensured by assuming a parametric form for the warp, such as the affine transformation, or by penalizing roughness, such as by using thin-plate splines. Depending on the application, matching might be specified by points which must be brought into alignment, by local measures of correlation between images, or by the coincidence of edges. Overviews of geometric transformations are given by Wolberg (1988), Bookstein (1991), Brown (1992) and Tang and Suen (1993).

In Section 2, we will progress through parametric models, from linear to non-linear ones, elucidating their properties. Then, in Section 3, we will consider a range of non-parametric models. Matching criteria will be covered as they arise in different applications.

2 Parametric transformations

Figure 2 shows a hierarchy of parametric transformations. In many applications, it is important to use a transformation which is no more general than it need be. We will consider the properties of each transformation in turn.

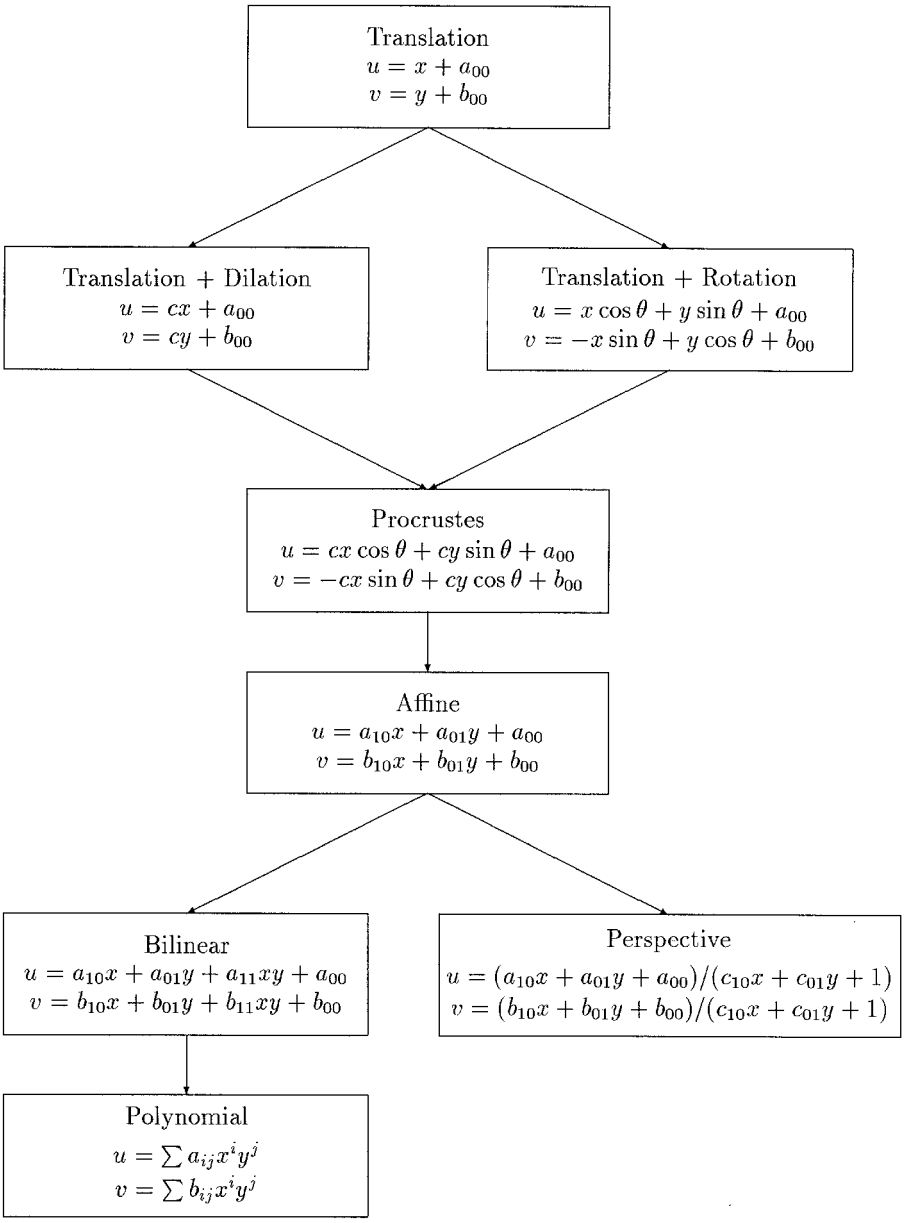

FIG. 2. A hierarchy of transformations. Arrows denote models which generalize others.

2.1 Translation

If the only difference between two images is one of translation, either along rows, along columns or a combination of both, then

$$u = x + b_{00} \quad \text{and} \quad v = y + b_{00}$$

performs the required mapping. Here, the top-left corner $(0, 0)$ of one image

matches location (a_{00}, b_{00}) in the other image. Agreement between images may be measured by

$$S = \sum_{x=0}^{n_x} \sum_{y=0}^{n_y} (I_{x,y} - I'_{u,v})^2$$

where $I_{x,y}$ denotes pixel value (x, y) in one image and $I'_{u,v}$ denotes pixel value (u, v) in the other image. In general, u and v will not be integers, and $I'_{u,v}$ is assigned the value of $I'_{[u+1/2],[v+1/2]}$, where [.] is used to denote the integer part of a number. Alternatively, bilinear interpolation could be used, with

$$I'_{u,v} = ([u]+1-u)([v]+1-v)I'_{[u],[v]} + (u-[u])([v]+1-v)I'_{[u+1],[v]}$$
$$+ ([u]+1-u)(v-[v])I'_{[u],[v+1]} + (u-[u])(v-[v])I'_{[u+1],[v+1]}$$

There are potential problems with S: if the dimensions of the second image (n_u, n_v) are such that $n_u \leqslant n_x$ and $n_v \leqslant n_y$, then some values of (u, v) will lie outside this range if $(a_{00}, b_{00}) \neq (0, 0)$. The simplest way around this difficulty is to restrict the summations to the area of overlap of the two images after transformation. But then it is also a good idea to standardize S by dividing by the area of overlap, because otherwise the measure takes a minimum value of zero when the two images do not overlap at all. Note that, although the transformation is functionally invertible, because (x, y) can be expressed as a translation of (u, v), in general a different result would be obtained if

$$S' = \sum_{u=0}^{n_u} \sum_{v=0}^{n_v} (I_{x,y} - I'_{u,v})^2$$

were minimized instead of S. One way to retain equivalence in the two images would be to minimize $S + S'$ instead.

There are many other measures of agreement (see, for example, Rosenfeld & Kak, 1982, Section 9.4), such as the covariance, which is proportional to

$$\sum_{x=0}^{n_x} \sum_{y=0}^{n_y} (I_{x,y} - \bar{I})(I'_{u,v} - \bar{I}')$$

where $\bar{I}$ denotes the average pixel value in the first image. This is a linear convolution, which can be evaluated very efficiently for all possible changes in origin simultaneously using a fast Fourier transform (Glasbey & Horgan, 1995, Section 3.2). The covariances are given by

$$\mathcal{F}^{-1}\{\mathcal{F}(I - \bar{I})\mathcal{F}^{C}(I' - \bar{I}')\}$$

where $\mathcal{F}$ denotes the Fourier transform, $\mathcal{F}^{C}$ denotes the complex conjugate of the transform and $\mathcal{F}^{-1}$ is the inverse transform. S can be computed in an analogous way. When shifts of location are all that are involved, this problem is sometimes referred to as template matching. Attention has been given to efficient computations; for example, by searching first for optimal values of a_{00} and b_{00} using a coarse grid, and then on a finer grid (Goshtasby *et al.*, 1984).

Phase correlation, defined as

$$\mathcal{F}^{-1}\left\{\frac{\mathcal{F}(I)\mathcal{F}^{C}(I')}{|\mathcal{F}(I)\mathcal{F}^{C}(I')|}\right\}$$

where $| |$ denotes modulus, was proposed by Kuglin and Hines (1975). It performs better than the covariance in cases where the differences between I and I' (after the appropriate shift in location) occur only at a subset of frequencies. This would be the case, for example, if the trend in illumination differs between the images, but not if significant levels of white noise are present. It is also possible to formulate criteria, intermediate between correlation and phase correlation, which are optimal when images agree only at a subset of frequencies.

2.2 Procrustes transformation

If there is some change in magnification between the images, and a rotation of θ degrees, then

$$u = cx\cos\theta + cy\sin\theta + a_{00} \quad \text{and} \quad v = -cx\sin\theta + cy\cos\theta + b_{00}$$

where the constant c performs a rescaling. A value of $c = 1$ corresponds to no change in magnification, whereas $c > 1$ is an enlargement and $c < 1$ is a shrinkage. In computer vision, location, scale and rotation are known collectively as pose. Once these differences have been removed, what remains is known as shape. This is the transformation used in Procrustes analysis (Goodall, 1991) to align labelled sets of points. It is a four-parameter transformation and can be uniquely defined from two points in the two images, or a least-squares solution can be obtained if there are more than two points. Extensions to more than two images are relatively straightforward. The simplest approach is to align all images with the first one. However, it is usually somewhat arbitrary to single out one image in this way. In generalized Procrustes analysis, all sets of points are matched to the average configuration. In other applications, such an approach could lead to much heavier computations.

It is possible to restrict the transformation to fewer parameters in specific situations, by omitting either the scaling or the rotation terms (see Fig. 2). For example, Glasbey and Martin (1996) used translation and scaling to align microscope images of algal cells obtained using different modalities, namely brightfield, phase contrast and differential interference contrast (DIC). This warping was both necessary and sufficient to compensate for changes in image alignment resulting from imperfect centration of the different lens systems and from slight differences in magnification between objective lenses with the same nominal magnification. Matching criteria such as correlation, which is based simply on image intensities, did not perform well because, in brightfield microscopy, algal cells appear dark, whereas in DIC one side of cells is dark and the other side is light. Positions where intensities change rapidly at the edges of cells do coincide, and were therefore used in the matching criterion. Edge information has also been used by Bajcsy and Kovacic (1989) and Moshfeghi (1991) to align medical images. Prewitt's edge filter (see, for example, Glasbey & Horgan, 1995, Section 3.4) was applied to all the microscope images. Then the cross-correlation between gradient images was computed using Fourier methods for each of a range of differences in magnifications between the images. Translation and magnification parameters which produced the highest cross-correlation were selected.

2.3 Affine transformation

The affine transformation is a six-parameter generalization of the Procrustes transformation. It permits different stretching along rows and columns of an image,

and shearing, and is the most general linear transformation:

$$u = a_{10}x + a_{01}y + a_{00} \quad \text{and} \quad v = b_{10}x + b_{01}y + b_{00}$$

Bookstein (1978, Chapter 6) gave a characterization of the affine transform: an orthogonal pair of directions in the x–y image remains orthogonal in the u–v image, and the transformation either stretches or shrinks in these two directions. Horgan *et al.* (1992) used this mapping to superimpose pairs of SDS-PAGE gel electrophoretograms, because differential stretching between rows and columns is possible in gels and some rotation may occur in digitizing the gels. Invariant spots were identified which were common to the two images. The centres of the spots were denoted by $(x_1, y_1), \ldots, (x_m, y_m)$ in the first gel, and by $(u_1, v_1), \ldots, (u_m, v_m)$ in the second gel. A regression algorithm was used to estimate the linear parameters which minimized the sum of squared differences between the spots:

$$\sum_{i=1}^{m} \{u_i - u(x_i, y_i)\}^2 + \sum_{i=1}^{m} \{v_i - v(x_i, y_i)\}^2$$

and therefore brought them approximately into alignment. Note that the problem is considerably harder if the points are unlabelled and need to be assigned (see, for example, Moss & Hancock, 1996; Stoddart *et al.*, 1996).

2.4 *Perspective transformation*

The perspective transformation arises if a planar object is viewed from a fixed point in space:

$$u = \frac{a_{10}x + a_{01}y + a_{00}}{c_{10}x + c_{01}y + 1} \quad \text{and} \quad v = \frac{b_{10}x + b_{01}y + b_{00}}{c_{10}x + c_{01}y + 1}$$

It is a non-linear transformation requiring eight parameters, and has the affine transformation as a limiting case as the viewing point becomes more distant and foreshortening effects diminish. Glasbey (1997) used the affine approximation as a first step in registering an airborne synthetic aperture radar image with a digital map. The perspective transformation is the most general transformation which maps straight lines at all orientations to straight lines, as do all the previously considered transformations. And it preserves conic sections, that is circles, ellipses, parabolas and hyperpolas, which have the general functional form

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0$$

Also in common with the earlier transformations, it is functionally invertible: the inverse transformation $(u, v) \to (x, y)$ has the same functional form. Therefore, the transformation is guaranteed to be bijective, i.e. it is impossible for folding to occur where two points in the x–y image map to the same point in the u–v image. Also, if the measurement of agreement between two images treats them equivalently, it is arbitrary which image is chosen to be mapped on to the other one. None of the transformations which follow has this property, except for the linear spline. Parameters can be chosen to align four landmarks, by solving the four pairs of linear equations of the following form:

$$c_{10}xu + c_{01}yu + u = a_{10}x + a_{01}y + a_{00}$$

$$c_{10}xv + c_{01}yv + v = b_{10}x + b_{01}y + b_{00}$$

However, to fit to more than four landmarks by least-squares in the image space requires an iterative approach.

2.5 Bilinear transformation

The bilinear transformation is another eight-parameter transformation generalization of the affine transformation (see Fig. 2), but with different properties:

$$u = a_{10}x + a_{01}y + a_{11}xy + a_{00} \quad \text{and} \quad v = b_{10}x + b_{01}y + b_{11}xy + b_{00}$$

Bookstein (1991, p. 253) gave illustrations of the difference between it and the perspective transformation. Straight lines in three particular directions are preserved, including lines parallel to either x- or y-axes. Therefore, the transformation is not rotationally invariant: it has the undesirable feature that, if both images are rotated, then the warping transformation between them would be different. Also, this transformation, and most of those to follow, are not guaranteed to be bijective, i.e. one-to-one. For continuously differentiable transformations, bijectivity is equivalent to the Jacobian

$$\left| \frac{\partial u}{\partial x}\frac{\partial v}{\partial y} - \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} \right|$$

being non-zero over the whole image. (Note that all the previously considered transformations have a constant Jacobian.) Fitzpatrick and Louze (1987) gave sufficient conditions for this to hold in the case of the bilinear transformation. As with the perspective transformation, four landmarks (except where collinearities occur) uniquely specify the bilinear transformation. In particular, if the four corners of the first image, that is $(0, 0)$, $(n_x, 0)$, $(0, n_y)$, and (n_x, n_y), map to (u_{00}, v_{00}), (u_{10}, v_{10}), (u_{01}, v_{01}) and (u_{11}, v_{11}) respectively, then

$$u = u_{00} + (u_{10} - u_{00})\frac{x}{n_x} + (u_{01} - u_{00})\frac{y}{n_y} + (u_{11} - u_{10} - u_{01} + u_{00})\frac{xy}{n_x n_y}$$

$$v = v_{00} + (v_{10} - v_{00})\frac{x}{n_x} + (v_{01} - v_{00})\frac{y}{n_y} + (v_{11} - v_{10} - v_{01} + v_{00})\frac{xy}{n_x n_y}$$

2.6 Polynomial and other transformations

Polynomial transformations of order p are specified by either

$$u = \sum_{i=0}^{p} \sum_{j=0}^{p-i} a_{ij} x^i y^j \quad \text{and} \quad v = \sum_{i=0}^{p} \sum_{j=0}^{p-i} b_{ij} x^i y^j$$

or

$$u = \sum_{i=0}^{p} \sum_{j=0}^{p} a_{ij} x^i y^j \quad \text{and} \quad v = \sum_{i=0}^{p} \sum_{j=0}^{p} b_{ij} x^i y^j$$

depending on the maximum order of interaction term included. These transformations include quadratic, biquadratic, cubic and bicubic ones as special cases (Tang & Suen, 1993). Bookstein (1991, Section 7.4) studied the properties of the quadratic transformation. In registration of remotely sensed images, polynomials of third and higher order are used (Richards, 1986). Bernstein (1976) proposed

evaluation of the transformation on a coarse grid only, followed by bilinear interpolation, to reduce the computations involved.

There are many alternative parametric transformations to polynomials. Glasbey *et al.* (1995) mapped arcs of concentric circles to straight lines in order to remove the flexion from images of fish. Bookstein (1991, Section 7.4.4) reviewed transformations appropriate for matching landmarks for spiral structures and other patterns of growth. Amit *et al.* (1991) used a transformation which maps the unit square on to itself:

$$u = x + \sum_{i=0}^{m}\sum_{j=0}^{m} a_{ij} \sin(\pi i x)\cos(\pi j y) \quad \text{and} \quad v = y + \sum_{i=0}^{m}\sum_{j=0}^{m} b_{ij} \cos(\pi i x)\sin(\pi j y)$$

obtained as the eigenvectors of a differential operator appropriate to a diffusion process, to match X-ray images of hands. Jain *et al.* (1996) used a similar transformation to align hand-drawn templates with objects in images. Glasbey and Wright (1994) removed warping distortions from multi-track electrophoretic gels by estimating the orientation direction of bands in different parts of a gel as a polynomial $f(x, y)$ in x and y, then integrating to obtain

$$g(u, v) = v + \int_0^u \frac{1}{\tan f(x, g(x, v))} \mathrm{d}x$$

and transforming $(x, g(x, v)) \to (x, v)$ in order to bring the bands into alignment with the x-axis. This transformation is guaranteed to be bijective.

Displays other than of the warped image can be used to see more clearly the effect of a transformation. Alternative displays include the result of applying the transformation to a regular lattice pattern (Grenander & Miller, 1994, Fig. 7(f)), a vector field of changed positions (Bookstein, 1991, Fig. 7.4.4), a bi-orthogonal grid of locally perpendicular directions which remain perpendicular after transformation (Bookstein, 1991, Fig. 6.6.1) and stretch marks which show values of (u, v) to which no points (x, y) map (Glasbey *et al.*, 1995, Fig. 5; Mardia & Hainsworth, 1993). To produce a display of the transformed image, it is usual to fill in these stretch marks by interpolating from adjacent values. Alternatively, the mapping can be inverted, by mapping $\{u(x, y), v(x, y)\}$ to $\{x, y\}$ for all integer values of x and y, in which case there will be no gaps.

3 Non-parametric transformations

Parametric transformations do not perform well in the presence of local distortions (Goshtasby, 1986). Piecewise affine transformations offer an alternative to polynomials in generalizing affine transformations. Given a set of matched landmarks in the two image, their Delaunay triangulation can be obtained, and an affine transformation defined by its vertices can be used inside each triangle (Goshtasby, 1986). Continuity of the transformation along the edges of triangles is thereby ensured. This is a first-order or linear spline. A cubic spline could be used to produce a smoother transformation (Goshtasby, 1987), although Maeland (1988) shows the *sinc* function to be a superior interpolant. If the landmarks in one image are the lattice points of a rectangular grid, then another option is to use a bilinear transformation within each rectangle. A related method was used by Conradsen and Pedersen (1992). They applied linear transformations at a progressively finer

series of scales, to maximize correlations between electrophoresis images. In none of these cases have we considered any smoothness constraints, and transformations can therefore be very rough. The matching of stereo pairs of images is a specialized area, because distortions between images need not be smooth nor continuous, even (see Weng *et al.*, 1993, for an optimal algorithm). Similarly, in situations where multiple objects in one image have each separately moved position in a second image, then a smooth warping between images is inappropriate. X-ray images of human chests is one such example, where the rib cage and internal organs can move relative to one another. Modelling of individual organs is then more appropriate. We will not consider such applications further in this paper.

In the following subsections, we will consider formulations of roughness based on elastic deformations, pairs of thin-plate splines and Bayesian priors. These lead to a range of non-parametric transformations.

3.1 Elastic deformations

One way to introduce smoothness constraints is to equate warping with the distortions of an elastic sheet or membrane. The elastic energy of a deformation is given by

$$\iint \frac{1}{2}(w_{xx}\sigma_{xx} + w_{yy}\sigma_{yy} + 2w_{xy}\sigma_{xy})\,\mathrm{d}x\,\mathrm{d}y$$

where (w_{xx}, w_{yy}, w_{xy}) is the strain tensor specified by

$$w_{xx} = \frac{\partial u}{\partial x}, \qquad w_{yy} = \frac{\partial v}{\partial y}, \qquad w_{xy} = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)$$

and $(\sigma_{xx}, \sigma_{yy}, \sigma_{xy})$ is the stress tensor

$$\sigma_{xx} = \frac{E}{1-\sigma^2}\left(\frac{\partial u}{\partial x} + \sigma\frac{\partial v}{\partial y}\right)$$

$$\sigma_{yy} = \frac{E}{1-\sigma^2}\left(\frac{\partial v}{\partial y} + \sigma\frac{\partial u}{\partial x}\right)$$

$$\sigma_{xy} = \frac{E}{2(1+\sigma)}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)$$

Here, E is Young's modulus and σ is Poisson's ratio (Landau & Lifshitz, 1986). In particular, if $\sigma = 0$, then the energy is proportional to

$$\iint \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 \mathrm{d}x\,\mathrm{d}y$$

whereas, as $\sigma \to -1$, the energy approaches

$$\iint \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2 \mathrm{d}x\,\mathrm{d}y$$

Many variants of such first-order differential equations have been proposed. For example, Burr (1981) used sums of Gaussian weight functions to interpolate between matched points, justifying them as an elastic Green's function in an appropriate medium. Tang and Suen (1993) found harmonic transformations which minimized

$$\int_\Omega\int\left\{\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2+\left(\frac{\partial v}{\partial y}\right)^2\right\}\mathrm{d}x\,\mathrm{d}y$$

subject to matching along a closed contour enclosing Ω. Amit (1994) proposed a quadratic form generated by a linear differential operator, and used wavelets to find a solution. Glasbey (1998) used variances of squared first-order derivatives as a roughness penalty when comparing the shapes of fish, the criterion having been chosen so that it was minimized uniquely by the shape-preserving Procrustes transformation. The approach can also be extended to three dimensions, for example to align three-dimensional medical images. Bajcsy and Kovacic (1989) registered three-dimensional X-ray computed tomography data with a brain atlas.

Barron *et al.* (1994) considered both first- and second-order derivatives for optical flow, to align a series of images in which there was relative motion among the several objects being imaged. Here, the physical analogy is with fluid flow. If deformations between successive images are small, then flow orthogonal to an edge can be estimated by combining the output from an edge filter with the result of subtracting one image from another. Smoothness constraints are introduced to regularize the problem, to estimate flow along edges. There are similarities with the use of snakes, linear templates which are distorted smoothly to align with image features (Kass *et al.*, 1988). First-order derivatives are regarded as 'tension' constraints and second-order derivatives as 'rigidity' constraints.

3.2 Thin-plate splines

Another physical analogue for warping is to regard it as a pair of two-dimensional surfaces, representing u as a function of x and y, and similarly v as a function of x and y. Mardia and Little (1994) and Mardia *et al.* (1996) have proposed kriging predictors with derivative information for u and v following the thin-plate spline deformation of Bookstein (1989) and Bookstein and Green (1993). Independently, Arad *et al.* (1994) have dealt with particular cases using radial basis function terminology, namelt the two surfaces are:

$$u=\sum_{i=1}^{m}c_i f\left(\sqrt{(x-x_i)^2+(y-y_i)^2}\right)+a_{10}x+a_{01}y+a_{00}$$

and similarly for v, where $(x_1,y_1),\ldots,(x_m,y_m)$ are a set of landmarks, and f is a function such as a multi-quadric

$$f(t)=(t^2+t_0^2)^\alpha,\qquad 0<\alpha<1$$

a shifted log

$$f(t)=\log\left(\sqrt{t^2+t_0^2}\right),\qquad t_0^2>1$$

a Gaussian density

$$f(t) = \exp[-t^2/(2\sigma^2)]$$

or a thin-plate spline

$$f(t) = t^2 \log t^2$$

Arad *et al.* (1994) found that the thin-plate spline could lead to problems if landmarks were sparse: small changes in landmark positions could produce global changes in the warping transformation.

Many choices of f correspond to the minimization of some functional. In particular, the thin-plate spline minimizes

$$\mathcal{J}(u) = \iint_{-\infty}^{\infty} \left\{ \left(\frac{\partial^2 u}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 u}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 u}{\partial y^2}\right)^2 \right\} \mathrm{d}x\, \mathrm{d}y$$

and similarly for v, subject to the transformations matching the m landmarks in the first image to $(u_1, v_1), \ldots, (u_m, v_m)$, respectively, in the second image (Bookstein, 1989; Green & Silverman, 1994, Section 7). The solution also satisfies the biharmonic equations

$$\frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = \frac{\partial^4 v}{\partial x^4} + 2\frac{\partial^4 v}{\partial x^2 \partial y^2} + \frac{\partial^4 v}{\partial y^4} = 0$$

except at the landmarks. (For solutions when other orders of derivative are used, see Wahba, 1990.) The transformation has been used by Bookstein (1991) to align medical scan images, and by Horgan *et al.* (1992) as a generalization of the affine transformation to align electrophoresis gels.

If landmarks are observed subject to noise, then it is possible to use thin-plate splines for smoothing rather than for interpolation, by seeking functions u and v which minimize the functional

$$\sum_{i=1}^{m} \{u_i - u(x_i, y_i)\}^2 + \sum_{i=1}^{m} \{v_i - v(x_i, y_i)\}^2 + \lambda\{\mathcal{J}(u) + \mathcal{J}(v)\}$$

for some non-negative choice of λ. The coefficients $c_1, \ldots, c_m, a_{10}, a_{01}, a_{00}$, and similarly those for v, can be obtained as the solution of $(m+3)$ simultaneous linear equations, resulting in a thin-plate smoothing spline. Larger values of λ produce smoother results, but with poorer alignment of the labelled points.

Amodei and Benbourhim (1991) generalized the method to give different weights to compressional and rotational transformations between images. Kent and Mardia (1994) showed that the thin-plate spline is equivalent to kriging with a specific spatial covariance structure. Evaluation of thin-plate spline transformations over a whole image is computationally expensive. Several ideas have been pursued to speed up these calculations, including the use of an adaptive grid with affine or bilinear interpolation within it (Flusser, 1992), and Fourier transforms (Berman, 1994). A closely related roughness penalty is the Laplacian (O'Sullivan, 1991). Sampson and Guttorp (1992) used pairs of thin-plate splines to warp the space of a Gaussian random field to make it stationary and isotropic. This approach is strongly linked with kriging (see Mardia *et al.*, 1991, 1996). In fact, it can be shown that these are particular cases either of stationary processes, or intrinsic

random fields, of appropriate order. Another stationary process of importance has covariance at distance d

$$\frac{1}{2^{\theta_2-1}\Gamma(\theta_2)}\left(\frac{2\sqrt{\theta_2 d}}{\theta_1}\right)^{\theta_2} K_{\theta_2}\left(\frac{2\sqrt{\theta_2 d}}{\theta_1}\right), \qquad \text{for } \theta_1, \theta_2 > 0$$

where $\Gamma(.)$ is the standard gamma function and $K_{\theta_2}(.)$ is the modified Bessel function of the third kind of order θ_2. Arad and Reisfield (1995) and Little *et al.* (1996) showed how to include rigid structure within a non-linear landmark-based deformation. Glasbey (1997) used a single thin-plate spline to represent ground height, in combination with an affine transformation, to register an airborne synthetic aperture radar image with a digital map.

3.3 Bayesian approach

A Bayesian approach offers further possibilities for specifying smoothness constraints. The first image (I) is regarded as a template, to be warped to align with the second image (I'), using a set of transformation parameters $W = \{w_j : j = 0, \ldots, n\}$. W is estimated using the posterior density, which can be expressed in terms of likelihood and prior density:

$$p(W \mid I', I) \propto p(I' \mid W, I)\, p(W)$$

The likelihood of the second image, conditional on the template and deformation, is specified by, for example

$$p(I' \mid W, I) = \frac{1}{(2\pi\sigma^2)^{(n_u+1)(n_v+1)/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{u=0}^{n_u}\sum_{v=0}^{n_v}(I_{x,y} - I'_{u,v})^2\right\}$$

There are many possible prior distributions, each corresponding to different prior beliefs. One possibility is the Gibbs distribution corresponding to the thin-plate spline

$$p(W) = \frac{1}{Z(\lambda)} \exp\{-\lambda(\mathcal{J}(u) + \mathcal{J}(v))\}$$

where λ is a non-negative prior parameter and Z is a normalizing constant. To produce an automatic procedure, we must have knowledge of the value of λ, for example from training data, or include its estimation in a fully Bayesian analysis by further introducing a hyper-prior distribution $p(\lambda)$ for λ. The posterior distribution then becomes

$$p(W, \lambda \mid I', I) \propto p(I' \mid W, I)\, p(W \mid \lambda)\, p(\lambda)$$

In image warping in general, but particularly with Bayesian approaches, parameter estimation is challenging because of a heavy computational load and the presence of local optima. Good starting values are important and have been obtained in many ways, such as the matching of low-order moments (Wong & Hall, 1978), conducting a grid search, and initially using a simpler model. Stochastic optimization has been used to escape from local optima. Multi-resolution strategies can also help, as can smoothing before matching (Kass *et al.*, 1988). A common approach to Bayesian estimation is to use the Metropolis–Hastings algorithm to approximate the posterior distribution. This is a Markov chain Monte Carlo

(MCMC) technique, in which an ergodic Markov chain is constructed which has the required posterior as its limiting distribution. For a discussion, see Hammersley and Hanscomb (1964), Hastings (1970), Geman (1991), Green and Han (1992) and Propp and Wilson (1996).

The Metropolis–Hastings algorithm is used in the following manner. Let the current set of warping parameters be $W = \{w_j : j = 0, \ldots, n\}$. A proposed new value for one of the parameters, w_i say, is drawn from a proposal distribution $q(w_i' | w_i)$. Although the choice of proposal distribution is (almost) arbitrary, a common choice is a normal distribution centred on the current parameter value, with spread parameter chosen to achieve acceptable convergence rates. Let the set of warping parameters including the proposed parameter be W', that is $W' = \{w_1, \ldots, w_i', \ldots, w_n\}$. The proposal is accepted, and the parameter values updated accordingly, with probability

$$\min\left\{1, \frac{p(W' | I', I) q(w_i' | w_i)}{p(W | I', I) q(w_i | w_i')}\right\}$$

Otherwise it is rejected and no change is made. Note that this proposal distribution is symmetric, that is $q(w_i' | w_i) = q(w_i | w_i')$, hence the ratio of these terms cancels in the above expression. Also, some simplification of the posterior ratio can usually be performed, producing a computationally cheap updating step.

Each of the warping parameters is considered in the same way and the whole iterative process repeated until stability is apparent. The most usual and simplest approach to detection of convergence is to monitor the value of one-dimensional functions of the evolving process: once these appear stable the Markov chain is assumed to have converged. Clearly, this approach is subjective, but usually works well. A discussion of formal convergence diagnostics can be found in Cowles and Carlin (1996). Another question is how many sweeps should be performed after the transient period has ended. One approach is to consider integrated autocorrelation of the one-dimensional monitoring statistics using a truncated periodogram estimator (Sokal, 1989). Once the pseudo-sample has been generated from the posterior distribution, a number of possible estimators are available: one choice is the posterior median, which can be estimated by the sample median of the pseudo-sample. Among the benefits of employing sampling techniques is that the pseudo-sample can be used to calculate interval estimates using sample percentiles, or in fact the whole of the posterior distribution can be examined.

Mardia and Hainsworth (1993) used a pair of thin-plate splines in a Bayesian approach to align images when landmarks are available in only one of the images. Carstensen (1996) used a Markov random field lattice model to correct for distortions in a hybridization filter. Lee *et al.* (1997) proposed a novel approach to warping for sequences of tagged magnetic resonance images of the heart, based on tracking quadrilaterals through time. The approach of Christensen *et al.* (1996) allows for large deformations. For example, their method has the capacity for a square to be warped to a letter 'C', and for letters to be regenerated from partial information. Many approaches to warping using thin-plate splines produce smooth deformations. Terzopoulos (1986), however, allows discontinuous deformation, giving, as an example, a broken bone deformed into the original, intact bone. In such a case, smoothness is inappropriate. Bookstein and Green (1993), Mardia *et al.* (1996) and Mardia and Little (1994) describe deformations when higher-derivative information, such as gradients and curvature, are available at landmarks.

The use of least-squares techniques with landmark-based warping can be greatly influenced by outliers: Dryden (1996) considers robust matching. The use of wavelet warping is examined by Aykroyd and Mardia (1996a,b) for curve deformation applied to a study of spinal scoliosis, including longitudinal growth studies. They use an MCMC approach to estimation which allows estimation of clinically important quantities such as curvature. The same approach, but using simulated annealing, was proposed by Downie *et al.* (1996) for deformation of binary bone templates. The use of wavelet expansions to describe the warping function in these approaches permits local deformations. In aligning a template with an image, such as a map of field boundaries with a SAR image, an alternative approach to having a rigid template in a flexible space is to allow the template itself to be flexible. See Grenander *et al.* (1991), Phillips and Smith (1994) and Cootes *et al.* (1995) for examples of this, and McInerney and Terzopoulos (1996) for a review of medical applications.

Acknowledgements

The first author's work was supported by funds from the Scottish Office Agriculture, Environment and Fisheries Department. The authors are grateful to Robert Aykroyd for his comments and to EPSRC for a research grant.

REFERENCES

AMIT, Y. (1994) A nonlinear variational problem for image matching, *SIAM Journal of Sci. Comput*, 15, pp. 207–224.

AMIT, Y., GRENANDER, U. & PICCIONI, M. (1991) Structural image restoration through deformable templates, *Journal of the American Statistical Association*, 86, pp. 376–387.

AMODEI, L. & BENBOURHIM, M. N. (1991) A vector spline approximation, *Journal of Approximation Theory*, 67, pp. 51–79.

ARAD, N. & REISFIELD, D. (1995) Image warping using few anchor points and radial functions, *Computer Graphics Forum*, 14, pp. 35–46.

ARAD, N., DYN, N., REISFELD, D. & YESHURUN, Y. (1994) Image warping by radial basis functions: applications to facial expressions, *CVGIP: Graphical Models and Image Processing*, 56, pp. 161–172.

AYKROYD, R. G. & MARDIA, K. V. (1996a) An MCMC approach to wavelet warping. *In:* K. V. MARDIA, C. A. GILL & I. L. DRYDEN (Eds), *Image Fusion and Shape Variability Techniques* (Leeds, Leeds University Press), pp. 129–140.

AYKROYD, R. G. & MARDIA, K. V. (1996b) Shape analysis of spinal curves by wavelet warping using an MCMC approach, Technical report, *STAT-96/10*, Department of Statistics, University of Leeds.

BAJCSY, R. & KOVACIC, S. (1989) Multiresolution elastic matching, *Computer Vision, Graphics and Image Processing*, 46, pp. 1–21.

BARRON, J. L., FLEET, D. J. & BEAUCHEMIN, S. S. (1994) Performance of optical flow techniques, *International Journal of Computer Vision*, 12, pp. 43–77.

BERMAN, M. (1994) Automated smoothing of image and other regularly spaced data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, pp. 460–468.

BERNSTEIN, R. (1976) Digital image processing of Earth observation sensor data, *IBM Journal of Research and Development*, 20, pp. 40–56.

BOOKSTEIN, F. L. (1978) *The Measurement of Biological Shape and Shape Change* (Berlin, Springer-Verlag).

BOOKSTEIN, F. L. (1989) Principal warps: thin-plate splines and the decomposition of deformations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, pp. 567–585.

BOOKSTEIN, F. L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology* (Cambridge, Cambridge University Press).

BOOKSTEIN, F. L. & GREEN, W. D. K. (1993) A thin-plate spline for deformations with specified derivatives. *In: Mathematical Methods in Medical Imaging.*

BROWN, L. G. (1992) A survey of image registration techniques, *ACM Computing Surveys*, 24, pp. 325–376.

BURR, D. J. (1981) A dynamic model for image registration, *Computer Graphics and Image processing*, 15, pp. 102–112.

BURR, D. J. (1983) Designing a handwriting reader, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5, pp. 554–559.

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

CHRISTENSEN, G. E., RABBITT, R. D. & MILLER, M. I. (1996) Deformable templates using large deformation kinetics, *IEEE Transactions on Image processing*, 5, pp. 1435–1447.

COLCHESTER, A. C. F. & HAWKES, D. J. (Eds) (1991) Information Processing in Medical Imaging. *In: Proceedings of the 12th International Conference on Information Processing in Medical Imaging* (Berlin, Springer-Verlag).

CONRADSEN, K. & PEDERSEN, J. (1992) Analysis of 2-dimensional electrophoretic gels, *Biometrics*, 48, pp. 1273–1287.

COOTES, T. F., TAYLOR, C. J., COOPER, D. H. & GRAHAM, J. (1995) Active shape models—their training and application, *Computer Vision and Image Understanding*, 61, pp. 38–59.

COWLES, M. K. & CARLIN, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association*, 91, pp. 883–904.

DOWNIE, T. R., SHEPSTONE, L. & SILVERMAN, B. W. (1996) A wavelet based approach to deformable templates. *In:* K. V. MARDIA, C. A. GILL & I. L. DRYDEN (Eds), *Image Fusion and Shape Variability Techniques* (Leeds, Leeds University Press), pp. 163–169.

DRYDEN, I. L. (1996) General shape and registration analysis. Technical report, *STAT 96/03*, Department of Statistics, University of Leeds.

FITZPATRICK, J. M. & LOUZE, M. R. (1987) A class of one-to-one two-dimensional transformations, *Computer Vision, Graphics and Image processing*, 39, pp. 369–382.

FLUSSER, J. (1992) An adaptive method for image registration, *Pattern Recognition*, 25, pp. 45–54.

GEMAN, D. (1991) Random fields and inverse problems in imaging, *Lecture Notes in Mathematics*, 1427, pp. 113–193.

GLASBEY, C. A. (1997) SAR image registration and segmentation using an estimated DEM. *In:* M. PELILLO & E. R. HANCOCK (Eds), *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 507–520 (Berlin, Springer-Verlag).

GLASBEY, C. A. (1998) Fish species recognition by image warping, *Bioss Report.*

GLASBEY, C. A. & WRIGHT, F. G. (1994) An algorithm for unwarping multitrack electrophoretic gels, *Electrophoresis*, 15, pp. 143–148.

GLASBEY, C. A. & HORGAN, G. W. (1995) *Image Analysis for the Biological Sciences* (Chichester, Wiley).

GLASBEY, C. A. & MARTIN, N. J. (1996) Multimodality microscopy by digital image processing, *Journal of Microscopy*, 181, pp. 225–237.

GLASBEY, C. A., HORGAN, G. W., GIBSON, G. J. & HITCHCOCK, D. (1995) Fish shape analysis using landmarks, *Biometrical Journal*, 37, pp. 481–495.

GOODALL, C. (1991) Procrustes methods in the statistical analysis of shape (with discussion), *Journal of the Royal Statistical Society, Series B*, 53, pp. 285–339.

GOSHTASBY, A. (1986) Piecewise linear mapping functions for image registration, *Pattern Recognition*, 19, pp. 459–466.

GOSHTASBY, A. (1987) Piecewise cubic mapping functions for image registration, *Pattern Recognition*, 20, pp. 525–533.

GOSHTASBY, A., GAGE, S. H. & BARTHOLIC, J. F. (1984) A two-stage cross correlation approach to template matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, pp. 374–378.

GREEN, P. J. & HAN, X.-L. (1992) Metropolis methods, Gaussian proposals, and antithetic variables, *Lecture Notes in Statistics*, 74, pp. 142–164.

GREEN, P. J. & SILVERMAN, B. W. (1994) *Nonparametric Regression and Generalized Linear Models* (London, Chapman and Hall).

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

GRENANDER, W., CHOW, Y. & KEENAN, D. M. (1991) *Hands: A Pattern Theoretic Study of Biological Shapes* (New York, Springer-Verlag).

HAMMERSLEY, J. M. & HANSCOMB, D. C. (1964) *Monte Carlo Methods* (London, Methuen).

HASTINGS, W. K. (1970) Monte Carlo sampling methods using Markov chains, and their applications, *Biometrika*, 57, pp. 97–109.

HEIKKILA, J. & SILVEN, O. (1997) Camera calibration and image correction using circular control points. *In: Scandinavian Image Analysis Conference SCIA97*, Lappeenranta, Finland, pp. 847–854.

HORGAN, G. W., CREASEY, A. M. & FENTON, B. (1992) Superimposing two-dimensional gels to study genetic variation in malaria parasites, *Electrophoresis*, 13, pp. 871–875.

HURN, M. A., MARDIA, K. V., HAINSWORTH, T. J., KIRKBRIDE, J. & BERRY, E. (1996) Bayesian fused classification of medical images, *IEEE Transactions on Medical Imaging*, 15, pp. 850–858.

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

KASS, M., WITKIN, A. & TERZOPOULOS, D. (1988) Snakes: active contour models, *International Journal of Computer Vision*, 1, pp. 321–331.

KUGLIN, C. D. & HINES, D. C. (1975) The phase correlation image alignment method. *In: Proceedings of the IEEE 1975 International Conference on Cybernetics and Society*, pp. 163–165.

KENT, J. T. & MARDIA, K. V. (1994) The link between kriging and thin plate splines. *In:* F. P. KELLY (Ed.), *Probability, Statistics and Optimization* (Wiley, New York), pp. 324–339.

LANDAU, L. D. & LIFSHITZ, E. M. (1986) *Theory of Elasticity* (Oxford, Pergamon Press).

LEE, D., KENT, J. T. & MARDIA, K. V. (1997) Tracking of tagged MR images by Bayesian analysis of a network of quads. *In:* J. DUNCAN & G. GINDI (Eds) *IPMI Conference Proceedings*, pp. 495–500 (Berlin, Springer-Verlog).

LITTLE, J. A., HILL, D. L. G. & HAWKES, D. J. (1996) Constraining rigid structures within a landmark based deformation. *In:* K. V. MARDIA, C. A. GILL & I. L. DRYDEN (Eds), *Image Fusion and Shape Variability Techniques* (Leeds, Leeds University Press), pp. 170–177.

MAELAND, E. (1988) On the comparison of interpolation methods, *IEEE Transactions on Medical Imaging*, 7, pp. 213–217.

MARDIA, K. V. & HAINSWORTH, T. J. (1993) Image warping and Bayesian reconstruction with grey-level templates, *Journal of Applied Statistics*, 20 (5&6), Special Issue, pp. 257–280.

MARDIA, K. V. & LITTLE, J. L. (1994) Image warping using derivative information. *In:* F. L. BOOKSTEIN, J. S. DUNCAN, N. LANGE & D. C. WILSON (Eds), *Mathematical Methods in Medical Imaging III*, *SPIE Proceedings*, 2299, pp. 16–31.

MARDIA, K. V., KENT, J. T., GOODALL, C. R. & LITTLE, J. L. (1996) Kriging and splines with derivative information, *Biometrika*, 83, pp. 207–221.

MARDIA, K. V., KENT, J. T. & WALDER, A. N. (1991) Statistical shape models in image analysis. *In:* E. M. KARAMIDAS (Ed.), *Computer Science and Statistics: proceedings of 23rd Symposium*, *Interface* (Interface Foundation, Fairfax Station), pp. 550–557.

MCCONNELL, R., KWOK, R., CURLANDER, J. C., KOBER, W. & PANG, S. S. (1991) Ψ-S correlation and dynamic time warping: two methods for tracking ice floes in SAR image, *IEEE Transactions on Geoscience and Remote Sensing*, 29, pp. 1004–1012.

MCINERNEY, T. & TERZOPOULOS, D. (1996) Deformable models in medical image analysis: a survey, *Medical Image Analysis*, 1, pp. 91–108.

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

MOSS, S. & HANCOCK, E. R. (1996) Registering incomplete radar images using the EM algorithm. *In: BMVC96—Seventh British Machine Vision Conference*, Edinburgh, pp. 685–694.

O'SULLIVAN, F. (1991) Discretized Laplacian smoothing by Fourier methods, *Journal of the American Statistical Association*, 86, pp. 634–642.

PHILLIPS, D. B. & SMITH, A. F. M. (1994) Bayesian faces via hierarchical template modelling, *Journal of the American Statistical Association*, 89, pp. 1151–1163.

PROPP, J. G. & WILSON, B. D. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics, *Symposium on Discrete Algorithms*.

RAMSAY, J. O. (1988) Monotone regression splines in action, *Statistical Science*, 3, pp. 425–461.

RICHARDS, J. A. (1986) *Remote Sensing Digital Image Analysis: An Introduction* (Berlin, Springer-Verlag).

ROSENFELD, A. & KAK, A. C. (1982) *Digital Picture Processing*, 2nd Edn (San Diego, Academic Press).

SAKOE, H. & CHIBA, S. (1978) Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26, pp. 43–49.

SAMPSON, P. D. & GUTTORP, P. (1992) Nonparametric estimation of nonstationary spatial covariance structures, *Journal of the American Statistical Association*, 87, pp. 108–119.

SKOVGAARD, I. M., JENSEN, K. & SONDERGAARD, I. (1995) From image processing to classification: III. Matching patterns by shifting and stretching, *Electrophoresis*, 16, pp. 1385–1389.

SOKAL, A. D. (1989) Monte Carlo methods in statistical mechanics: foundations and new algorithms. *In: Cours de Troisième Cycle de la Physique en Suisse Romande*, Lausanne.

STODDART, A. J., BRUNNSTROM, K., DATAKONSULTER, S. & DATAKONSULTER, T. (1996) Free-form surface matching using mean field theory. *In: BMVC96—Seventh British Machine Vision Conference*, Edinburgh, pp. 33–42.

TANG, Y. T. & SUEN, C. Y. (1993) Image transformation approach to nonlinear shape restoration, *IEEE Transactions on Systems, Man and Cybernetics*, 23, pp. 155–171.

TERZOPOULOS, D. (1986) Regularization of inverse visual problems involving discontinuities, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, pp. 413–424.

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

WANG, K. & GASSER, T. (1997) Alignment of curves by dynamic time warping, *Annals of Statistics*, 25, pp. 1251–1276.

WENG, J., AHUJA, N. & HUANG, T. S. (1993) Optimal motion and structure estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, pp. 864–884.

WOLBERG, G. (1988) Geometric transformation techniques for digital images: a survey, Technical report, *CUCS-390-88*, Columbia University, New York.

WONG, R. Y. & HALL, E. L. (1978) Scene matching with invariant moments, *Computer Graphics and Image Processing*, 8, pp. 16–24.