Sankhyā : The Indian Journal of Statistics
Special issue on Wavelets
2001, Volume 63, Series B, Pt. 2, pp. 181–198

A WAVELET MIXTURE APPROACH TO THE ESTIMATION OF IMAGE DEFORMATION FUNCTIONS

By T.R. DOWNIE
University College London, London, U.K.
and
B.W. SILVERMAN
University of Bristol, Bristol, U.K.

SUMMARY. Deriving a function that maps one image on to another similar image, or template, is a useful method in statistical image analysis. Such deformation functions can be modelled in terms of a suitable basis expansion. In a wavelet basis, it is reasonable to assume that most coefficients are zero, and this leads to a model where each wavelet coefficient has a mixture distribution with an atom of probability at zero. The model is used within a penalized least squares framework to fit a deformation to the image. The resulting deformation is explored by various visualization methods. The multiresolution property of the wavelet decomposition yields summary statistics of the deformation in terms of its energy at different scales and in different parts of the image. Inferences are then made on samples where each data point is itself an image. The methods are developed and illustrated in the context of a set of images of human femora, obtained from a palaeopathological study of osteoarthritis of the knee.

1. **Introduction**

A common method of analysing images is to model them as deformed versions of a "standard image," called a template. The method is often applied to medical images and can be used to analyse shape variation, particularly that due to some pathology. Our methodology was developed with reference to a particular data set of human femora arising from a palaeopathological study of osteoarthritis of the knee, but it is intended to be of much more general applicability.

Paper received. March 2001.
AMS (1991) *subject classification.* Primary 62H35; secondary 42C40.
Key words and phrases. Arthritis, deformed template, image analysis, iterated conditional mode, mixture distribution, palaeopathology, penalized least squares, shape statistics, wavelets.

It is usual in statistical work to model the deformation in terms of its expansion in a functional basis such as a Fourier series, a thin-plate spline basis or, more recently, a wavelet basis. We propose a wavelet mixture approach that models the set of deformation functions making use of the prior assumption that nearly all the wavelet coefficients will be zero. This improves on existing wavelet approaches to deformable templates by exploiting a general property of wavelet expansions of functions likely to occur in practice. The basic idea is that the model should give an economical representation of most deformations, particularly those with localized features. A variety of methodological issues are explored; our conclusion is that good results are obtained by using an integrated Haar wavelet basis and an iterated conditional mode optimization algorithm.

Once the deformation has been fitted, the economy of the wavelet representation makes it possible to conceptualize important aspects of the deformation, for example by examining the position of the largest coefficients within the wavelet array. We introduce a method of visualizing the wavelet coefficients that facilitates this. The localization of wavelets both in time and frequency makes it possible to quantify the energy of the deformation at different scales and/or in different parts of the image. We illustrate how the resulting statistics of the deformations can be used in an analysis of a set of images, for example allowing population comparisons to be made between the arthritic and non-arthritic specimens in the data set.

In Section 2, we review the basic ideas of deformation functions and of wavelet expansions, and also describe the palaeopathological femoral data. Wavelet approaches specifically to deformable templates are discussed in Section 3, and our own mixture model approach introduced. Details of possible implementations are set out in Section 4. In Section 5, approaches to the visualization of the fitted deformation are demonstrated. The main results (Section 6) fall into two parts. Firstly, the issue of what wavelet and what optimization method to use are explored through a detailed study of a particular image. Secondly, deformations are found for every image in the data set, and some statistical analysis of the wavelet expansions of the resulting deformations is carried out.

2. **Preliminaries**

2.1 *Deformation functions.* Consider an observed *image* I and a *template* T. As in Amit *et al.* (1991), I and T are defined to be real-valued functions on the unit square $[0,1]^2$. Also assume there exists a *deformation function* $f : [0,1]^2 \to [0,1]^2$, satisfying

$$I(x,y) \approx T\left\{f(x,y) + (x,y)\right\}, \qquad (x,y) \in [0,1]^2. \tag{1}$$

The deformation function models the relation between the image and template, and is useful for analysing the shape variation between different observed images. For any deformation function $f = (f_x, f_y)$, we call $DT_f(x, y) = T\{f(x, y) + (x, y)\}$ the *deformed template* defined by f. For many applications the "best" deformation will not give equality between the deformed template and the image, and the regularity of f, in some sense, will also be considered.

A probability model for f can be specified by considering a suitable basis expansion. For example, Amit *et al.* (1991) propose a Fourier prior model, setting

$$f_x(x, y) = \sum_{m,n=0}^{\infty} \xi_{mn}^x \lambda_{mn} \sin(m\pi x) \cos(n\pi y)$$

and

$$f_y(x, y) = \sum_{m,n=0}^{\infty} \xi_{mn}^y \lambda_{mn} \sin(n\pi y) \cos(m\pi x).$$

The coefficients ξ_{mn}^x and ξ_{mn}^y are assumed to be independent $N(0, 1)$ random variables, and the regularity of deformations drawn from the model depends on the constants λ_{mn}.

They then construct the likelihood $l(I|f)$ of observing I given a deformation f by assuming that, independently for each (x, y) on a finite grid $G \subset [0, 1]^2$, $I(x, y)$ has a normal distribution with mean $DT_f(x, y)$ and variance τ^2. They obtain a local maximum a posteriori (MAP) estimate using a gradient descent method. This requires $T(x, y)$ to be a smooth function rather than the piecewise constant function that will result from using a pixel image, whether binary, gray scale, or colour, as the template.

For a comprehensive survey and bibliography of work on image warping, see Glasbey and Mardia (1998). In a recent paper by the same authors, a penalized likelihood approach is used to estimate the deformation function (Glasbey and Mardia, 2001). Thin plate spline penalties, and their extensions, are used on the functions f_x and f_y, and alternatives to the likelihood $l(I|f)$ are also considered.

2.2 *Wavelet expansions.* A one-dimensional wavelet basis is generated by dilations j, and translations k, of a *mother wavelet* ψ (Daubechies, 1992), to give wavelets $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j} - k)$. A general one dimensional function, $f : \mathbb{R} \mapsto \mathbb{R}$, can then be expressed as a wavelet expansion

$$f(t) = \sum_{j,k \in \mathbb{Z}} w_{j,k} \psi_{j,k}(t) \tag{2}$$

for some $w_{j,k} \in \mathbb{R}$. The mother wavelet has short effective support both in the time and frequency domains, and so each coefficient $w_{j,k}$ contains information about f

localized both in time and frequency. Each Fourier coefficient, on the other hand, contains information about f across the whole range of time t.

For a two-dimensional function, $f : \mathbb{R}^2 \mapsto \mathbb{R}$, three mother wavelets are required, corresponding to a horizontal, vertical and diagonal *orientation* (Mallat, 1989). Each wavelet is a dilation j, and translation in two dimensions k_1, k_2, of one of the three mother wavelets indexed by $l = 1, 2$ or 3. So the two dimensional equivalent to (2) is

$$f(x, y) = \sum_{\kappa} w_\kappa \Psi_\kappa(x, y), \tag{3}$$

where κ is an index $\kappa = (j, k_1, k_2, l)$ taking values in the index set $\mathbb{Z}^3 \times \{1, 2, 3\}$. The standard practice is to consider functions defined over a finite window at a discrete grid of points, and use a discrete wavelet transform (DWT). This yields a corresponding expansion to (3) but with a finite index set $\mathcal{K}$.

A function f that can be expressed (or well approximated) using a small number of non-zero coefficients is said to have an *economical representation*. The localization property of wavelets means that a wide class of functions useful in practice, including those that are piecewise smooth with local irregularities or discontinuities, have economical wavelet representations. For further reading on wavelets see Chui (1992) or Daubechies (1992). Nason and Silverman (1994), and Downie (1997) both discuss statistical methods that have utilized wavelet theory. The volume Silverman and Vassilicos (1999) contains several papers discussing recent applications of wavelet methods.

In principle, wavelets can be used to model deformations in three (or more) dimensions. In three dimensions there will be seven possible orientations of the wavelets and a three-dimensional integer translation. As a result, the index $\kappa = (j, k_1, k_2, k_3, l)$ will take values in $\mathbb{Z}^4 \times \{1, 2, \ldots, 7\}$. In this case, even a relatively economical transformation may have a large number of nonzero coefficients, but any difficulties in the extension of our work are likely to be computational rather than conceptual.

2.3 *The archaeological data*. The data used to develop the wavelet deformation approach arose from the analysis of a sample of human femora (thigh bones), taken from an archaeological collection of skeletal remains. See Shepstone *et al.* (1999) for details of the data and their collection. Of particular interest is the relationship between osteoarthritis of the knee and the shape of the distal (knee) end of the femur. Osteoarthritis can be diagnosed on the bones by the presence of 'eburnation', polished areas of articular surface. It has been speculated by researchers in the field that the shape of the knee may predispose to the development of osteoarthritis (Bullough, 1981; Cook *et al.*, 1997). For the present study, two-dimensional images of the femora, viewed axially, were available. Pixels corresponding to various bone

changes were identified, but our main interest will be in the binary image yielding the overall shape of the joint. One bone, unaffected by any pathology, was arbitrarily designated as the template. The registration of the images to one another was performed during the data collection phase; the bones were positioned relative to the camera in a consistent way, as explained in Shepstone *et al.* (1999).

Our methodology—equally applicable to gray scale or colour images—is to fit a deformation function for each of the images using a wavelet approach. Because of the economy of the wavelet expansion, individual large wavelet coefficients give particular information about the deformation. Furthermore, the localization properties of the wavelet expansion make it possible to find statistics of the deformation, for example the amount of 'energy' at particular scales and/or in particular parts of the image. In our application, these measures will be related to osteoarthritis and also to postmortem bone damage.

3. **Wavelet Deformation Models**

3.1 *Previous approaches*. Using a wavelet basis to model a deformation function is a natural extension of the Fourier model, especially bearing in mind that many deformations will have localized features, and therefore will be more appropriately represented by an economical wavelet expansion than by a Fourier expansion. A wavelet basis is used as a model for the deformation in Amit (1994). Amit uses a penalized least squares (PLS) approach, defined as follows. Let $\text{SSE}_f = \sum_{(x,y)\in G}\{I(x,y) - DT_f(x,y)\}^2$, the sum of squares of errors between DT and I over a fixed fine grid G. For binary images SSE_f is simply the number of grid points at which the image and deformed template differ; for colour images, on the other hand, it will be desirable to extend the definition appropriately.

In general, the PLS approach (see Green and Silverman (1994) for example) finds a minimum of $S(f,\alpha) = \text{SSE}_f + \alpha\,\text{PEN}(f)$, where $\text{PEN}(f)$ is a function that quantifies the roughness or complexity of f, and α is a smoothing parameter that controls the trade-off between roughness and fit to the data. In a Bayesian setting, the penalty corresponds to the negative log prior density for f. Amit's work uses $\text{PEN}(f) = \sum_\kappa w_\kappa^{x2} + w_\kappa^{y2}$, where w_κ^x and w_κ^y are the wavelet coefficients for f_x and f_y respectively. This corresponds to a prior where the wavelet coefficients have independent normal distributions, and the notion that appropriate deformations will have economical wavelet expansions is not incorporated. Amit uses periodic boundary conditions, and optimizes the PLS score by a steepest descent method that requires T to be differentiable.

Aykroyd and Mardia (1996) use wavelets to model one-dimensional deformation functions. They study spinal images where both the data and deformation are

expressed in one dimension and each vertebra is a well defined landmark. Their prior is defined in the domain of the deformation, not of the wavelet coefficients, and depends on the squared differences in the deformation between neighbouring vertebrae, with $\text{PEN}(f) = \sum_i \{f(x_{i+1}) - f(x_i)\}^2$. Posterior estimates are obtained using MCMC methods, and the posterior mean is used as a point estimate for each wavelet coefficient.

3.2 *A wavelet mixture model*. Neither model discussed in Section 3.1 yields an economical wavelet expansion for the prior deformation. The resulting coefficients may be thresholded after an optimal deformation has been obtained, but there is no guarantee that this 'smoothing' will result in an acceptable deformation. In order to model economy throughout the procedure we model the coefficients w_κ^x and w_κ^y as independent random variables with the following mixture distribution:

$$w_\kappa^{\cdot} = \begin{cases} 0 & \text{w.p.} \quad p_j \\ Z_\kappa & \text{w.p.} \quad 1 - p_j, \end{cases} \quad \text{where } Z_\kappa \sim N(0, \sigma_j^2). \tag{4}$$

A coefficient will be exactly zero with probability p_j; or else it has a normal distribution with mean zero and variance σ_j^2. The parameters p_j and σ_j may depend on the resolution level j of the associated wavelet function. The one-dimensional Bayesian thresholding method of Abramovich, Sapatinas and Silverman (1998) also uses this mixture distribution. They discuss some specific analytical aspects of this prior, for example, the relationship of the values of p_j and σ_j to the smoothness and differentiability properties of the functions generated by the prior.

3.3 *PLS and Bayesian optimization*. We proceed by using a PLS approach, with penalty term given by $\text{PEN}(f) = -\log \pi(f)$ where $\pi(f)$ is the density of the wavelet mixture model (4)

$$\pi(f) = \prod_{\nu \in \{x,y\}} \prod_{\kappa \in \mathcal{K}} \{p_j \delta_{0,w_\kappa^\nu} + (1 - p_j)\phi(w_\kappa^\nu/\sigma_j)(1 - \delta_{0,w_\kappa^\nu})\}.$$

Here ϕ is the standard normal density function. To be precise, π is the Radon-Nikodym derivative with respect to an appropriate mixed dominating measure.

Under the assumption that the errors are i.i.d. normal over the grid G, the Bayesian posterior density of a deformation given the image would be $\pi(f|I) \propto \exp\{-\frac{1}{2}\tau^{-2}\text{SSE}_f + \log \pi(f)\}$ and so, as usual, finding the posterior mode is equivalent to minimizing the PLS score. Since the assumption of independent errors on the grid G is not always appropriate, we prefer to regard our approach as a PLS method motivated by Bayesian considerations, rather than as a strict Bayesian method.

3.4 *Global deformations*. An interesting feature of wavelet expansions is that if a function f is linear, then the mother wavelet coefficients of f are all zero, provided

the wavelet has at least one vanishing moment, and ignoring boundary effects. Examples of deformations that give linear functions f_x and f_y are global rotations, translations, shearing transformations, and rescaling, possibly by different amounts in different directions. Thin plate spline roughness penalties have an analogous property, exploited by Glasbey and Mardia (2001): the roughness penalty of any linear transformation is zero. In both cases, boundary effects come into play when the image and template are both considered over fixed finite regions. However, neglecting boundary conditions, any deformation that is predominantly a linear transformation will certainly have an economical wavelet expansion, provided suitable wavelets are used.

In the present work, we consider images that have been registered to fit closely to one another, and so this feature is not immediately relevant, but it is an attractive aspect of the wavelet approach more generally. It is indeed possible to avoid any boundary conditions by defining the transformation over the whole plane, keeping track only of the nonzero wavelet coefficients. For an application of wavelets where such an algorithmic approach is used, see Herrick, Nason and Silverman (2001).

4. **Implementation**

Of the many minimization algorithms that obtain a deformation with a low PLS score, we have implemented three: ICM, simulated annealing, and a greedy algorithm. More detailed discussion is given by Downie (1997). In particular, the structure of the wavelet transform makes various economies possible. Each algorithm is devised to work with binary, coloured or gray level images, and no smoothness assumptions are made. Because of the multiresolution properties of wavelets, it is sensible to incorporate a cascade approach, as applied in conventional image restoration problems by Hurn and Jennison (1995). In the wavelet context, where appropriate we first optimize over the low frequency coefficients, and then add (cascade) higher frequency levels to obtain a more refined solution.

ICM. The iterated co-ordinatewise maximization (ICM) routine of Besag (1986) minimizes the PLS score over each coefficient in turn, updating the coefficient values and the deformation at each stage. This is repeated for a fixed number of iterations, considering the coefficients in 'cascade' order. A simple grid search method is used for each coefficient, so there is no need for the template to be smooth.

Simulated Annealing. We can optimize the PLS score by a simulated annealing approach using the Metropolis algorithm (see Gilks *et al.*, 1996). This approach

should allow the deformation to wander in the high dimensional space of wavelet coefficients, gradually converging to a region with a low PLS score. The simulated annealing 'temperature' is reset to one at the start of each cascade.

Greedy algorithm. This approach recursively accepts coefficient values that individually reduce SSE_f as far as possible. To ease the computational burden, an approximation is used: the whole wavelet table is searched and the coefficients that individually decrease SSE_f are listed in order of their effect on SSE_f. We then introduce the coefficients in this order, at each stage finding the best possible reduction in SSE_f, stopping when no decrease in SSE_f is obtained. This process is repeated, and in practice converges after a very small number of iterations.

Both ICM and Simulated Annealing require a choice of smoothing parameter. For the Greedy algorithm, a smoother or more economical deformation is found either by stopping after finding a certain number of coefficients, or by discarding the coefficients that gave the smallest change in SSE_f and only constraining SSE_f to remain below a specified quantity. The parameters of the prior p_j and σ_j^2 also have to be chosen. When deforming the femoral condyle images, the parameters were chosen by trial and error, allowing a low SSE_f to be attained if fitting a small smothing parameter. The effect of varying the priors is investigated and discussed in Downie (1997). More sophisticated ways of chosing the prior parameter values could be used, at the cost of increased computation.

Two different wavelets were used, the Haar Wavelet and the Daubechies D4 Wavelet (Daubechies, 1992). In addition to expressing f_x and f_y as wavelet expansions, we also considered an *Integrated Haar* expansion, where $\partial f_x/\partial x$ and $\partial f_y/\partial y$ are expanded in Haar wavelets, giving

$$\begin{aligned} f_x(x,y) &= \sum_\kappa w_\kappa^x \int_0^x \Psi_\kappa(u,y)du \\ \text{and } f_y(x,y) &= \sum_\kappa w_\kappa^y \int_0^y \Psi_\kappa(x,v)dv. \end{aligned}$$

This basis is a two dimensional linear spline basis. The basis functions are piecewise linear with very short support, but are not orthogonal wavelets. Considering the effect of integrating each basis function, we find that only two wavelet orientations are required. For example, the horizontal Haar wavelet is (up to a constant) -1 in the top half of its support and $+1$ in the bottom half. Integrating this in the x-direction gives a basis function that is uniformly decreasing in the top half

and increasing in the bottom half. Thus some points may be mapped to a position that is outside the span of the wavelet, creating a 'fold' in the deformed template. However, integrating the same wavelet in the y-direction does not have this folding effect so is an acceptable basis function. Correspondingly the vertical wavelet is applicable for deformations in the x-direction, but not in the y-direction, and the diagonal wavelet is valid in both directions. For convenience, we shall use the term *primary orientation* for the vertical wavelet in the x-direction, and the horizontal wavelet in the y-direction.

When using the Haar and Integrated Haar bases no boundary corrections are required, because the wavelet functions all have support within the unit square. For the Daubechies wavelets, symmetric boundary conditions are used.

5. **Visualization**

The aims of the deformable template method should not merely be to warp the image to the template, but also to convey information or understanding about the deformation itself. In many contexts the deformation carries important information, for example about the shape of the bone underlying the image. To some extent, the utility of our presentations is subjective, and particular readers may well find some presentations more valuable than others. We briefly explain the three types of visualization used.

Figure 1 depicts the deformation function directly by overlaying arrows on the template. The points $f(x_s, y_t) + (x_s, y_t)$ in the template (the arrow tails) become

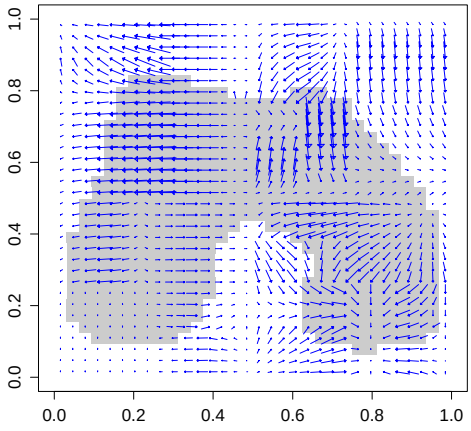

Figure 1. The deformation function, using ICM with $\alpha = 1.5$.

the grid points (x_s, y_t) of the deformed template (the arrowheads). So the value of the deformed template at the position of each arrowhead is the value of the template underneath the corresponding arrow tail. Figure 2 shows how well the deformation fits the template to the image. Four plots are shown, the template, the image, "DT\image" and "Image\DT". The plot DT\image shows the deformed template, highlighting those pixels which are within the outline of the femur in the deformed template but not in the image. Conversely Image\DT shows the image, highlighting those pixels that are within the original femoral outline but not in the deformed template.

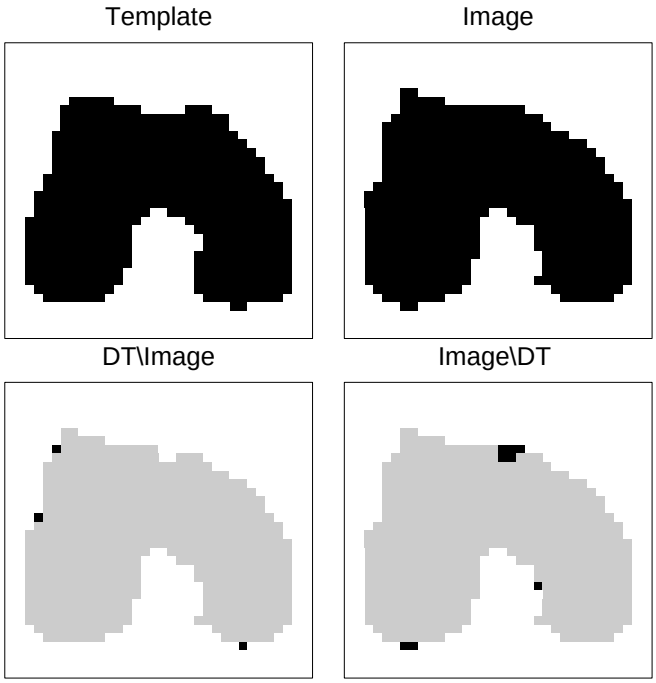

Figure 2. The femoral template, image, and plots comparing the deformed template with the image, using ICM with $\alpha = 1.5$. In the plot DT\image the pixels in both the deformed template and the image are plotted in grey, while the black pixels are in the deformed template alone. The black pixels in the plot image\DT are in the image but not in the deformed template.

The third type of plot, as in Figure 3, displays the non-zero coefficients in the wavelet expansion of the deformation. The two components (x and y) are displayed on two separate diagrams. Each letter appears at the centre of the relevant wavelet's support. The letter represents the resolution level (A is level 0, E is level 4) and hence the support of the wavelet; for example, for the Haar

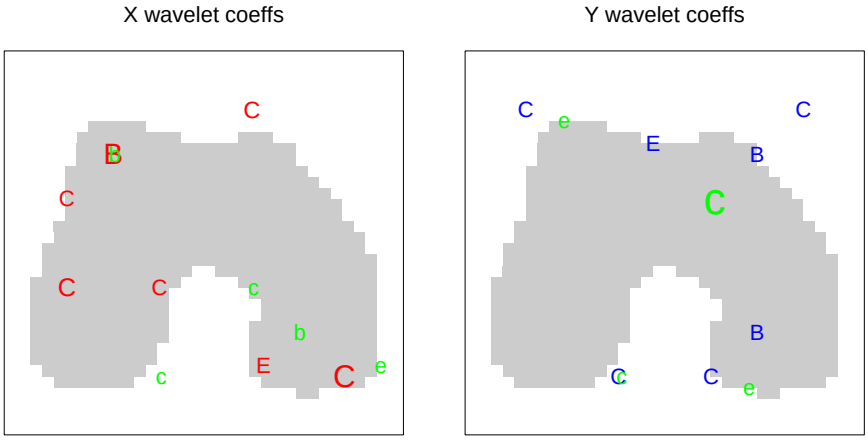

Figure 3. The wavelet coefficient positions, using ICM ($\alpha = 1.5$).

wavelet, the support of a wavelet indicated by the letter D is a square of side $\frac{1}{8} = 2^{-3}$. The size at which the letter is printed represents the absolute magnitude of the coefficient. Lower case letters correspond to the diagonal wavelet and upper case letters to the primary orientation. In coloured formats the colour can also be used to indicate the wavelet's orientation; see Downie (1997) and the web page `www.stats.bris.ac.uk/~bernard/deftem.html`. The salient features and interpretation of these figures are discussed in the following section.

6. **Results**

We first investigate a single image in detail, to demonstrate the visualization approaches discussed in Section 5 above, and also to explore the various algorithms and wavelets. For a more detailed analysis of the methods see Downie (1997). We then move on to an analysis of the femoral images, demonstrating ways that the wavelet approach allows inferences to be made on the shape variation among the images. It is possible both to pick out individual cases with extreme characteristics of various kinds, and to explore population differences between different groups in the data.

6.1 *Methodological investigation*. We concentrate on a coarse version (32^2 pixels) of a particular image and a corresponding version of the template, illustrated in Figure 2. At this resolution, there are five wavelet resolution levels; $j = 0$ (coarsest) to $j = 4$ (finest). In our discussion we shall use the phrase 'number of

coefficients' to denote the number of *nonzero* coefficients in the wavelet expansion of a deformation.

Of the two model-based methods, the considerable additional computational burden of simulated annealing did not give appreciably better results than ICM. On the other hand, the greedy algorithm does not allow the control over the quality of fit possible when the mixture-model penalty function is used. Therefore we recommend use of the ICM algorithm, and only report the ICM results in detail.

Table 1. USING THE ICM ALGORITHM WITH DIFFERENT VALUES OF α.

Smoothing parameter α	∞	2.0	1.5	1.25	1.0	0.5	0.1
Sum of squares of errors	53	20	11	6	6	2	2
Number of nonzero coefficients	0	11	23	27	31	34	34

Table 1 shows the results from the ICM algorithm for different values of α, using the integrated Haar basis. As α decreases, SSE_f decreases and more coefficients are used, until $\alpha \leq 0.5$ when the algorithm reaches its best match. For $\alpha = 1.0$ the SSE_f is the same as for $\alpha = 1.25$, but uses more coefficients in the deformation. In some examples SSE_f has been observed to be not quite strictly monotone in α. This reflects the fact that the algorithm may only obtain a local minimum of $S(f, \alpha)$, and the deformation obtained for large α may not simply be a contraction of the deformation for a smaller α.

For $\alpha = 1.5$, Figure 1 gives the deformation itself, in conventional form, and Figure 2 shows a comparison of the deformed template and the image. Figure 3 presents the coefficients in the manner described in Section 5 above. It can be seen that most of the coefficients are near the edge of the femur, demonstrating the localization properties of the wavelet basis. At the bottom left of the y component, coefficients from resolution levels one, two and four (denoted B, C and e in the figure) are present, indicating that both broad scale (levels 1 and 2) and detailed (level 4) warping is required. Figures 4 and 5 show the deformation and coefficients for $\alpha = 0.5$. Comparing Figure 5 with Figure 3, again most of the coefficients lie around the outline but for smaller α there are more coefficients. Though most, but not all, of the coefficients in Figure 3 also appear in Figure 5, there are some coefficients in Figure 3 not present in Figure 5, confirming that the deformation when $\alpha = 1.5$ is not merely a smoother version of the deformation obtained when $\alpha = 0.5$.

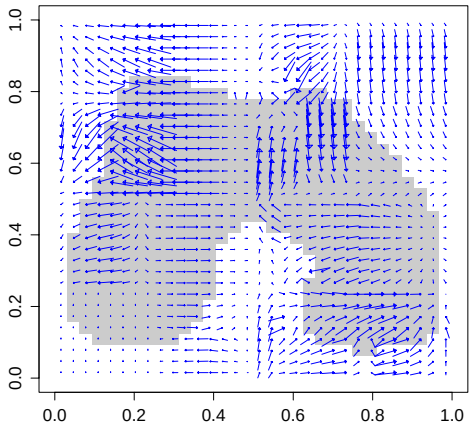

Figure 4. The deformation function, using ICM with $\alpha = 0.5$.

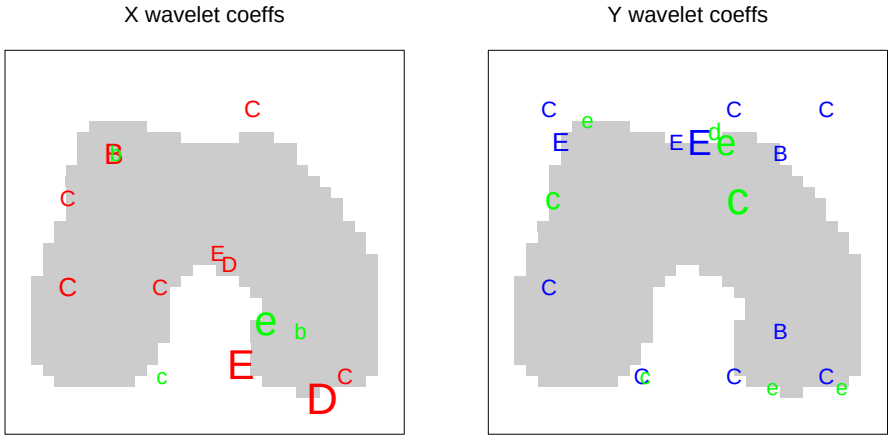

Figure 5. The wavelet coefficient positions, using ICM ($\alpha = 0.5$).

We now turn to the choice of wavelet. Using ICM with the Haar wavelet, for $\alpha = 0.5$ we achieved $\text{SSE}_f = 3$ with 34 nonzero coefficients, while for $\alpha = 1.25$, $\text{SSE}_f = 17$ with 19 nonzero coefficients. Comparing these results with Table 1 (for the same α), the Haar Wavelet gave a worse deformation than the Integrated Haar Basis in terms of SSE_f. The Daubechies D4 wavelet gave slightly better results than the Haar but was still inferior to the integrated Haar basis, which generally gave better fitting deformations with fewer coefficients. Therefore this basis was used for future investigation, even though it is numerically somewhat more intensive to fit than an orthogonal wavelet basis.

6.2 *Analysing images*. We used the ICM algorithm with the Integrated Haar Basis to obtain deformations on the whole population of 121 images. Of this population, 75 (including the template) were intact and were not eburnated, and so show no indication of osteoarthritis. With the template omitted, these form a 'control' group. Of the remainder, 22 were broken, and 24 (the 'test group') were intact but eburnated. The control/test/broken information was not used in the fitting of the deformations. The images are defined on a 128^2 pixel grid, and the deformation was evaluated at 64^2 points. A small smoothing parameter was used ($\alpha = 0.1$) to ensure a good fit between the deformed template and image. Five resolution levels were searched over, so that the finest resolution level has not been fitted. Omitting the finest level gives a considerable saving in algorithm running time, at very little cost in terms of fit.

An important feature of the wavelet approach is the possibility of finding the energy of the deformation—the sum of squares of relevant subsets of coefficients—corresponding to different parts of the image and/or effects on different scales. We considered the energy in the wavelet transform broken down according to resolution levels 0 to 4 and the four quadrants of the image (labelled A: top left, B: top right, C: bottom left, D: bottom right) as well as by the orientation of the wavelet and the x and y components of the transform. The coefficients at level zero span the whole image and so cannot be assigned to any quadrant. We are able to analyse localizations of the deformation function both in time and frequency, a direct consequence of using a wavelet approach.

Table 2. THE MEAN TOTAL ENERGY FOR IMAGES IN THE VARIOUS GROUPS. THE ENERGIES ARE WELL SYMMETRIZED BY A LOG TRANSFORMATION. AFTER THIS HAS BEEN PERFORMED, T-TESTS SHOW THE DIFFERENCE BETWEEN THE EBURNATED AND CONTROL GROUPS TO BE SIGNIFICANT WITH $p = 0.02$ AND THE DIFFERENCE BETWEEN THE BROKEN AND INTACT BONES TO BE SIGNIFICANT WITH $p < 0.0001$. THE SIGNIFICANCE LEVELS ARE CALCULATED INDEPENDENTLY FOR EACH TEST.

	mean	s.e.
broken	2.414	0.297
eburnated	0.664	0.055
control	0.541	0.038

The broken images had the highest total energy, since a large deformation is needed to fit the break effectively. Table 2 gives the mean energies in the various groups. A logarithmic transformation symmetrizes the energy within groups and stabilizes the variances. On the basis of t-tests, the difference between the broken and the intact bones is significant beyond the 0.01% level, and the difference between the eburnated and control group is significant at the 2% level. The coarse

levels of the transform showed particularly large differences between the the broken and the intact bones; on the average, at levels 0 and 1 the deformations for broken bones had about nine times as much energy as those for the intact ones. In the subsequent analysis, we considered only the intact bones.

For each bone, it is recorded whether the bone is a left or right femur. (The images for the left femora are reflected before any processing takes place.) Linear models were fitted to the arrays of energy broken down by the various factors, and no significant effects or interactions involving the left/right variable were found. Thus, after reflecting the left knee images, there is no evidence in this data set of any difference in shape between left and right knees.

Table 3. MEAN ENERGIES ($\times 1000$) WITHIN EACH QUADRANT FOR CONTROL AND EBURNATED BONES. THE SIGNIFICANCE LEVELS OF THE DIFFERENCES ARE OBTAINED FROM T-TESTS OF THE LOG ENERGIES WITHIN QUADRANTS. THE SIGNIFICANCE LEVELS ARE CALCULATED INDEPENDENTLY FOR EACH TEST. AN ANALYSIS OF VARIANCE SHOWS THAT THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN THE QUADRANTS FOR THE CONTROL GROUP.

Quadrant	mean (controls)	mean (eburnated)	p-value
A	131	226	0.001
B	129	111	NS
C	110	171	0.019
D	155	133	NS

Where do the eburnated bones differ in shape from the controls? Breaking down the energy by quadrants is instructive. In Table 3 the mean energy by quadrant for the eburnated and control bones is shown; coefficients at level 0 cannot be assigned to any quadrant and are not considered in this analysis. It can be seen that there are significant differences between the two populations in quadrants A and C, the left part of the image, but not in the corresponding quadrants on the right hand half; this is an interesting topic for further consideration by rheumatologists.

Now consider breaking down the energy by the level of the transform. The proportion of energy at different levels highlights different types of shape variation. Images with a large proportion of energy at level 0 have large-scale distortion from the template. Figure 6 shows an example of this. Images with a high proportion of energy in resolution level 4 have undulated edges, which require small scale deformations in order to match to these. An example of such a deformation is shown in Figure 7. These figures demonstrate the ability of the wavelet method to pick out particular cases for individual examination.

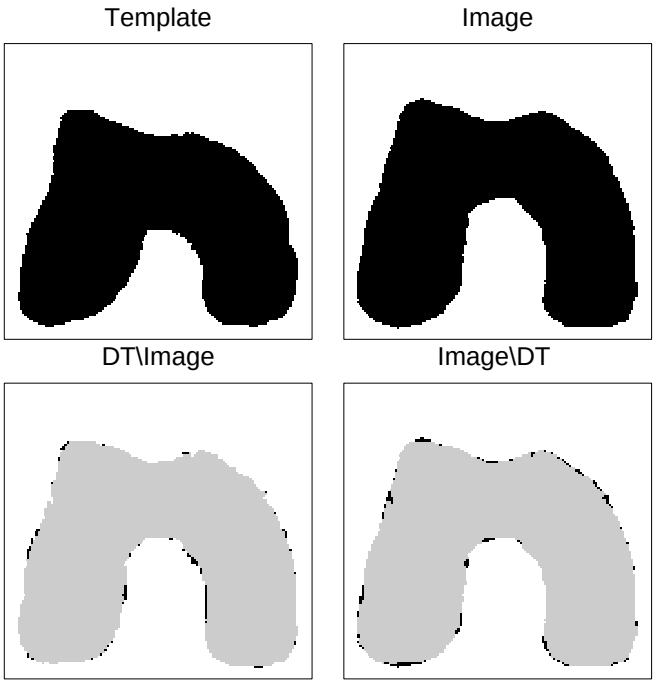

Figure 6. A deformed template that has a high proportion of energy at level 0.

Investigating the energy at different resolution levels over the whole sample, Table 4 demonstrates that the important differences between the energy for the test and control groups are at the larger scale levels. Thus, on the population level, it is at the larger-scale levels that the arthritic bones show differences in shape. The osteophyte growth characteristic of arthritic bones does not appear to have contributed significantly to the smaller scale effects.

Table 4. MEAN ENERGIES ($\times 1000$) WITHIN LEVELS FOR CONTROL AND EBURNATED BONES. THE SIGNIFICANCE LEVELS OF THE DIFFERENCES ARE OBTAINED FROM SEPARATE T-TESTS OF THE LOG ENERGIES AT EACH LEVEL; AT LEVEL 0 A RANK-SUM TEST WAS USED BECAUSE OF THE ZEROES IN THE DATA. TWO-SIDED p-VALUES ARE QUOTED, SO THE DIFFERENCE AT LEVEL 2 IS SIGNIFICANT ON A ONE-SIDED TEST.

Level	mean (controls)	mean (eburnated)	p-value
0	16.5	23.3	NS
1	104	166	0.001
2	313	367	0.07
3	88.4	87.9	NS
4	20.2	20.1	NS

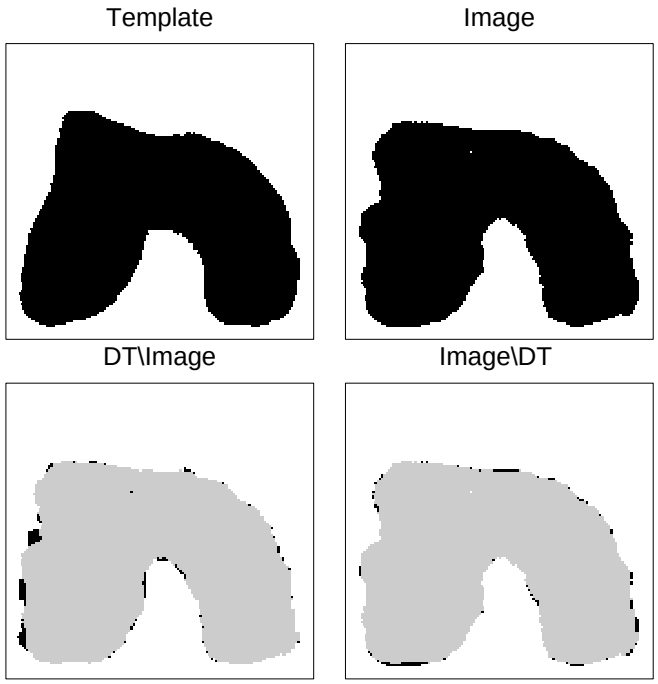

Figure 7. A deformed template that has a high proportion of energy at level 4.

Acknowledgements. This work originated as part of Tim Downie's PhD project at the Department of Mathematics, University of Bristol. Part of the work was carried out while Bernard Silverman was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, supported by NSF grant number SBR-9601236. Financial support was also provided by CSIRO and EPSRC. We are very grateful to Lee Shepstone for providing the palaeopathology data he collected as part of his PhD work under the supervision of John Kirwan and Bernard Silverman.

References

ABRAMOVICH, F., SAPATINAS, T. and SILVERMAN, B.W. (1998). Wavelet thresholding via a Bayesian approach, *Journal of the Royal Statistical Society*, Series B, **60**, 725–749.

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

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

AYKROYD, R.G. and MARDIA, K.V. (1996). An MCMC approach to wavelet warping. *Pages 129–140 of:* Mardia, K.V., Gill, C.A. and Dryden, I.L. (eds), *Image Fusion and Shape Variability Techniques*. Leeds University Press, Leeds.

BESAG, J. (1986). On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society,* Series B, **48**, 259–302.

BULLOUGH, P.G. (1981). The geometry of diarthrodial joints, its physiologic maintenance and the possible significance of age-related changes in geometry-to-load distribution and the development of osteoarthritis, *Clinical Orthopaedics and Related Research*, **256**, 61–66.

CHUI, C.K. (1992). *An Introduction to Wavelets*. Academic Press, London.

COOKE, D., SCUDAMORE, A., LI, J., WYSS, U., BRYANT, T. and COSTIGAN, P. (1997). Axial lower-limb alignment: comparison of knee geometry in normal volunteers and OA patients, *Osteoarthritis and Cartilage*, **5**, 39–47.

DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia.

DOWNIE, T.R. (1997). *Wavelet Methods in Statistics*. Ph.D. thesis, University of Bristol.

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

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

—— (2001). A penalized likelihood approach to image warping (with discussion). *Journal of the Royal Statistical Society,* Series B, **63**, 465–514.

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

HERRICK, D.R.M., NASON, G.P. and SILVERMAN, B.W. (2001). Some new methods for wavelet density estimation. *Sankhyā* Series A, **63**.

HURN, M. and JENNISON, C. (1995). A study of simulated annealing and a revised cascade algorithm for image reconstruction, *Statistics and Computing*, **5**, 175–190.

MALLAT, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **11**, 674–693.

NASON, G.P. and SILVERMAN, B.W. (1994). The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, **3**, 163–191.

SHEPSTONE, L., ROGERS, J., KIRWAN, J. and SILVERMAN, B.W. (1999). The shape of the distal femur: a palaeopathological comparison of eburnated and non-eburnated femora, *Annals of the Rheumatic Diseases*, **58**, 72–78.

SILVERMAN, B.W. and VASSILICOS, J.C., editors (1999). Wavelets: the key to intermittent information? *Philosophical Transactions of the Royal Society of London,* Series A, **357**, 2393–2625.

T.R. DOWNIE
DEPARTMENT OF STATISTICAL SCIENCE
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON WC1E 6BT, UK.
E-mail: tim@stats.ucl.ac.uk

B.W. SILVERMAN
INSTITUTE FOR ADVANCED STUDIES
UNIVERSITY OF BRISTOL
ROYAL FORT HOUSE
BRISTOL BS8 1UJ, UK.
E-mail: b.w.silverman@bristol.ac.uk