Economical Representation of Image Deformation Functions Using a Wavelet Mixture Model

T. R. Downie¹*, L. Shepstone² and B. W. Silverman³

¹CSIRO Mathematical and Information Sciences, Locked Bag 17, North Ryde, NSW 1760, Australia.

²Rheumatology Unit, Division of Medicine, University of Bristol, Bristol, U.K.

³School of Mathematics, University of Bristol, Bristol BS8 1TW.

SUMMARY

Deriving a function that maps one image on to another similar image is a useful method in medical imaging. The modelling of this deformation function in terms of a functional basis leads to numerical methods of obtaining a good deformation. It is advantageous to be able to represent a wide class commonly observed deformations economically, i.e. where most coefficients are zero. We proposed a wavelet model for the deformation, where each wavelet coefficient has the mixture distribution of the form \( p\delta_0 + (1 - p)N(0, \sigma^2) \). This distribution reflects our prior belief in the wavelet coefficients and results in an economical representation for the deformation. To implement this method, a penalised least squares methodology is adopted and three algorithms are devised. A numerical assessment of the method is made by applying the algorithms to images of femoral condyles, in which the deformations have predominantly localised features. A new method of visualising the wavelet decomposition of the deformation is presented.

*Email address: tim.downie@cmis.csiro.au.

Key Words: Wavelet expansions, deformed template, mixture distribution, penalised least squares, Markov Chain Monte Carlo, femoral condyle.
1 Introduction

A common method of analysing an image is to model it as a deformed version of a “standard image,” called a template. The assumption is that the image is a realisation of the template, with observed variability being due to sample variation, or imaging imperfections such as blurring or noise. The method is often applied to medical images and can be used to analyse components of shape variation, particularly those due to some pathology. It is usual to model the deformation in terms of a functional basis such as a Fourier basis, thin-plate spline bases or more recently a wavelet basis. After brief introductions to deformation functions and wavelet bases (this section), some wavelet models for deformations are reviewed (Section 2). We propose a wavelet mixture model, which assigns a probability density to the set of deformation functions (Section 3). This model is consistent with the a priori assumption that nearly all wavelet coefficients will be zero and so exploits a general property of wavelet expressions better than previous wavelet models. The basic idea is that the model should give an economical representation of most deformations, particularly those with localised deformations. Deformations that can be expressed in only a few coefficients are desirable when analysing the shape variation of many images.

Our methodology was developed with reference to a data set of femoral condyle (knee bone) images, many of which are of a similar shape except at the edges (Section 4). A penalised least squares method of evaluating any specific deformation is defined (Section 5) and three algorithms are developed to obtain a deformation with desired smoothness (Section 6). Three different methods of visualising the deformation function and the deformed template are used, one of which is specific to the economical wavelet representation derived earlier in this paper. To demonstrate the method, results from these algorithms are presented and discussed in Section 7.
1.1 Deformation Functions

Consider an observed image $I$ and a template $T$. As in Amit et al. (1991), assume that $I$ and $T$ are real valued and defined on the unit square $[0, 1]^2$. Also assume there exists a deformation function

$$f : [0, 1]^2 \rightarrow [0, 1]^2,$$

which is a bijection satisfying

$$I(x, y) = T \{ f(x, y) + (x, y) \}, \quad (x, y) \in [0, 1]^2. \quad (1)$$

The effect of the deformation function is to warp the Cartesian grid upon which the template is defined. The purpose of obtaining a deformation function is to model the difference between the image and template, useful for analysing the shape variation between different observed images. The function $f$ is expressed as an expansion relative to a suitable basis. The probability modelling of $f$ may be carried out in terms of its basis coefficients.

In Equation (1) the template is warped by $f$ so that it fits the image exactly, however, any valid deformation may define a deformed template. 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; usually the smoothness of $f$, in some sense, will also need to be considered.

Amit et al. (1991) propose a probability model for $f$ expressed in terms of a Fourier basis.

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

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 $\xi_{mn}^x$ and $\xi_{mn}^y$ are assumed to be $N(0, 1)$ random variables, which defines the prior distribution of $f$. The regularity of deformations drawn from this model depends on the
constants $\lambda_{mn}$. This gives the Bayesian prior density of $f$:

$$
\pi(f) \propto \exp \left[ -\frac{1}{2} \sum_{m,n} \{ (\xi^x_{mn})^2 + (\xi^y_{mn})^2 \} \right].
$$

The likelihood $l(I|f)$ of observing $I$ given a deformation $f$ is constructed by Amit et al. on the assumption that $I(x, y)$ has a normal distribution with mean $DT_f(x, y)$ and variance $\tau^2$ independent for each point $(x, y)$ on a finite grid $G \subset [0, 1]^2$. Thus

$$
l(I|f) \propto \exp \left[ -\frac{1}{2\tau^2} \sum_{(x,y) \in G} \{ I(x, y) - DT_f(x, y) \}^2 \right].
$$

Using a Markov Chain Monte Carlo (MCMC) routine, a sample from the posterior distribution or a point estimate can be obtained. Amit et al. obtain a local maximum a posteriori (MAP) estimate using a gradient descent method, which requires partial derivatives of $T(x, y)$ to be known.

### 1.2 Wavelet Expansions

The standard Fourier basis is generated by dilations of a sine and a cosine function, so each basis function has a specific frequency. A one-dimensional wavelet basis is generated by dilations $j$, and translations $k$, of one function called a mother wavelet $\psi$ (Daubechies, 1992). Each basis function is called a wavelet and is defined by

$$
\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j} - k).
$$

A general one-dimensional function, $f : \mathbb{R} \mapsto \mathbb{R}$, can be expressed as a wavelet expansion

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

for some $w_{j,k} \in \mathbb{R}$. One property of the mother wavelet is that it has short support in the time and frequency domains, hence each coefficient value $d_{j,k}$ contains information about $f$ over a specific frequency and time interval.
For a two-dimensional function, $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, the wavelet expansion is similar to above. Now 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 Equation (2) is

$$f(x, y) = \sum_{\kappa \in \mathcal{K}} w_{\kappa} \Psi_{\kappa}(x, y),$$

where $\kappa$ is an index $\kappa = (j, k_1, k_2, l)$ taking values in the index set $\mathcal{K}$.

A function $f$ that can be expressed using a small number of non-zero coefficients has an economical representation. The localisation property of wavelets mean that a wide class of functions useful in practice, including those that are piecewise smooth with local irregularities or discontinuities, can be well expressed or well approximated by economical wavelet representations. For further reading on wavelet theory see Chui (1992) or Daubechies (1992). Nason & Silverman (1994), and Downie (1997) both discuss statistical methods that have utilised wavelet theory.

## 2 Wavelet Deformation Models

Using a wavelet basis to model a deformation function is a natural extension of the Fourier model. In particular it is reasonable to assume that many deformations will have localised features, which will be represented more economically as a wavelet basis than as a Fourier basis. A wavelet basis is used as a model for the deformation in Amit (1994), where a deterministic approach is used. The norm of the function corresponds to the prior density in the Bayesian setting and is given by $||f||^2 = \sum_\kappa w_{\kappa}^x w_{\kappa}^x + w_{\kappa}^y w_{\kappa}^y$, where $w_{\kappa}^x$ and $w_{\kappa}^y$ are the wavelet coefficients for the $x$ and $y$ components of $f$ respectively. In a probabilistic setting, the distribution of each wavelet coefficient would be $N(0,1)$ and a priori the deformation $f$ would not have an
The data driven term is proportional to the sum of squares error between the deformed template and image, \( \frac{1}{2} \sum_{x,y \in G} \{I(x,y) - DT_f(x,y)\}^2 \). The “optimal” deformation minimises the sum of these two terms. Amit assumes that the image and template are defined on the unit torus, allowing periodic boundary conditions to be used. The method of obtaining the optimal deformation uses a steepest descent method, which again needs the partial derivatives of \( T \) to be known.

Another example of using wavelets to model deformation functions (albeit of a simpler form) is given in Aykroyd & Mardia (1996). They study spinal images where both the data and deformation are expressed in one dimension and each vertebra is a well defined landmark. The prior model depends on the squared difference in the deformation between neighboring vertebrae \( \pi(f) \propto \exp \left[ - \sum_i \{ f(x_{i+1}) - f(x_i) \}^2 \right] \). This prior is defined in the domain of the deformation and not the domain of the wavelet coefficients. For each image and template pair, posterior estimates are obtained using MCMC methods. The variability of each coefficient is observed and the mean of all realisations is used for a point estimate for each wavelet coefficient.

3 A Wavelet Mixture Model

Neither of the models discussed in the previous section yield 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 ensure economy throughout the procedure we model the coefficients \( w_x^\tau \) and \( w_y^\tau \) as independent random variables with the following mixture distribution:

\[
\begin{align*}
   w_x^\tau &= \begin{cases} 
   0 & \text{w.p. } p_j \\
   Z_k & \text{w.p. } 1 - p_j, \text{ \ where } Z_k \sim N(0, \sigma^2_j) \end{cases}
\end{align*}
\]
A coefficient will be exactly zero with probability \( p_j \) else it has a normal distribution with mean zero and variance \( \sigma_j^2 \). The parameters \( p_j \) and \( \sigma_j \) may depend on the resolution level \( j \) of the associated wavelet function. A Bayesian thresholding method for one-dimensional signals has been proposed by Abramovich, Sapatinas & Silverman (1998), which also uses this mixture distribution. Some specific analytical aspects of using this prior are discussed by them. For example, the choice of \( p_j \) and \( \sigma_j \) determines exactly which Besov spaces a deformation belongs to.

4 Femoral Condyle Images

The original idea for developing the above wavelet model arose from the desire to analyse a sample of human femora (thigh bones), taken from an archaeological collection of skeletal remains. This sample of bones was collected as part of a study exploring the relationship between osteoarthritis of the knee and the shape of the distal end of the femur (i.e. the shape of the thigh bone at the knee). It has been speculated by researchers in the field that the shape of the knee may predispose to the development of osteoarthritis (Bullough, 1981) (Cooke et al., 1997). Osteoarthritis is a joint disease characterised by the loss of articular cartilage. In cases where cartilage is lost completely the articular surfaces of the opposing bones of the joint rub together producing areas of smooth, polished bone. This phenomena is known as eburnation and, in skeletal material is diagnostic of late-stage osteoarthritis. Another characteristic feature of osteoarthritis is the development of osteophytes. These are bony outgrowths around the edge of the joint which will cause local shape deformations. Very mature osteophytes are often indistinguishable from the articular surface of the bone. Thus, shape variation in this sample will be both natural and pathological.

The following method was used to capture two-dimensional images of the femora. The
distal end of each bone, viewed axially, was recorded using a video camera (Figure 1). A consistent alignment of each bone, relative to the camera, was used to minimise spurious shape variation in the plane of the two-dimensional projection. Such variation may be due to rotation in the planes perpendicular to the plane of projection, for example. Each bone was allowed to lie naturally upon a horizontal surface, resting on the feet (the most anterior points) of the condyles at the distal end, and upon the lesser and greater trochanters at the proximal end. The articular surface of the condyles was then squared against a vertical plane with the intercondylar notch (the arch in the center) directly in line with the lens (Figure 2). Each video image was subsequently digitised to bitmap images and any areas of eburnation and osteophytes, were highlighted. For the purposes of developing and using the deformable template method, the images were then transformed into 32 by 32 pixel images (Figure 3). One bone, unaffected by any pathology, was arbitrarily designated as the template.

The wavelet coefficients of a deformation function give a multivariate description of the difference between the template and each image. By analysing the results using multivariate methods, it should be possible to quantify the variation in shape within natural or diseased subjects and possibly to form a discrimination rule to identify the two groups. For this to be effective, it is paramount that the difference between the template and the image be described as economically as possible.

5 PLS and Bayesian Optimisation

Ideally we want $f$ to be a good fit for the data and yet be as smooth as possible, but in general these two requirements are contradictory. The penalised least squares (PLS) method obtains a deformation function which compromises between the two (see Green & Silverman (1994) for
Define the penalised least squares score as

\[ S(f, \alpha) = \sum_{(x,y) \in G} \{ (I(x,y) - DT_f(x,y))^2 - \alpha \log \{ \pi(f) \} \}. \]

The roughness penalty \( \log \pi(f) \) is constructed by letting \( \pi(f) \) be the density of the wavelet mixture model (3):

\[ \pi(f) = \prod_{\nu \in \{x,y\}} \prod_{\kappa \in \mathcal{K}} \left\{ p_j \delta_{0, w_k} + (1 - p_j) \phi \left( \frac{w_{\kappa}}{\sigma_j} \right) (1 - \delta_{0, w_k}) \right\}, \]

where \( \phi(\cdot) \) is the standard normal density function. We would like to find a deformation that minimises \( S(f, \alpha) \). The smoothing parameter \( \alpha \geq 0 \) gives the relative importance between the sum of squared errors and the roughness penalty.

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

\[ \pi(f|I) \propto \exp \left[ -\frac{1}{2\tau^2} \sum_{(x,y) \in G} \{ (I(x,y) - DT_f(x,y))^2 + \log \{ \pi(f) \} \} \right]. \]

Thus there is a bijection between \( \pi(f|I) \) and \( S(f, \alpha) \) and so minimising the PLS score is equivalent to finding the deformation that maximises the posterior density. The assumption of independent errors on the grid \( G \) may not always be realistic in practice, for example with the femoral condyle images, where the condyle region region is large and uniform. We therefore regard our approach as a PLS method with a Bayesian motivated roughness penalty, rather than a strict Bayesian method.

6 Implementation

There are many minimisation algorithms that obtain a deformation with a low PLS score.

We have implemented three different algorithms; an ICM algorithm, a simulated annealing algorithm and a greedy algorithm. We give a brief explanation of each algorithm below; for
a more detailed explanation see Downie (1997). Each algorithm is devised to work with any image with $2^n \times 2^n$ pixels. The template and image may be binary, colored or have grey levels, no smoothness assumptions are made and the image may be noisy or noise free. If periodic boundary conditions are used the deformation is theoretically defined on the unit torus and features near one edge of the template will wrap around to the opposite edge. To avoid this problem symmetric boundary conditions are used.

Due to the multiresolution properties of wavelet bases, it is sensible to apply a cascade method to the optimisation routine. See Hurn & Jennison (1995) for a discussion of cascade methods applied to conventional image restoration problems. Because there are only a small number of coefficients at the low frequencies it is efficient to optimise over those low frequencies first to obtain an approximate solution quickly. Then gradually adding (cascading) higher frequency levels to obtain a more refined solution. In addition a more economical expansion is likely to arise by optimising over the low frequencies first.

### 6.1 An ICM Algorithm

The first algorithm applies an iterated co-ordinatewise maximisation (ICM) routine (equivalently called iterated conditional mode), see Besag (1986). This takes each coefficient and sets it to the value that minimises the PLS score, updating the coefficient values and the deformation at each stage. This is repeated for a fixed number of iterations. At the end of each iteration the number of resolution levels that are adapted may cascade. The optimisation procedure in Amit (1994) requires the partial derivatives of the template $T$ to be known. When the template is a black and white or a colored image, the partial derivatives of $T$ need not exist over much of the unit square. Thus, for each coefficient its approximate minimising value is found by naively testing equally spaced values from within a search window. This search window initially has
length $4\sigma_j$ centred on the current coefficient value and is contracted after each iteration.

In order to ensure that a coefficient becomes non-zero only if there is a corresponding decrease in the error term, we impose the condition $p_j \geq (1 - p_j)/\sqrt{2\pi \sigma_j^2}$ on the parameters $p_j$ and $\sigma_j$. This ensures that the probability a coefficient is zero is larger than the density of it taking any non-zero value.

If we were using a Bayesian methodology, this algorithm would converge to the iterated conditional mode (ICM) of the posterior distribution, (conditioned on all coefficients initially equalling zero and the mode is relative to the previously defined mixed dominating measure).

### 6.2 A Simulated Annealing Algorithm

Using the Metropolis Algorithm we can implement a simulated annealing method (see Gilks et al. (1996)). This method simulates coefficients from the mixture distribution given in (3) and accepts the coefficient if the resulting deformation gives a lower PLS score. If a higher PLS score ensues, the coefficient may still be accepted at random with a probability that depends on the ratio of the old and new PLS scores and a temperature schedule. The temperature is a monotonic function decreasing with iteration number. This allows suboptimal coefficients to be chosen with relatively high probability early in the routine but with only very low probability late in the routine. This approach allows the deformation to wander in the high dimensional space of wavelet coefficients, gradually converging to a region with a low PLS score. The temperature is reset to one at the start of each cascade and at the end of each cascade, redundant non-zero coefficients are removed as with the ICM algorithm.
A Greedy algorithm accepts coefficient values that reduce the sum of squares error (SSE) 
\[ \sum_{(x,y) \in G} \{ T(x, y) - DT_f(x, y) \}^2, \] 
irrespective of how rough the deformation is (equivalent to \( \alpha = 0 \)). However, if a range of coefficient values all give the same minimum SSE the value 
closest to zero is chosen.

The coefficients which individually decrease the SSE are found and listed, without updating 
them. Next, update the coefficient found to give the biggest change in SSE. Then coefficient 
found to give the next largest decrease in SSE is updated – provided that a decrease in the SSE 
is still obtained, after inclusion of the previous coefficient – and so on, for each coefficient in 
the list. This process is iterated a fixed number of times. Note that, this method may update 
any number of coefficients in one iteration. Other versions of the greedy algorithm may only 
update one coefficient per iteration, the one that gives the biggest change in SSE.

For the ICM and Simulated Annealing methods the smoothing parameter is integral to the 
program and needs to be specified. But, for the Greedy Algorithm, it is not necessary to run 
the whole algorithm again to obtain a smoother deformation. To smooth the deformation, some 
of the non-zero wavelet coefficients may be reset to zero. Either a fixed number of coefficients 
could be removed or the SSE could be constrained to remain below a specified amount. In 
either case the coefficients that give the smallest change in SSE will be removed first.

6.4 Types of Wavelet

Two different wavelets were modelled, the Haar Wavelet and the Daubechies D4 Wavelet (Daubechies, 

In addition to expressing \( f_x \) and \( f_y \) as a wavelet expansion. We also consider \( \partial f_x / \partial x \) and
\[ f_x(x, y) = \sum_{\kappa} w^{x}_{\kappa} \int \Psi_{\kappa}(x, y) \, dx \quad \text{and} \quad f_y(x, y) = \sum_{\kappa} w^{y}_{\kappa} \int \Psi_{\kappa}(x, y) \, dy. \]

The basis functions are integrated two dimensional Haar wavelet functions, and so we call this basis the *Integrated Haar* basis. This basis is a two dimensional linear spline basis. Each basis function is piecewise linear with very short support, however they are not wavelets as they don’t integrate to zero (either w.r.t. \( x \) or \( y \)) and the basis is not orthogonal. When using the Haar and *Integrated Haar* bases no “additional” coefficients are required for the symmetric boundary conditions. This is because no wavelet functions ever cross the boundary of the unit square.

### 6.5 Visualisation

The aims of the deformable template method should not merely be to warp the image to the template, but also to convey some 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.

We use two conventional methods of visualising the deformation function and we present a new method specific to the wavelet representation. Figure 4 shows the deformation function itself, superimposed on the template. The arrowheads are the grid points \((x_s, y_t)\) and the arrow tails are the points \(f(x_s, y_t) + (x_s, y_t)\). The value of the deformed template at the position of the arrowhead is the value of the template underneath each arrow tail. Figure 5 is concerned with the efficacy of the deformation in fitting the image to the template, and shows the template, the image, the deformed template and the *overlay*. The overlay shows where the deformed template and image differ. The colored parts denote the pixels where the SSE is non-zero.

The third type of plot, as in Figure 6, shows the position of the non-zero wavelet coefficients and hence the localisation of the deformation. The two components \((x \text{ and } y)\) are displayed on
two separate diagrams. Each number appears at the center of the relevant wavelet’s support, the number represents the resolution level (higher numbers correspond to wavelets with smaller support) and the size represents the magnitude of the coefficient. In addition the numbers may be colored to correspond to the orientation of the wavelet, as in Downie (1997). The salient features and interpretation of these figures are discussed in the following section.

7 Results

In order to demonstrate the method, we applied the above three algorithms to just two of the femoral condyle images. These data can be obtained from the web page http://www.stats.bris.ac.uk/~trd/fcddata.html. The resolution of both the image and the template was $32 \times 32$ pixels and the deformation was evaluated at $32 \times 32$ points. This requires five wavelet resolution levels; $j = 0$ (low frequency) to $j = 4$ (high frequency). The algorithms were compared and the effects of fitting different parameter values and wavelet bases above were considered.

The speed of each algorithm does not depend on the resolution of the image and template, so the coarse resolution of this image and template has little effect on the results obtained. However the resolution of the deformation does affect the algorithm run times (if the deformation is evaluated at $n^2$ points, then the algorithm is $O(n^2)$). Since the aim is to get an economical description of the difference in shape between the image and template, it is unlikely that evaluating the deformation at a large number of points would give a worthwhile improvement. We claim that a small grid for the deformation should be sufficient for many datasets.

The total number of wavelet coefficients required to define the deformation depends on the type of wavelet fitted. For the D4 wavelet 2700 coefficients were required and 2048 coefficients defined deformations using the Haar or Integrated Haar Basis. We shall use the phrase “number
of coefficients” as an abbreviation for the number of non-zero wavelet coefficients for the relevant deformation. The algorithms were written in C and compiled and run on a Sun UltraSparc. For Simulated Annealing the Wichmann & Hill random number generator was used with a different seed for each deformation (Wichmann & Hill, 1982).

7.1 ICM Algorithm

Table 1 shows the results from the ICM algorithm for different values of \( \alpha \). Convergence always occurred within seven iterations and the mean CPU time taken to run this algorithm was 8 minutes 45 seconds. As \( \alpha \) increases, the SSE increases but fewer coefficients are used to define \( f \). When \( \alpha \leq 0.5 \) the algorithm finds an exact match. Notice that when \( \alpha = 1.25 \) the SSE is less than when \( \alpha = 1.0 \), contrary to what one would expect. This demonstrates that the algorithm does not necessarily obtain the global minimum of \( S(f, \alpha) \), but merely finds a local minimum. Furthermore, the deformation obtained for large \( \alpha \) is not simply the deformation for a smaller \( \alpha \) with some of the non-zero coefficients shrunk towards zero. Figures 4, 5 and 6 show the deformation function, the deformed template and the coefficient positions respectively for \( \alpha = 1.5 \). Note that in Figure 6 most of the coefficients are near the condyle outline, which demonstrates the localisation properties of the wavelet basis. At the top of the condyle in the \( y \) component, coefficients from resolution levels one to four are present, indicating that both broad scale and detailed warping is required, and this is dominated by two coefficients. Figures 7 and 8 show the deformation and coefficients for \( \alpha = 0.5 \). Comparing Figure 8 with Figure 6, Again most of the coefficients lie around the outline but for smaller \( \alpha \) there are more coefficients. Notice that there are some coefficients in Figure 6 not present in Figure 8, confirming that the deformation when \( \alpha = 1.5 \) is not merely a smoother version of the deformation obtained when \( \alpha = 0.5 \).
7.2 Simulated Annealing

The results obtained when using the Simulated Annealing method with 350 iterations are given in Table 2. Again, when $\alpha$ increases, fewer coefficients are fitted and the SSE increases, but Simulated Annealing fits more coefficients than ICM does for the same SSE. The most noticeable difference between the two methods was the time taken to run the algorithm. The mean time to run 350 iterations was nearly 33 minutes.

We investigated the effect of varying the number of iterations in the Simulated Annealing algorithm. There was some evidence that the number of coefficients required to obtain a similar SSE falls with an increased number of iterations. However, since this was by no means conclusive, the increased CPU time required did not seem warranted.

7.3 Greedy Algorithm

Using the Greedy Algorithm the minimum SSE obtained was 2. This deformation was defined by 36 non-zero coefficients and took nearly nine minutes to run. We were unable to get a perfect match between the two femoral condyle images when using any version of the Greedy Algorithm. The Greedy smoothing routine typically takes 5 seconds to run, so once the algorithm has run on one image and template pair, the preferred smoothness can be obtained quickly. The deformation which has been smoothed to have an SSE of five is shown in Figure 9. This deformation looks smoother than the ICM deformation in Figure 4, even though the SSE is smaller.

7.4 Wavelet and Parameter Choice

Using ICM with the Haar Wavelet the following results were obtained. When $\alpha = 0.5$ the SSE was three, fitting 34 coefficients and taking nearly nine minutes to run. For $\alpha = 1.25$
the SSE obtained was 17, using 19 coefficients and taking thirteen minutes to run. Comparing these results with Table 1 (for the same $\alpha$), one can see that the Haar Wavelet gave a worse deformation than the Integrated Haar Basis in terms of both SSE and the number of coefficients. Running the Greedy Algorithm with the Haar Wavelet, a deformation giving an SSE of five with 36 coefficients was obtained. This is worse than using the Integrated Haar Basis, but the program only took three minutes thirty-five seconds to run. The Greedy Algorithm using the Daubechies Extremal Phase wavelet with two vanishing moments gave a lower SSE than the Haar Wavelet. The SSE was three using 41 coefficients and taking three minutes twenty-two seconds to run. Of the three wavelets fitted, the Integrated Haar Basis appears to give better fitting deformations using fewer coefficients. However, obtaining a deformation from an Integrated Haar Basis expression is numerically more intensive than using a wavelet basis, thus making the time to run the algorithms longer.

The parameter $p$ corresponds to the cost of a coefficient being non-zero. Thus the larger the value of $p$, the less non-zero coefficients one would expect to see in the resulting deformation. So the effect of increasing $p$ is similar to the effect of increasing $\alpha$. The variance of the non-zero wavelet coefficients is $\sigma^2$. The optimal value of $\sigma^2$ will be large enough to allow the wavelets to define a good deformation, but not so large that convergence to the right value is slow. Too large a variance will also over penalise all non-zero coefficients. Changing the value of $\sigma^2$ affects the resulting deformation. The effect is non-linear and the optimal value for $\sigma^2$ will depend on each particular template and image pair.

8 Conclusions

We have demonstrated that modelling a deformation function using a wavelet basis, where the roughness penalty depends on a mixture model can give an economical representation for a
deformation function. We implemented three methods of minimising a penalised least squares score making no assumptions about the form of the image and template. Motivated by the desire to analyse a particular data set of femoral condyle images, the algorithms were compared using two of these images.

The ICM method was best, with the Greedy Algorithm giving only slightly worse results. The ICM method is acceptably quick and gives a good fit with a low number of coefficients. The method, however, does not necessarily attain the global PLS minimum. The Greedy Algorithm also gives an acceptable deformation with a similar number of coefficients for a given SSE compared to a ICM deformation. However, an exact fit was not obtained using this method. The advantage of using the Greedy Algorithm is that deformations with different smoothness are quickly obtained once the main program has run. The Simulated Annealing algorithm is worse than the other methods in terms of the number of coefficients fitted for the same SSE and took much longer to run. This is probably because most good deformations, when expressed in the wavelet domain, are local to the zero deformation. Thus allowing the deformation to “wander around” in a high dimensional vector space achieves very little.

The Integrated Haar Basis was better than the Haar and D4 wavelets, both in terms of minimising the squared error and for parsimony of coefficients. The Integrated Haar “wavelet” has support of [0, 1] and two vanishing moments. The fact that the Integrated Haar Basis is non-orthogonal and each “wavelet” has a non-zero integral is of little importance in this setting. A practical problem with all three algorithms was that the deformation obtained was highly dependent on the model and program parameters. Until more detailed investigation any practical use of this method would require the parameters to be chosen subjectively.

We introduce a means of visualising the wavelet coefficients for a given deformation, which can be used as an aid to interpreting the deformation. The full implications of this type of plot remain a subject for future research; in some contexts the coefficients and their values will
be candidates for subsequent statistical analysis, while elsewhere they will be valuable for the
intuition that they give into the position and scale of important aspects of the deformation.

Acknowledgements

This work was undertaken as part of Tim Downie’s PhD project at the Department of Math-
ematics, University of Bristol and funded by the Engineering and Physical Sciences Research
Council, U. K.

References

Abramovich, F., Sapatinas, T., & Silverman, B.W. 1998. Wavelet Thresholding Via a Bayesian


140 of: Mardia, K.V., Gill, C.A., & Dryden, I.L. (eds), Image Fusion and Shape Variability


Bullough, P.G. 1981. The Geometry of Diarthrodial Joints, its Physiologic Maintenance and the


Table 1: Using the ICM algorithm with different values of $\alpha$.

<table>
<thead>
<tr>
<th>Smoothing parameter $\alpha$</th>
<th>SSE</th>
<th># coeffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\infty$</td>
<td>53</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td>1.25</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>1.0</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>37</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 2: Simulated Annealing with different values of $\alpha$.

<table>
<thead>
<tr>
<th>Smoothing parameter $\alpha$</th>
<th>SSE</th>
<th># coeffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\infty$</td>
<td>53</td>
<td>0</td>
</tr>
<tr>
<td>2.0</td>
<td>17</td>
<td>30</td>
</tr>
<tr>
<td>1.5</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>1.25</td>
<td>12</td>
<td>44</td>
</tr>
<tr>
<td>1.0</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>0.5</td>
<td>2</td>
<td>62</td>
</tr>
<tr>
<td>0.25</td>
<td>0</td>
<td>64</td>
</tr>
</tbody>
</table>

Figure 1: Femoral condyles from the distal end of the femur.
Figure 2: Diagram of experimental setup.

Figure 3: Digitised Femoral Condyle Images.

Top row: Original images for a normal condyle (left) and a diseased condyle (right), shaded regions denote eburnation. Second row: coarse 32 x 32 pixel images of two normal condyles, used for developing the algorithms.
Figure 4: The deformation function using ICM with $\alpha = 1.5$.

Figure 5: The femoral condyle template and image, the deformed template and overlay using ICM with $\alpha = 1.5$. 
Figure 6: The wavelet coefficient positions using ICM ($\alpha = 1.5$).

Figure 7: The deformation function using ICM with $\alpha = 0.5$.

Figure 8: The wavelet coefficient positions using ICM ($\alpha = 0.5$).
Figure 9: The deformation function obtained using the Greedy algorithm and smoothed to have an SSE of five.