PRACTICAL METHODOLOGY FOR INCLUSION OF MODALITY-SPECIFIC MODIFICATIONS IN A HIERARCHICAL BAYESIAN DEFORMATION MODEL

by

Jacob Kooter Laading

Institute for Statistics and Decision Sciences
Duke University

Date: ______________________________
Approved:

____
Valen E. Johnson, Supervisor

____
Carey E. Floyd Jr.

____
David Higdon

____
Brani Vidakovic

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Institute for Statistics and Decision Sciences
in the Graduate School of
Duke University

1999

ABSTRACT

(Statistics)

PRACTICAL METHODOLOGY FOR INCLUSION OF MODALITY-SPECIFIC MODIFICATIONS IN A HIERARCHICAL BAYESIAN DEFORMATION MODEL

by

Jacob Kooter Laading

Institute for Statistics and Decision Sciences
Duke University

Date: ______________________________
Approved:

____
Valen E. Johnson, Supervisor

____
Carey E. Floyd Jr.

____
David Higdon

____
Brani Vidakovic

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the
Institute for Statistics and Decision Sciences in the Graduate School of
Duke University

1999

Abstract

An approach is presented which allows the incorporation of application-specific modifications to a general hierarchical shape deformation model. The general methodology models the perception of labeled points, or facets, across an image class through a joint distribution on facet position and image feature value. The modification introduces a set of parameters which represents the relative overall size of an image scene directly into the statistical model for the deformation. Through this, a more flexible and descriptive model is achieved without introducing an unmanageable computational burden. The methods are applied to and the modifications based on the application cardiac gated single photon emission computed tomography (SPECT). For this modality, a contraction factor and a center of contraction have real physical significance and are therefore included in the modeling. Results consistent with known heart behavior are seen for these quantities as well as for clinical quantities derived from the estimation results. A meaningful representation of the timeseries set of data is shown and improved results over traditional methods are seen. The method is also tested on two phantom datasets, one clinical and one mathematical, in order to quantify the ability of the method to track shapes and individual points, and good correspondence with known truth is seen. The methodology as described in detail is useful for any situation where automatic scaleability is useful. It also offers an instructive example of how the general hierarchical shape deformation model can be extended to incorporate application-specific information within the natural structure of the statistical model.

Acknowledgements

First I would like to thank my advisor and mentor Valen Johnson, who have taught me so much about statistics and life. I thank Carey Floyd for bringing me to Duke and sticking with me throughout. Thanks to David Higdon for his enthusiastic participation on my committee and interesting insights, and to Brani Vidakovic for stepping in at the last minute and participating fully.

Colin McCulloch; thanks for the insights, the collaboration and the coffee breaks.

Ron Jaszczak, David Gilland, Jörg Peter and Kim Greer;
thanks for the data and the insightful comments.

My friends and roommates Erik, Rene, Sarah and Scott;
thanks for being great friends, putting up with me and keeping me somewhat sane.

All my friends at ISDS; I will miss you dearly.

My friends in BME and Radiology, past and present ; thanks for the company.

My parents, who gave me what I needed within to get here;
I dedicate this thesis to you.

My wife Marianne, who has stood by me lovingly through all these years, through good times and bad; **ærte**.

My son Ole; you provided the inspiration to press on, no matter what.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 General Overview

The utility of statistical models in imaging has been established over the past decades. Recently, methods which utilize deformable structures have made great advances in the sub-field which can be labeled image understanding. This area is distinguished from the related fields of image restoration, reconstruction and enhancement through the notion that a more conceptual structure relating to the image information is incorporated into the modeling. For instance, most restoration models are based on either physical image generation models or models which treat the image information as signal in a noisy background, with the goal to remove the noise. In image understanding methods, the image is typically taken as is, and the focus is on the perception of the image, that is, how can the image be analyzed so as to maximize the utility of the information we derive from the data. In that sense, the field is closely related to computer vision and other pattern recognition methodology. Note that image processing methods of the enhancement, restoration and reconstruction methods can be used as integral parts of or pre-processing for an image understanding method. A skeleton example of how an image understanding method might ideally work is to imagine an image class for which an atlas is defined. If a method existed which automatically could map one part of the image scene to the same part in the atlas, all operations performed on the atlas could be automatically transferred to the new realization. This could include segmentation, volume measurements or descriptive summaries of object shape deformation.

The applications for such methods are numerous; a short list includes automatic segmentation of medical images such as brains for diagnostic purposes, shape analysis for industrial quality control and image-based person recognition systems. A particular type of deformation problem, treated in detail here, is that of images containing a time-varying structure with a well-known large-scale behavior.

The imaging modality gated cardiac single photon emission computed tomography (gated cardiac SPECT) consists of a sequence of images which, taken together, provide both strong evidence of where in the heart tissue might or might not be functioning properly and also rich data about the motion of the heart, both in global (overall shape) and local (where did certain parts of tissue move?) terms. To use this data properly, however, the motion of points in and around the heart needs to be estimated, and that gives the deformation models described in this thesis real utility for this imaging problem. Because several things are known about the expected motion of the heart, information can be incorporated into the model directly. It is also of vital importance that the complexity of the model is kept at a level which is computationally feasible, such that the potential for clinical implementation is maintained. The methodology applied grows naturally out of previous work in spatial statistics, computer vision and medical imaging as described below, and is based on the hierarchical deformation modeling as described in McCulloch (1998).

1.2 Thesis Organization

Related methods and historical background will be discussed in Chapter 2 together with an exposition of the medical modality from which our image data has been provided. Following this, our methodology is described and developed in detail. First, the baseline model is defined, followed by the modifications made to accomodate the current application. The computational techniques used to maximize and sample the resulting predictive probability distribution are then described, after which a few measures of performance are explained. In the following chapter, Chapter 4, the particular datasets the methodology was applied to are described. This is followed immediately by the results, organized as qualitative followed by quantitative. The former dominates in volume, because of the impossibility of obtaining truth regarding real patient data. Since the patient is alive, the true motion of heart tissue can not be physically ascertained through markers or direct measurements. More quantitative analysis is performed, however, on two phantom datasets for which some truth is known, and these results immediately follow the patient data. Finally, results are summarized and discussed, and future extensions of the methodology are proposed.

Chapter 2

BACKGROUND

The field of image analysis as applied to medical imaging has grown out of several avenues of research. Most notable and relevant to the work presented here are the fields of computer vision research and statistical image analysis. While this methodology clearly falls in the latter category, it has enough similarities to computer vision methods and has borrowed some particular ideas from that field which makes a description and comparison appropriate.

2.1 Computer Vision Methods

These methods are fundamentally based on the notion of making a computer capable of seeing; that is, perceiving an image scene in a sense similar to that of a human. Human vision research by Young (1986) indicates that the human eye in many aspects acts as a Gaussian filter bank, and thus the idea of a hierarchical Gaussian convolution pyramid gained acceptance as a viable candidate for human vision modeling. Particularly, the field of scale-space analysis grew out of this insight.

2.1.1 Scale-Space

The concept of scale-space is based on the idea that we perceive objects and image scenes at different scales or resolutions. (See Lindeberg (1994) for a full description.) Thus, it makes sense to blur out irrelevant details when considering an image at a certain scale. To achieve this algorithmically, let $I(x;t)$ be the $(n+1)$-dimensional extension of an n-dimensional image created by convolving the original image $I(x;t=0) \equiv I(x)$ with a Gaussian kernel of variance t^2, $G(0;t) \propto e^{-\frac{x^2}{2t^2}}$. Then, define the scale-space as

$$I(x;t) = I(x) * G(0;t) = \int_{y=-\infty}^{\infty} I(y)G(x-y;t)dy \quad ,t > 0 \; . \tag{2.1}$$

Under this definition, then, we can think of objects in the image as having inherent scales associated with them. To see this, consider the spatial Laplacian

$$L_{ii}^{I}(x;t) = \sum_{i=1}^{n} \left(\frac{\delta^2}{\delta x_i^2} \right) I(x;t) \tag{2.2}$$

across scales. This quantity has been shown to relate to “middleness” (ter Haar Romeny *et al.*, 1991) in an image or object; that is, a local maximum or minimum corresponds to a center for regions of high or low intensity, respectively. For example, in a face, the nose exists at a lower scale than the face as a whole, and when tracing the scale-space Laplacian L_{ii} this will give rise to a local maximum at the center of the nose at some scale t_n and to another local maximum for the face as a whole at some scale $t_f > t_n$ (for the scale-space of a facial image, see Figure 2.1).

Many other scale-space quantities of interest exist and have been studied (for an excellent exposition, see ter Haar Romeny (1991)), for example boundariness, cornerness and higher-order invariants. The scale-space concept has been the subject or tool of choice for many relevant image analysis methods in computer vision. An example of a simple and direct application is seen in Lifshitz and Pizer (1990), where

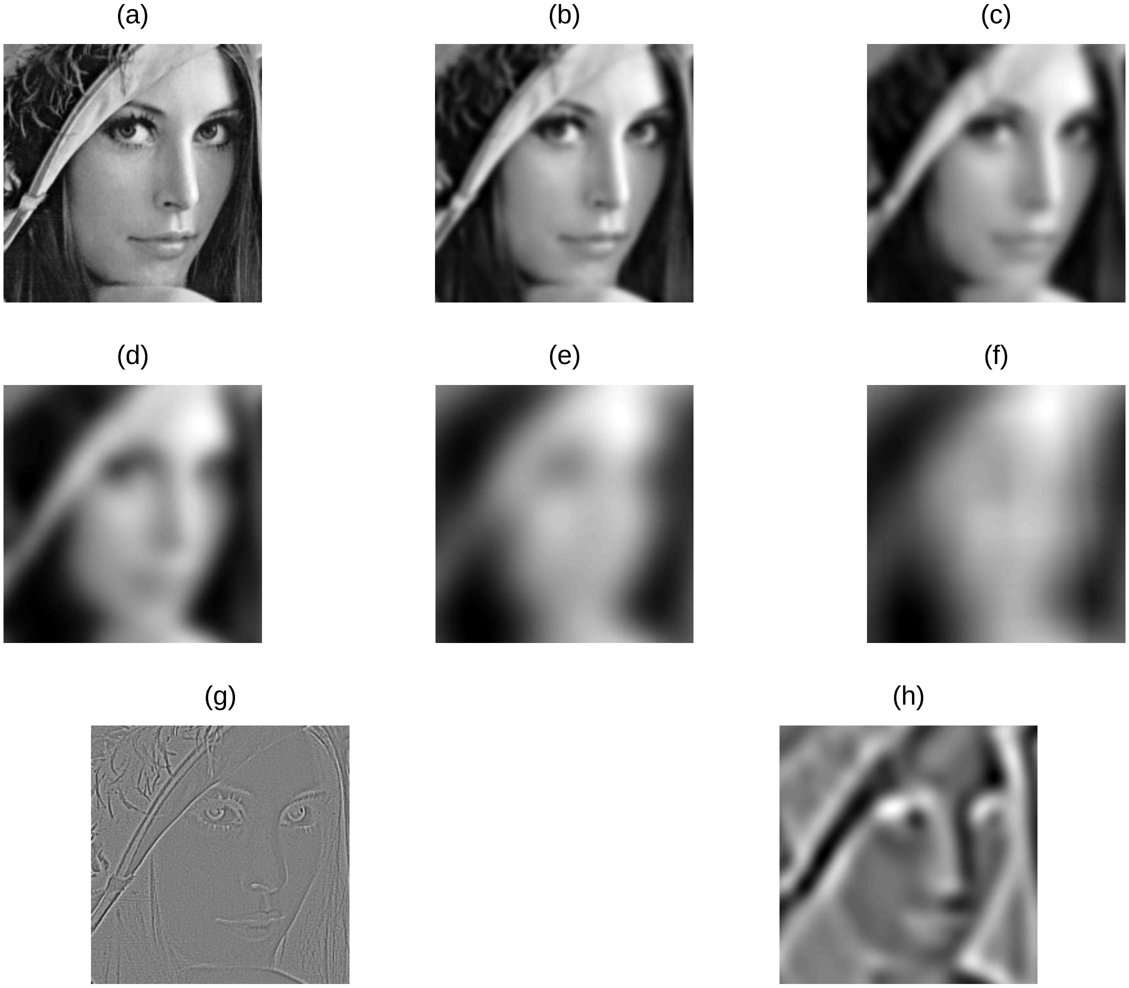

Figure 2.1: An image of a face observed at several scales. From left to right, top to bottom, scale t (in pixels) is (a) original (b) 2 (c) 4 (d) 8 (e) 16 (f) 32 . In addition, the Laplacian at scale (g) 0 and (h) 8 is shown. Here, the mouth has an associated scale of approximately 4, the eyes nad nose are dominant at scale approximately 8, and the face as a whole is associated with scale approximately 32. We also see how the zero-scale Laplacian acts as a simple edge detector and how the higher-scale Laplacian operator focuses in on “blobs” of high and low intensity in the image.

flowlines in scale-space (isointensities of Laplacian, for instance) are traced spatially through scale towards zero (the original image) and used as a basis for segmenting regions of interest in the images.

More relevant to our methodology, the scale-space of an image has been used for hierarchical maximization of objective functions ranging from individual image region segmentation (encourage regions of smoothness within boundaries) to atlas mapping methods (encourage similarities between matching anatomical locations on the two images). These methodologies typically utilize a model structure which in physical terms can be thought of as an external force (the image data) being weighed against an internal force (model-imposed structure or regularization).

2.1.2 Multiscale Registration Algorithm – ANIMAL

A very good example of atlas mapping methodology is seen in the ANIMAL algorithm developed by Collins et al (1995). They utilize a hierarchy of grids and the scale-space of an image to perform a constrained maximization which matches images to an atlas for the relevant image modality. This is done by defining the objective function

$$O(x, y; t) \propto \sum_{(x' \in C_x, y' \in C_y)} A(x'; t) I(y'; t) \tag{2.3}$$

where x is the coordinate in the atlas A, y is the corresponding grid position in the image under consideration I, and t the current scale for grid maximization. C_x is a neighborhood of predefined extent about the point x. For each level in the hierarchy this correlation measure between local features is maximized with respect to grid deformation (location of grid points in new image) site by site much like in the statistically based iterated conditional modes algorithm (ICM) ((Besag, 1974); see Section 2.2 for a more complete explanation). The maximization is performed subject to regularization constraints, such that the full deformation indicated by the objective function O is relaxed by a factor α towards the mean deformation in the neighborhood of a site on the grid,

$$d' = \alpha d + (1-\alpha)\frac{1}{N_C}\sum_C d \ . \tag{2.4}$$

Here, d is the full estimated deformation vector, d' the regularized value, C a set of neighboring site for the d being considered and N_C the number of neighbors. After completing maximization for a level, the resulting deformation estimates d' are then used as a starting point for the next finer resolution level. The process is repeated until heuristic convergence is seen. These methods have produced good results for human brain mapping (3-dimensional magnetic resonance imaging) in the context of automatic atlas-based segmentation.

2.1.3 Active Contour Models

Another approach to segmentation applied both with and without the scale-space framework, as well as being an example of external-internal force maximization is the so-called active contour approach used for example by Kass et al (1988) and McEachen and Duncan (1997). This methodology uses the concept of fitting a smooth contour to a boundary in an image by maximizing the perpendicular gradient or a similar measure along the contour while defining a regularity condition in the form of a smoothness constraint on the curve. One such constraint would be to impose smoothness in the directional derivative along the contour. Using a two-dimensional example (Kass *et al.*, 1988), consider the maximization of the functional

$$E(v) = S(v) + P(v) \ , \tag{2.5}$$

of a contour $v(s) = (x(s), y(s))$ where the internal energy term S is defined as

$$S(v) = \int_0^1 \{w_1(s)\|\frac{\delta v}{\delta s}\|^2 + w_2(s)\|\frac{\delta^2 v}{\delta s^2}\|^2 ds\} \ , \tag{2.6}$$

and the external energy term P is taken as

$$P(v) = \int_0^1 P(v(s))ds, P(x, y) = -c\|\nabla I(x, y; t)\| \ . \tag{2.7}$$

Here, ∇ is the 2-dimensional gradient operator. This particular setup would encourage the "snake" to find boundaries, with the internal energy term encouraging the curve to bend smoothly. This approach can be extended to 3 dimensions by using similarly defined deformable surfaces, see for example Clarysse et al (1997) or Park et al (1996). This has produced interesting results, particularly in situations where there are relatively clear boundaries to be found (e.g. brain ventricles).

2.2 Statistical Image Analysis Methods

Since the method proposed here is statistical in nature, it is reasonable to start with the beginning of statistical image analysis, mentioning some of the key developments, before moving on to the specific areas particularly relevant to the methodology at hand.

2.2.1 Origins and Markov Random Fields

The current field of statistical image analysis owes much to the work of Julian Besag who first presented the idea of specifying statistical lattice models conditionally (Besag, 1974). Through this, he outlined the necessary ingredients for the Markov Random Field (MRF) (Dobrushin, 1968) as applied to statistical imaging and related subsequent developments. This involves the realization, through the Hammersley-Clifford theorem[1], that given appropriate constraints, a valid model can be specified via potential functions $V(x_1, \ldots, x_n)$ so that the full conditional distributions are completely specified by a set of functions which each depend on only a limited subset

---

[1] named for an unpublished 1971 paper by Hammersley and Clifford

of variables in the model. Referring to Besag (1974) for the proof itself, the joint specification on variable vector $x = (x_1, \dots, x_n)$ given parameters θ, expressed as

$$p(x_1, \dots, x_n|\theta) = Z^{-1}(\theta) \exp \left\{ \sum_{m=1}^{M} V_m(x_{i_m}, \theta) \right\} , \tag{2.8}$$

will give rise to a valid joint distribution as long as certain conditions, given in the paper (Besag, 1974), are met. Here, the V_m are a limited number of potential functions which depend only on interactions between the variables $x_{i_m} \subset x$, and Z is the partition function, also known as the normalizing constant, which is independent of x. M is the number of cliques in the model, that is, the total number of interaction terms specified between the various elements of x.

This representation enables efficient computation in situations where joint specification would be far too complex or computationally intensive to implement. In this representation, each variable can be updated or maximized based only on a subset of the other variables in the model. Marginally, of course, there can be significant correlations between all variables. These correlations can, in fact, in some cases seriously affect the computational performance when using the MRF conditional specification (slow exploration of the supported parameter space).

2.2.2 Implementation

Besag also outlined many practical aspects of implementing MRFs, such as the method of Iterated Conditional Modes (ICM) (Besag, 1986), in which the value of each variable is set iteratively to the value which gives a maximum of its full conditional distribution. This algorithm is guaranteed to converge to at least a local maximum of the joint distribution on x. Updating is done according to the iterative setup

$$\{\forall i \in I : \max_{x_i} p(x_i|x_{-i}) = \max_{x_i} p(x_i|x_{c_i})\}, \tag{2.9}$$

where i is the index of the current variable or variables under consideration and I is the full set of indices in the model. The variable x_{-i} is the full set of variables except for x_i, and x_{c_i} denotes the subset of x which is in the neighborhood of x_i, that is, which has interactions with x_i directly through the model specification, be it MRF or otherwise. The maximization is cycled several times through all the variables until convergence is obtained.

To ensure global convergence, other methods such as simulated annealing (Kirkpatrick *et al.*, 1982) must be used. In simulated annealing, a temperature parameter T is defined which expresses a smoothing applied to the probability distributions. This terminology is drawn from thermal physics, where such a parameter is used to characterize ensemble particle motion (statistical mechanics). In this context, we maximize $(p(x))^{1/T}$ rather than $p(x)$ iteratively over x, and an outer loop is included in the algorithm in which T is allowed to cool towards zero. Given a slow enough schedule of cooling, this converges surely to the global maximum of p, but there is no guarantee of convergence in a reasonable amount of time. In many problems, therefore, estimating a “good” local maximum of the joint distribution is the only option available due to the complexity of the problem. Simulated annealing was first proposed for the imaging problem by Geman and Geman (1984). In the same paper, they proposed the Gibbs sampler as a tool of choice for updating and maximization in lattice problems, cast as an alternative or complement to the already popular Metropolis algorithm.

The Metropolis algorithm or its extension, the Metropolis-Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970), works on the principle that we obtain samples y from a true distribution $\pi(y)$ by generating candidates from a distribution $p(y)$ which is easy to sample from (e.g. the normal) and then accepting or rejecting that sample with probability q (for symmetric proposal distributions, $p(y_i|y_j) = p(y_j|y_i)$)

$$q = \min\left\{1, \frac{\pi(y_{i+1})}{\pi(y_i)}\right\} . \tag{2.10}$$

For asymmetric proposals p (i.e. $p(y_i|y_j) \neq p(y_j|y_i)$), this is modified slightly via a so-called Hastings correction, to

$$q = \min\left\{1, \frac{\pi(y_{i+1})p(y_i|y_{i+1})}{\pi(y_i)p(y_{i+1}|y_i)}\right\} . \tag{2.11}$$

The Gibbs sampler can be seen as a special case of this, where you know and can sample directly from π, such that $p(y_i) = \pi(y_i)$, which, as is intuitively obvious, leads to always accepting the proposed value. Note that $\pi(y_i)$ can depend on other variables in the model (conditioning suppressed for compactness when writing out equations), such that the benefit of the conditional specification of the MRF still can be taken advantage of fully. The Gibbs sampler is thus just the sequential updating of the variables in the model directly using known full conditional distributions from which we are able to generate samples.

From the above, it is obvious that the computational efficiency of the application of these methods depends critically on the form of the distributions, whether the joint or full conditional form is being used. In that context, the idea of data augmentation, as described well in the general modeling sense in Tanner (1994), has become important in fields like medical imaging. The idea is to introduce underlying unobservables in order to make the model more straightforward in implementation and statistical structure, if somewhat less compact. For example, the relationship between underlying image formation parameters (attenuation, mean radiation) and the observed image recorded can be quite complex. If an intermediary set of parameters representing the radiation from a certain position directed towards a particular image location is introduced, then the resulting model structure can be fit into a more standard analysis and implementation becomes much more straightforward.

2.2.3 Classes of Imaging Problems

A clear distinction should be drawn here between classes of imaging-related problems and the corresponding motivation and models. Early on in the field, most work was performed in image enhancement or restoration. In these problems, the task was to modify the intensities of the original image to make the pictures clearer or make certain features stand out. This typically involved a physical model for the image likelihood $p(y|x)$, where y is the image data and x is the true parameter describing the image. Priors $p(x)$ are then defined in an MRF representation to enforce local noise contraints of the type given by the potential function (referring back to equation (2.8))

$$V(x_i, x_j) = -(x_i - x_j)^2 \ , \tag{2.12}$$

where x_i and x_j are adjoining pixels in a 2-dimensional image, for instance. The posterior estimate of x could then be obtained from the product of the two through Bayes rule, $p(x|y) \propto p(y|x)p(x)$. More complex potentials V have been developed, of course, but the spirit remains the same; to use the MRF structure to regularize the image scene towards behavior which we expect a priori.

There has also been much statistical work in image reconstruction, a field which concerns itself with the creation of an n-dimensional image from $(n-1)$-dimensional projections. Except for the more complicated physical model (the compilation of reconstructions from the projections to obtain the image scene), this field was also dominated by incorporation of physical image formation models and noise-constraining regularization. A good example of recent methodology cast in the Bayesian framework can be found in Higdon et al (1996).

Extensions and further work in such statistical image analysis are too numerous to fully describe here. One of many excellent reviews is available in Besag et al (1995). One particularly notable extension is that suggested in the previously mentioned paper by Geman and Geman (1984). For purposes of segmentation, that is, to allow sudden changes in intensities as correspond to region boundaries, they introduce so-called line-sites which exist between pixel sites and determine whether a site is a neighbor (line-site set to 0) of its nearest sites or not (line-site value 1). This gives rise to a somewhat complex distribution with normalizing and stability problems. Johnson et al (1991) proposed a modified version of this approach with continuous-valued line-sites (line-sites valued on the interval $[0, 1]$).

Johnson (1994) also later introduced a model for segmentation based on hierarchical thinking. Priors for image regions were specified hierarchically through potential functions defined on the entire graph (all sites) as a function of the number of distinct regions. This was done to penalize a high number of regions, thus eliminating spurious region formation. In addition, a more local portion of the prior was defined on cliques formed as "rings" around a given site at a certain selected radius. This portion was designed to discourage breaks in labels around that ring, thus encouraging regions of size at least the diameter of the ring. By varying this parameter, the types of regions being encouraged varies accordingly, from large smoothly outlined areas to smaller intertwined fingers of intensity. The use of a hexagonal grid for purposes of space-filling and symmetry reasons was utilized and the model thus defined was applied to real and simulated examples with interesting results in terms of capturing different types of image structures (large regions with smooth boundaries versus intertwined snake-like patterns).

All of the statistical methods thus far described can be viewed as pixel-based methods, that is, they focus on changing or classifying the pixel values of an image to achieve an objective in terms of image quality. The latter examples, however, point in the direction of representing images in the modeling sense to represent meaningful conceptual things about the image. The linesite model can be thought of as segmenting images into regions by finding boundaries, and the hierarchical region modeling can be thought of as deciding what kind of information the image contains (how many regions, what type of regions). This type of thinking has lead to the more recent and complimentary discipline of image understanding, models where the emphasis, regardless of image type or inherent noise structure, is on modeling the image scene in such a way as to maximize the conceptual and practical information content extracted from the image.

2.2.4 Curve-Based Image Understanding Methods

This latter methodology often includes reducing the image to a set of salient features, for example a set of curves representing a shape in the image. This was proposed for facial image analysis by Phillips and Smith (1994), who also cast it in a Bayesian framework. They defined a model in which such curves, defined by a small number of parameters θ, uniquely identified a pixel i as being in one of n regions in the face, giving it a label x_i. They then defined a likelihood $p(y|x)$ as

$$p(y|x, \tau, \mu) \propto \exp -\frac{1}{2} \sum_{i \in I} \tau_{x_i}(y_i - \mu_{x_i}) \ , \tag{2.13}$$

where y_i is the pixel value at location i and I is the full set of pixel sites.

Thus, the likelihood is based directly on the uniformity of intensity within curve-enclosed regions. Vague priors on the parameters θ which determine the curve shapes (e.g. ellipse size or breakpoints) were defined. Also, a set of hierarchical constraint were introduced which enforced the notion that various parts of the face would be correctly positioned relative to each other, such as the eye inside the face outline, the pupil within the eye, etc. Reasonable results were seen for this simple curve representation of the face.

Referring back to Section 2.1.3, the active contour methodology is sometimes also cast in a similar framework. A likelihood $p(y|x)$ on the curve x with respect to the image data y is defined as a function of the gradient values along the points defining the curve, and the internal energy of the contour is replaced by a prior distribution $p(x)$, serving the same purpose in constraining the snake shape. A good example of such an approach is found in Cootes (1994), with applications to segmentation of objects such as the ventricles in magnetic resonance images of the brain and the heart chambers in ultrasound.

2.2.5 Landmark Methods

The field of landmark analysis grew out of the understanding that in images of the same kind, certain points may have special significance across the image class, and that these points could be incorporated into statistical models to better understand and represent the image scenes. Bookstein (1986) defines three types of landmarks as follows:

Type I : a point with special anatomical significance, typically the intersection of structures or a branch point;

Type II : a point representing a local mathematical feature, such as a local maximum in some feature such as the curvature or derivative;

type III : a point at an extremal value of some feature.

Within the same framework, the pseudo-landmark can be defined as a point defined only by its position in relation to landmarks of types I, II or III.

Dryden and Mardia (1996) use a similar division, in which anatomically based, mathematically based and derived (from the two others) landmarks are used as categories. To measure the relative positions of the landmarks in a statistical fashion, a joint distribution, often defined as an MRF, can be specified on the landmark positions. This distribution can then be used in the context of an example image or atlas configuration to estimate landmark positions in a new image, or to measure in some sense the shape deviation between two images in the same class. For example, Amit and Kong (1994) extracted mathematical landmarks from an image and pruned them to the same number for each image. A regularizing prior specified via potential functions defined on triangular cliques was then used to limit shape deviations from an atlas when estimating the landmark positions.

Wilson (1995) developed another, scale-space method for the use of landmarks in medical images. Here, several scale-space features were used to define mathematical landmarks in scale-space on an object in a medical image (ventricles in a brain MRI). An MRF graph was then defined to link these landmarks and enable deformation analysis between images of the same class.

One of the unique aspects of this approach to image analysis was the inclusion of scale in the specification of a landmark so that it is situated in the scale-space of the image. The landmark for a two-dimensional image then has associated coordinates $x = \{s_x, s_y, t\}$ (s are spatial coordinates, t is scale). A potential function for an MRF type specification can then be based on the scale-space orientation and distance between two landmarks' position in a template configuration and in the current image. Let d_{ss} denote the scale-space distance between two landmarks, let x_i and x_j be two landmark positions in the image under consideration for which a clique is specified, and let μ_i signify the position of landmark i in the template. Then, define a potential function as

$$\begin{aligned} V_m = \{ & d_{ss}(x_i, x_j)^2 + d_{ss}(\mu_i, \mu_j)^2 \\ & - 2 d_{ss}(x_i, x_j) d_{ss}(\mu_i, \mu_j) cos(x_i - x_j, \mu_i - \mu_j)\} \ . \end{aligned} \tag{2.14}$$

This specification encourages the vectors connecting landmarks in the template and observed image to have similar magnitude and direction.

In addition, a feature-based distribution p_I was defined for each landmark independently on some scale-space feature, chosen to represent things such as "middleness" (Laplacian extremum) or "boundary" (spatial first derivative magnitude). The predictive distribution on landmark positions in a new image for consideration was then defined as the product of these two components. (This contains L independent landmark feature distributions and M MRF-type clique-based distributions, where L is the total number of landmarks and M is the total number of interactions between landmarks in the shape model, or the total number of cliques.)

McCulloch et al (1996) developed this notion further by using a different shape-descriptor with larger cliques. The Procrustes distance d_P measures the deviation of two objects after rotating and scaling linearly to the best match for the two configurations being evaluated. It can then be used as a basis for the potential functions. This approach worked well for simple objects, but the inclusion of scale as a parameter introduced a sizeable computational obstacle. To perform useful estimation for clinical tasks such as segmentation and volume estimation requires a large set of points. When the number of landmarks used is increased to a number sufficient to do this, the computational time increases to levels beyond clinical utility.

Landmark-type analysis in the context of atlas mapping can be taken further by, for instance, incorporating a thin-plate spline (Bookstein, 1991) into the model, basically extending the methodology to estimate a warp of the entire coordinate system. This approach provides a very rich and useful representation of the understanding of an image, but carries a high computational burden. An alternative approach based on modeling the deformation as a Gaussian field has been developed by Amit et al (1991).

McCulloch et al (McCulloch *et al.*, 1997; McCulloch, 1998; Laading *et al.*, 1999) developed a method which extends the landmark idea to one of generalized landmarks, or facets, which are labeled points without any pre-assigned significance attached to them. Only when a set of facets is put down in an atlas or reference image do they gain significance, and even so only implicitly through the interpretation of the atlas. This methodology is described in detail in the next chapter, and it has provided promising results in several medical imaging modalities (e.g. MRI, SPECT).

2.3 Medical Imaging - Cardiac Gated Single Photon Emission Computed Tomography

The application in this thesis is that of gated cardiac single photon emission computed tomography (gated cardiac SPECT). Since inherent features of this modality are crucial to the specific modifications made to the general methodology, the modality and related image-processing methodology applied to this modality is presented here.

2.3.1 Explanation of Modality

SPECT imaging is based on the idea of measuring the physiological activity in tissue by measuring uptake of injected radiochemicals designed to act similarly to substances such as sugar and water. The selected compounds then decay radioactively, causing a high-energy photon to be emitted from a location in the patient. These photons are detected (through a setup involving collimators, screens and photo-multiplier tubes) and form n-dimensional images ($n \in \{1, 2\}$) of the activity in the patient. These n-dimensional images, termed projections, are then used as a basis for an $(n+1)$-dimensional reconstruction of the activity in the 3-dimensional volume ($n = 2$) or the 2-dimensional slice ($n = 1$) being imaged. The reconstruction of the emission intensity is itself the subject of much study, statistical and otherwise, but this aspect of the modality is not considered here.

In cardiac SPECT, the heart is imaged using compounds which go where the heart is consuming energy, thus enabling detection of malfunctioning heart muscle as areas of abnormally low activity. It is typically used to identify extent and location of cardiac ischemia (lack of blood supply to the heart muscle) before heart surgery. The area of greatest interest is the left ventricular (LV) heart muscle, which is responsible for pumping blood out of the LV chamber, through the aorta and out into the body. To create the SPECT image, the patient is injected with a radioactive compound and data is acquired over several beat cycles, causing obvious blurring problems due to the beating of the heart and patient movement. This is necessary to provide enough photons (counts) to image the heart adequately; there is an epidemiological tradeoff between image quality and radiation risk to the patient.

In gated cardiac SPECT, part of this blurring problem is alleviated by partitioning the acquisition of data according to the patient's electrocardiogram (ECG). This is acquired for some beat cycles and divided into k segments. After injection of the radionuclear substance, the detected activity is then assigned to k different images, so that each image represents one specific portion of the patient's heart beat cycle. The acquisition is triggered at a certain point on the ECG (start of contraction or start-systole), and because the partitioning of the cycle is less accurate than detecting its start, the images shortly after that point typically have better characteristics than the ones late in the series. It is important to note that while this method avoids part of the motion-blurring problem, each of the images suffers from the reduction in the number of detected photons per image to approximately $\frac{1}{k}$ of the original number. Thus, the noise characteristics associated with number of photons detected in the individual images are inferior to that of the single image.

This problem forms the motivation for our approach to this image class. The motivation will lead us to estimate the motion of a large number of points on the heart through the image timeseries such that the intensity can be compiled correctly and summarized for better clinical utility than in the traditional SPECT case. Since the gross behavior of the heart is much-studied and relatively well-known, we have an accurate model for its expected behavior, and this model can be used in performance evaluation. Particularly, it is known (Potel *et al.*, 1984) that most of the contraction seen in the heart is radially towards a moving center of contraction, that most of the motion in the LV wall occurs in the outer LV wall, and that the apex (bottom portion) of the LV chamber remains relatively stationary throughout the heartbeat.

2.3.2 Previous Methods Applied to Gated Image Series

The challenge of using the gated images for clinically useful tasks have been adressed in several ways. A recent approach is found in Klein et al (1997), who use an optical flow methodology to trace intensities between two gates taken from a timeseries of positron emission tomography (PET) images. Using this method, they sum the intensities from two images according to the motion field, rather than voxel-wise, to achieve a composite anatomical estimate of the heart radiotracer uptake. Motion estimates are obtained by minimizing the objective function defined as a sum over all voxels of terms of the form

$$e = e_I + e_s + e_d \ . \tag{2.15}$$

Define an image motion field $u(x) = \{u_1(x), u_2(x), u_3(x)\}$, $x \in \mathcal{X}$ from image I_1 to image I_2, where $\mathcal{X}$ is the image domain. Then define a squared difference image error function e_I on the image intensities at a location in image I_2 and the deformed version of I_1. In other words,

$$e_I = \lambda_I \left(I_2(x) - I_1(x-u) \right)^2 \ . \tag{2.16}$$

Also, define a smoothness-based function

$$e_s = \lambda_s \sum_{i=1}^{3} \left(\frac{\delta}{\delta x_i} u_i \right)^2 \ . \tag{2.17}$$

Finally, e_d is defined to discourage by an overfitting to the image data enforcing an incompressibility constraint; that is, a penalty on moving points closer together. Thus, e_I encourages image similarity in a direct intensity sense, e_s enforces smoothness in motion between adjoining voxels, and e_d pulls the configuration towards the original image, limiting the magnitude of motion allowed. The parameter vector λ adjusts the smoothness of the deformation estimate allowed. Like Collins et al (1995), a multiscale minimization approach was utilized to speed convergence. The results for this method were promising, if somewhat sparse.

Other researchers (McEachen and Duncan, 1997; Clarysse *et al.*, 1997) have used shape-based methodology similar to that described in Section 2.1.3, applied to gated magnetic resonance image timeseries of the heart, with some success.

Chapter 3

METHODS

The motivation behind my model is to statistically represent **an intelligent observer's placement of labeled points in images of a certain class**. This implies a focus on the image as is, rather than the physical formation of the images. There is no attempt made to model the image acquisition itself. Rather, the perception of the image scene is modeled based on using ideas inspired by the methodologies outlined in the previous chapter.

3.1 Hierarchical Deformation Model - Facets

The model is developed around the concept of a facet, which should be interpreted as a type of generalized landmark or a labeled point. Facets differ from landmarks (following Bookstein (1991)) in that facets do not correspond to specific pre-defined anatomical or mathematical features. Instead, each facet's label is generally inferred by its location in a reference or atlas image. A facet model can be thought of as a representation of the image as a concept, with the stated goal of providing a general framework for representing the image as contextual perceived information rather than a set of individual pixels. This includes the image intensity resulting from the image formation as a basis for the analysis, but in addition includes a wider context, both relative to other images in the image class and between points in the image. The basis of the modeling is that we have a class of images containing similar structures. We then want to identify points that represent the same conceptual quantity; that is, we are interested in obtaining the map of perceived deformation

$$\{M : (\mu, \phi) \to (x, f)\} \ , \tag{3.1}$$

with

$$\{\mu, x\} \in \mathcal{X} \subset R^d, \{\phi, f\} \in \mathcal{F} \subset R^p \ . \tag{3.2}$$

Here, $\mathcal{X}$ is the space spanned by the image grid, and $\mathcal{F}$ the space spanned by the image features used in modeling the perception as described below. Furthermore, d is the dimensionality of the images and p is the length of a feature vector derived from the image data. The positions of all facets are contained in the vector x, and the corresponding positions in the reference image are held in vector μ. Similarly, relevant image information pertaining to placement of the facets is contained in the vectors f and ϕ for the image under consideration and the reference, respectively. In this model, an individual facet labeled i has associated position x_i. It also has an associated feature value or values f_i. To obtain the map M, we define a joint predictive probability distribution p on $\{x, f\}$ given hyperparameters $\theta = \{\theta_x, \theta_f\}$, with x and f conditionally independent, typically so that (see Section 3.1.3)

$$p(x, f|\theta) = p_S(x|\theta_x) p_I(f|\theta_f) \ . \tag{3.3}$$

The parameter vector θ_x contains location and scale parameters $\{\mu, \kappa, \sigma^2\}$ for the facet location (shape) portion p_S of the distribution, and θ_f similarly contains parameters $\{\phi, \tau\}$ for the feature (image) part p_I of the distribution. We separate the distribution in this way to reflect that two distinct processes are at work; one which favors a predefined facet configuration and one which encourages similarity in some image-derived feature.

3.1.1 Shape Distribution

The baseline shape distribution p_S is now defined. As before, let x be the vector of facet positions. Let x_l indicate the vector of facets at level l in the hierarchy and let x_{lj} refer to an individual facet j in that level. Similarly, let μ be the vector of corresponding facet locations in the reference image T. The distribution $p_S(x|\theta_x)$ is assumed to have a hierarchical normal structure defined by equations (3.4) and (3.5). Each level l has N_l facets, and d is the dimensionality of the images in the class the model is applied to. For $(L+1)$ levels in the hierarchy, $l \in \{0, \ldots, L\}$, define:

$$p_S(x|\theta_x) = p_S(x_0|\theta_x) p_S(x_1|x_0;\theta_x) \ldots p_S(x_L|x_{L-1};\theta_x) \tag{3.4}$$

where each factor is a density of the form:

$$\begin{aligned} p_S(x_0|\kappa,\mu) &= \mathrm{MVN}(\mu_0, \kappa\sigma_0^2 \mathrm{I}_d) \\ p_S(x_1|x_0;\kappa,\mu) &= \mathrm{MVN}(\mu_1 + A_1\Delta x_0, \kappa\sigma_1^2 \mathrm{I}_{N_1 d}) \\ &\vdots \\ p_S(x_l|x_{l-1};\kappa,\mu) &= \mathrm{MVN}(\mu_l + A_l\Delta x_{l-1}, \kappa\sigma_l^2 \mathrm{I}_{N_l d}) \\ &\vdots \\ p_S(x_L|x_{L-1};\kappa,\mu) &= \mathrm{MVN}(\mu_L + A_L\Delta x_{L-1}, \kappa\sigma_L^2 \mathrm{I}_{N_L d}) \; . \end{aligned} \tag{3.5}$$

Here, $\mathrm{MVN}(a, \Sigma)$ denotes the multivariate normal density with mean vector a and covariance matrix Σ. The vectors x_l and μ_l have identical lengths, $N_l d$. The difference between facet placement and the reference position at a level l is given by $\Delta x_l = x_l - \mu_l$. I_n denotes the n x n identity matrix. A_l is an $N_l d$ x $N_{l-1} d$ design matrix for the hierarchical model, which in one dimension would schematically have entries indexed as follows,

$$A_l = \begin{bmatrix} w_{11}^{(l)} & \dots & w_{1k}^{(l)} & \dots & w_{1N_{l-1}}^{(l)} \\ w_{21}^{(l)} & \dots & w_{2k}^{(l)} & \dots & w_{2N_{l-1}}^{(l)} \\ & & \vdots & & \\ w_{j1}^{(l)} & \dots & w_{jk}^{(l)} & \dots & w_{jN_{l-1}}^{(l)} \\ & & \vdots & & \\ w_{N_l1}^{(l)} & \dots & w_{N_lk}^{(l)} & \dots & w_{N_lN_{l-1}}^{(l)} \end{bmatrix} . \tag{3.6}$$

Finally, σ_l^2 is a conditional scale parameter for a facet on level l given knowledge of the locations of facets on level $(l-1)$. The parameter κ is an overall scale factor to allow for adjustments in the weighting of shape (p_S) versus image (p_I) portions of the density in equation (3.3). The effective hierarchical conditional variance for a facet on level l is given by the product, $\kappa\sigma^2$.

This form for p_S captures the deformation on several levels of scale, thus easing the exploration of configuration space. This means that gross deformations are modeled by upper-level facets and that lower-level facets will not be subsequently penalized for the same movement. This makes sense in the same way scale-space does; that image perception should be modeled as occuring at a variety of resolutions, and that once a gross deformation has been detected, we move our expectations about deformation on finer resolutions accordingly. An alternate way of writing p_S is as a distribution on deviations, namely

$$\begin{aligned} p_S(\Delta x_0|\kappa,\mu) &= \mathrm{MVN}(0, \kappa\sigma_0^2 \mathrm{I}_d) \\ p_S(\Delta x_1|\Delta x_0;\kappa,\mu) &= \mathrm{MVN}(A_1\Delta x_0, \kappa\sigma_1^2 \mathrm{I}_{N_1d}) \\ &\vdots \\ p_S(\Delta x_l|\Delta x_{l-1};\kappa,\mu) &= \mathrm{MVN}(A_l\Delta x_{l-1}, \kappa\sigma_l^2 \mathrm{I}_{N_ld}) \\ &\vdots \\ p_S(\Delta x_L|\Delta x_{L-1};\kappa,\mu) &= \mathrm{MVN}(A_L\Delta x_{L-1}, \kappa\sigma_L^2 \mathrm{I}_{N_Ld}) \ . \end{aligned} \tag{3.7}$$

This emphasizes the fact that the model only penalizes deviations once; that is, if an upper level motion occurs, then low-scale deviations are recentered to a non-zero value according to this motion.

Choice of Design Matrix

The design matrix A which defines the interactions between facets may be constructed in several ways. The most straightforward formulation and the least costly computationally is the single connection approach, where each row in A_l has only one entry. Schematically, a one-dimensional example of that would be the form

$$A_l = \left[\begin{array}{ccccccc} & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \ldots & 0 & 1 & 0 & 0 & 0 & \ldots \\ \ldots & 0 & 1 & 0 & 0 & 0 & \ldots \\ \ldots & 0 & 0 & 1 & 0 & 0 & \ldots \\ \ldots & 0 & 0 & 1 & 0 & 0 & \ldots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \end{array} \right] . \tag{3.8}$$

This approach provides quite compact full conditional distributions (see Section 3.1.1), in that there are no direct interactions between facets on the same level. These downwards conditional dependencies are shown as a tree-structure in Figure 3.1.

There is a great deal of anisotropy spatially with this approach, however, and so model-driven rather than data-driven blocking of deformation has a tendency to occur. To see this, consider the marginal distribution of a hierarchical normal model as described in Gelman et al (1995). Given vectors U and V such that

$$\begin{aligned} U|V &\sim MVN(AV, \Sigma_{U|V}) \\ V &\sim MVN(\mu_V, \Sigma_V) \ , \end{aligned}$$

then the joint density on (U, V) has the form

$$\left[\begin{array}{c} U \\ V \end{array} \right] \sim MVN \left(\left[\begin{array}{c} A\mu_V \\ \mu_V \end{array} \right], \left[\begin{array}{cc} A\Sigma_V A' & \Sigma_{U|V} \\ \Sigma'_{U|V} & \Sigma_V \end{array} \right] \right) . \tag{3.9}$$

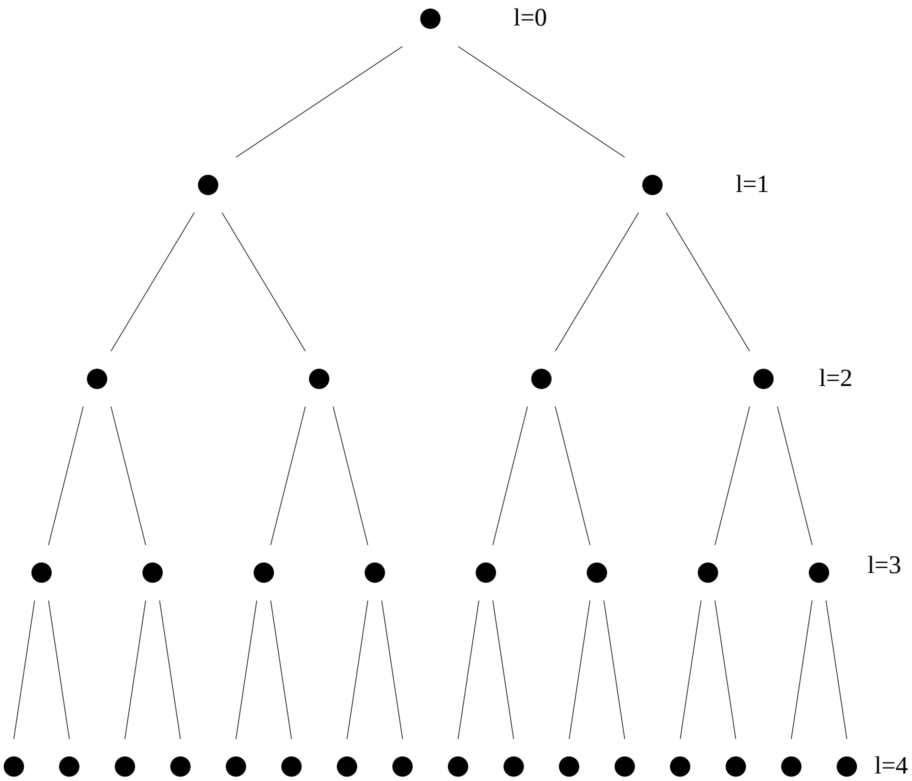

Figure 3.1: The conditional structure of the model when defined as in equation 3.8, that is, with each facet on a level connected only to one parent on the level above.

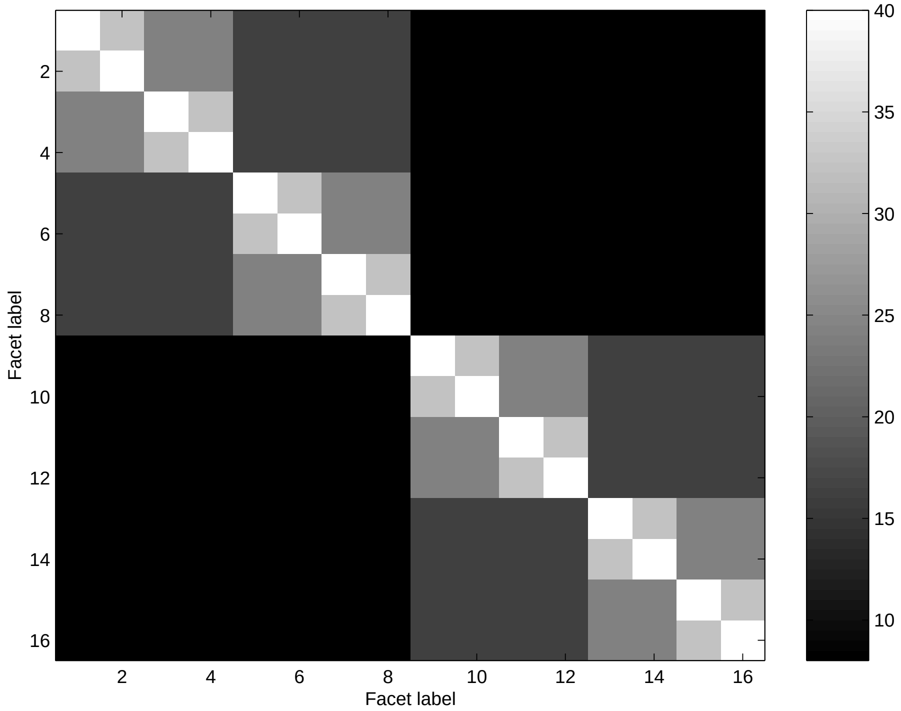

Figure 3.2: The marginal covariance structure for the form of the design matrix as seen in equation (3.8). We see the blockedness inherent in this model, which will give rise to model-driven artifacts when applied to real image data. The labeling runs from one side to the other in Figure 3.1.

This, repeatedly applied as per the context of the hierarchical model defined in equation (3.5) gives a marginal covariance matrix Σ_l on level l of the nested form

$$\Sigma_l = \kappa^{-1} \left\{ \sum_{j=0}^{l-1} \sigma_j^{-2} \left[\prod_{i=j+1}^{l} A_i \right] \left[\prod_{i=j}^{l-1} A'_{l-i} \right] + \sigma_l^{-2} I_{N_l d} \right\} . \tag{3.10}$$

Figure 3.2 illustrates this effect for a one-dimensional vector x with 4 levels in the hierarchy.

If we have knowledge about the objects in the image class we are applying the method to, linkages in the hierarchy (non-zero entries in A) could be tied to which object a facet belongs to. The blocking would then be natural rather than artificial. Such specific definitions would require more prior knowledge and less generality than we wish to assume here, however. While some attempts at automatic object linkage generation via maxima in the scale-space of a reference image have been made (see

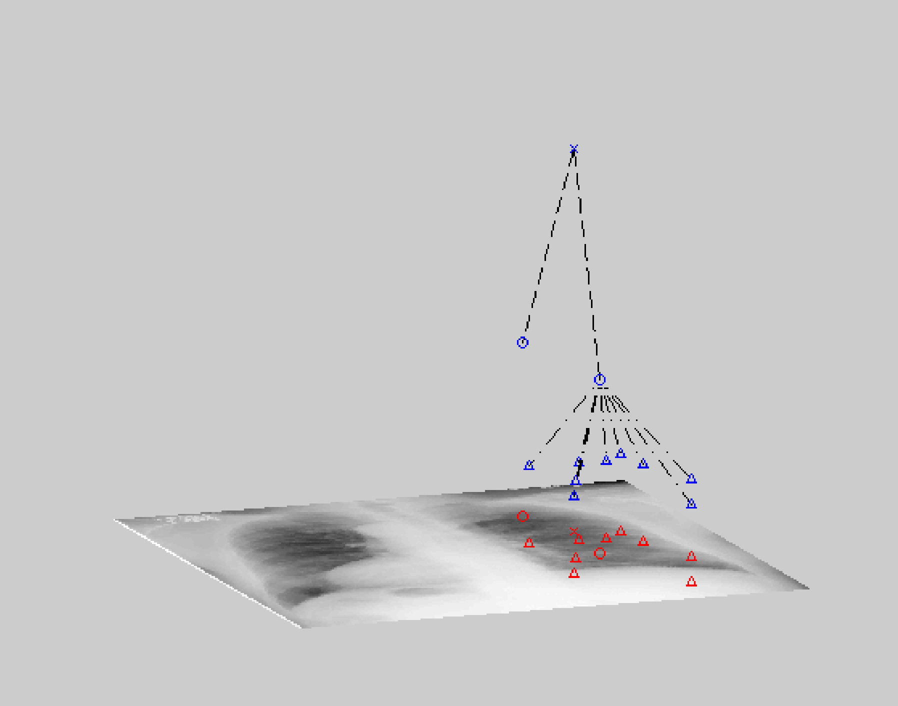

Figure 3.3: A portion of the resulting design matrix from tracing extrema in the spatial Laplacian for the image shown (a digital chest radiograph). The top three levels for the connections which turned out to represent the left lung are shown, displaying the decreasing number of extrema in scale-space as scale increases (upward direction from image plane in figure). The A_l matrices would then be set to ones for entries corresponding to the linkages seen between levels.

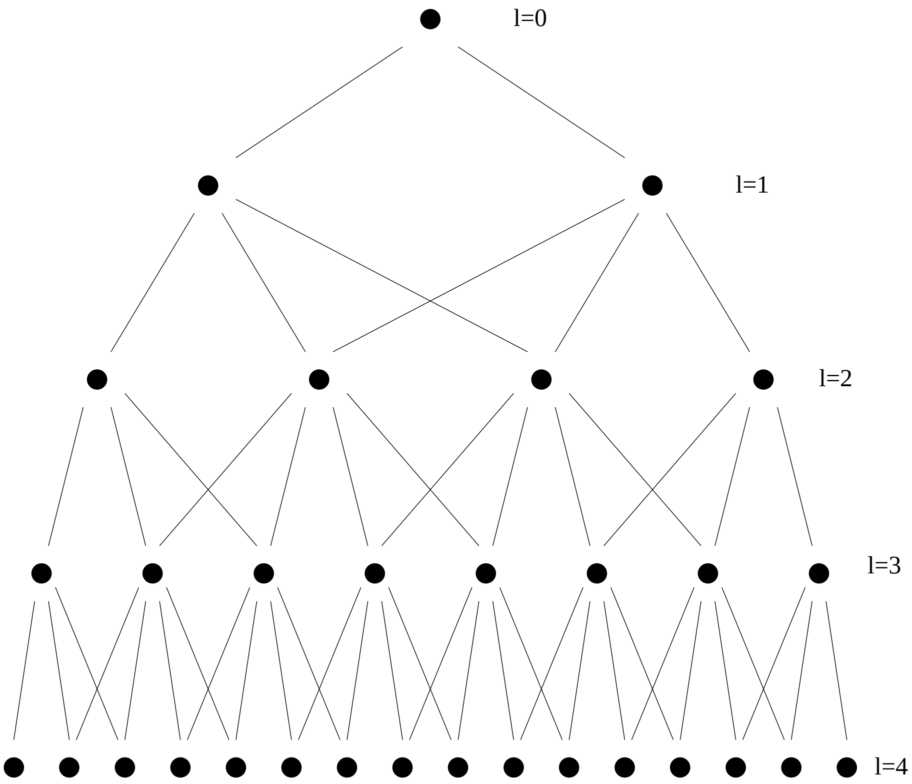

Figure 3.4: The conditional structure of the model when defined as in equation 3.11, that is, each facet on a level is now connected to the 2^d nearest parents on the level above.

Laading et al (1997) and Figure 3.3 for an example), this has not to date been proven to be a practical approach.

Approximate isotropy can be achieved if A is expanded in complexity and taken schematically (in one dimension) to be as follows,

$$A_l = \begin{bmatrix} & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \dots & 0 & w^{(l)}_{(j-1)1} & w^{(l)}_{(j-1)2} & 0 & 0 & \dots \\ \dots & 0 & w^{(l)}_{j1} & w^{(l)}_{j2} & 0 & 0 & \dots \\ \dots & 0 & 0 & w^{(l)}_{(j+1)1} & w^{(l)}_{(j+1)2} & 0 & \dots \\ \dots & 0 & 0 & w^{(l)}_{(j+2)1} & w^{(l)}_{(j+2)2} & 0 & \dots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}. \qquad (3.11)$$

This form of A_l enforces some smoothness on the deformation; that is, the marginal covariance between any pair of facets on a level l becomes a smooth decreasing func-

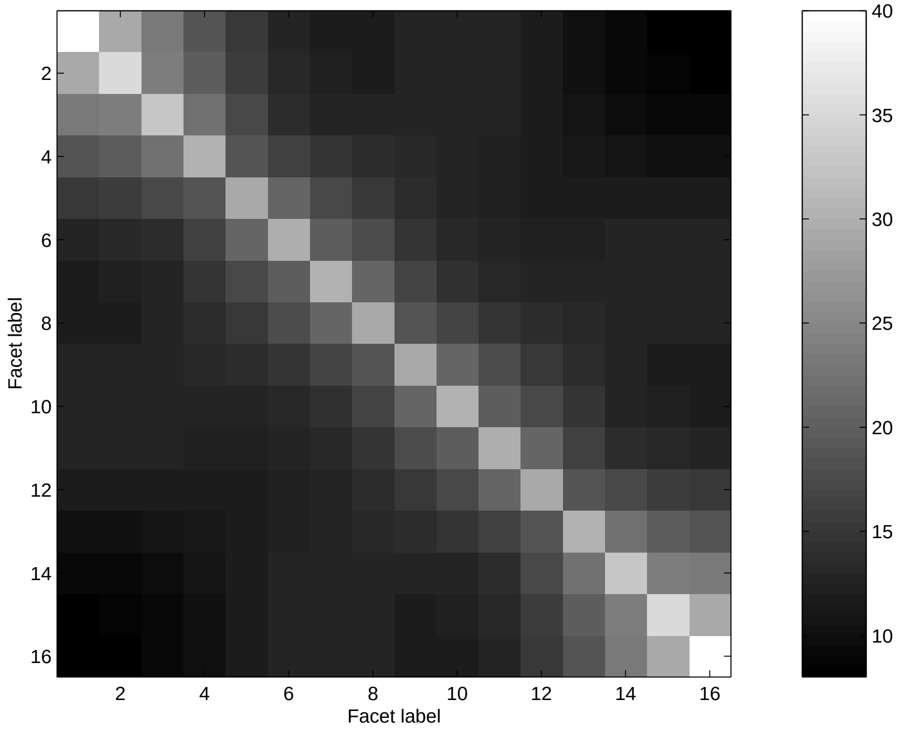

Figure 3.5: The marginal covariance structure for the form of the design matrix as seen in equation 3.11. We see the that the blockedness seen in Figure 3.2 has been alleviated in this model. We still see some edge effects, corresponding to locations where multiple parents are unavailable. In the interior, the structure is nearly isotropic. The labeling of facets runs from one side to the other in Figure 3.4

tion of their template (μ) distance, provided that the w are appropriately selected. To handle the boundaries in the model (facets on the outer edge of the tree structure), facets are tied more strongly to the closest available parents in the sense seen schematically in one dimension in Figure 3.4.

One way in which the near-isotropy can be achieved is to set the values of the $w_{jk}^{(l)}$ proportional to the inverse distance between μ_{lj} and $\mu_{(l-1)k}$, constrained by $\sum_k w_{jk}^{(l)} = 1$ and such that only the 2^d closest facets on level $(l-1)$ to facet lj are given non-zero weights $w_{jk}^{(l)}$. If the w are determined in this way, as illustrated in Figure 3.5, most of the blocking problem is mitigated. A visual example of how the hierarchy propagates observed deformations is seen in Figure 3.6. More accurate isotropy could be achieved by increasing the number of non-zero entries, but this

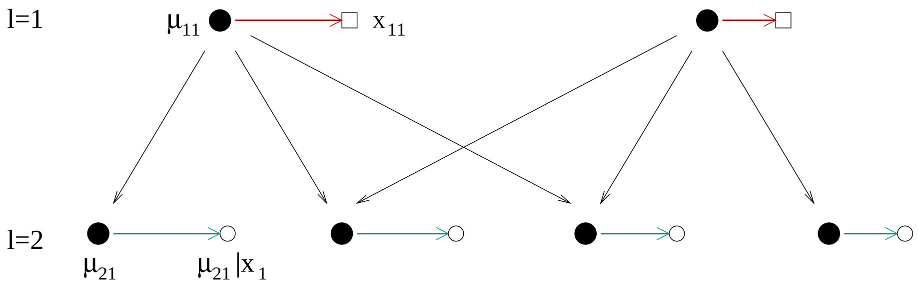

Figure 3.6: The levels $l = \{1, 2\}$ in the one-dimensional example with means for all facets shown as solid circles. Given the observed values of level $l = 1$, the deviation is propagated in expectation to level $l = 2$ through the hierarchical model.

minor improvement in the approximation of isotropy has not proven to be worth the extra computation required to include further interactions into the model.

To achieve the same improvement in isotropy for the 3-dimensional application presented in the next chapter, such entries in a row of A_l would be kept limited to 24 (or 8 per direction) in a parallel fashion to that seen above. Outer facets would be handled similarly with the exception that, when thinking of a cube of facets, corner facets would have one parent facet available, edge facets would have two parents available and face facets would have four parents. The edge effects diminish as you go to higher numbers of levels in the model.

Full Conditional Distributions

For both the implementation and understanding of the model, it is important to look closely at the full conditionals resulting from the shape model above. Consider Figure 3.7 together with Figure 3.4 for a coarse one-dimensional concept illustration of a facet's conditional connectedness. Using standard normal theory and assuming the structure of A as indicated above, this is easily shown to result in the following conditional distribution for $p(x_{lj}|x_{-lj})$. For $l \neq \{0, L\}$,

$$p(x_{lj}|x_{-lj}) = N(\mu_{lj}|\{x_{-lj}\}, \sigma^2_{lj}|\{x_{-lj}\}) \ , \tag{3.12}$$

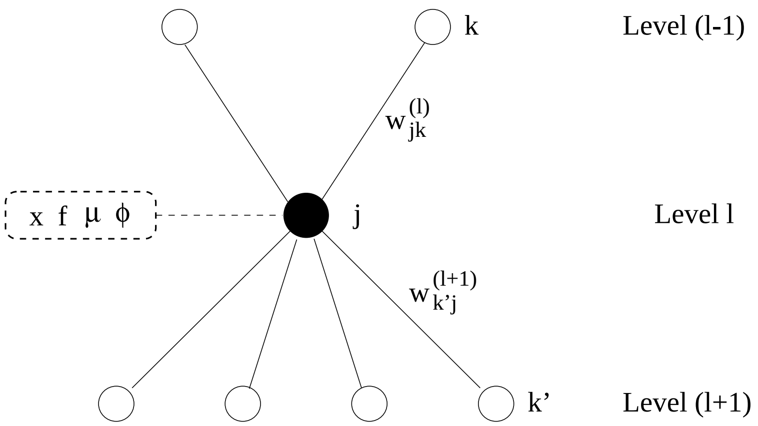

Figure 3.7: One-dimensional schematic for a facet in the hierarchical conditional specification of the model. A facet in the interior on level $L > l > 1$ is shown. When thinking in terms of the full conditional distributions or the equivalent MRF specification, horizontal linkages are added to the ones seen in this figure.

with

$$\begin{aligned} \mu_{lj}|\{x_{-lj}\} = \mu_{lj} \quad &+ \quad \frac{\sigma_l^2}{\sigma_{lj}^2|\{x_{-lj}\}} \sum_{k\in P_{lj}} w_{jk}^{(l)}(x_{(l-1)k} - \mu_{(l-1)k}) \\ &+ \quad \frac{\sigma_{l+1}^2}{\sigma_{lj}^2|\{x_{-lj}\}} \sum_{k\in D_{lj}} w_{kj}^{(l+1)}\xi_{lk} \ , \end{aligned} \tag{3.13}$$

where

$$\xi_{lk} = \left\{ (x_{(l+1)k} - \mu_{(l+1)k}) - \sum_{j'\in P'_{(l+1)k}} w_{j'k}^{(l)}(x_{lj'} - \mu_{lj'}) \right\} \ , \tag{3.14}$$

and

$$\sigma_{lj}^{-2}|\{x_{-lj}\} = \sigma_l^{-2} + \sigma_{l+1}^{-2} \sum_{k\in D_{lj}} (w_{kj}^{(l+1)})^2 \ . \tag{3.15}$$

Here, the subscript $-lj$ indicates all variables in x except for those associated with facet lj. All other conditioning is suppressed in the notation. The term $\mu_{lj}|\{x_{-lj}\}$ denotes the mean of the full conditional distribution, not to be confused with the placement of a facet in the reference image, μ_{lj}. In the summations, P_{lj} denotes the set of facets on level $(l-1)$ which contribute to the mean (via the design matrix A_l) in the full conditional distribution on x_{lj}. The index k runs over that set. Similarly, D_{lj} is the set of facets on level $(l+1)$ related directly to facet lj through the design matrix A_{l+1} between those levels.

The top level has the simpler form

$$\mu_0|\{x_{-0}\} = \mu_0 + \frac{\sigma_1^2}{\sigma_0^2|\{x_{-0}\}} \sum_{k \in D_0} w_{kj}^{(1)} \xi_{0k} \tag{3.16}$$

and

$$\sigma_0^{-2}|\{x_{-0}\} = \sigma_0^{-2} + \sigma_1^{-2} \sum_{k \in D_0} (w_{kj}^{(1)})^2 \ . \tag{3.17}$$

The bottom level is finally specified by

$$\mu_{Lj}|\{x_{-Lj}\} = \mu_{Lj} + \sum_{k \in P_{Lj}} w_{jk}^{(L)} (x_{(L-1)k} - \mu_{(L-1)k}) \tag{3.18}$$

since obviously

$$\sigma_{Lj}^2|\{x_{-Lj}\} = \sigma_L^2 \ . \tag{3.19}$$

Graphical Model Representation

It is instructive at this point to draw attention to the equivalent MRF model, an equivalence investigated by, among others, Lavine (1998). It is shown, through calculation of the full conditional distributions as seen in equation (3.12), that a facet hierarchy in which there are multiple connections between facets on adjacent levels corresponds to an equivalent model graph in which facets are connected not only to the non-zero contributors on the level above (parents) and below (children), but also to the facets on the same level with which it has common children. This effect conceptually comes from the fact that an adjustment in a parent's position is directly related to all the facets which are its descendants, but only to the extent not explained by

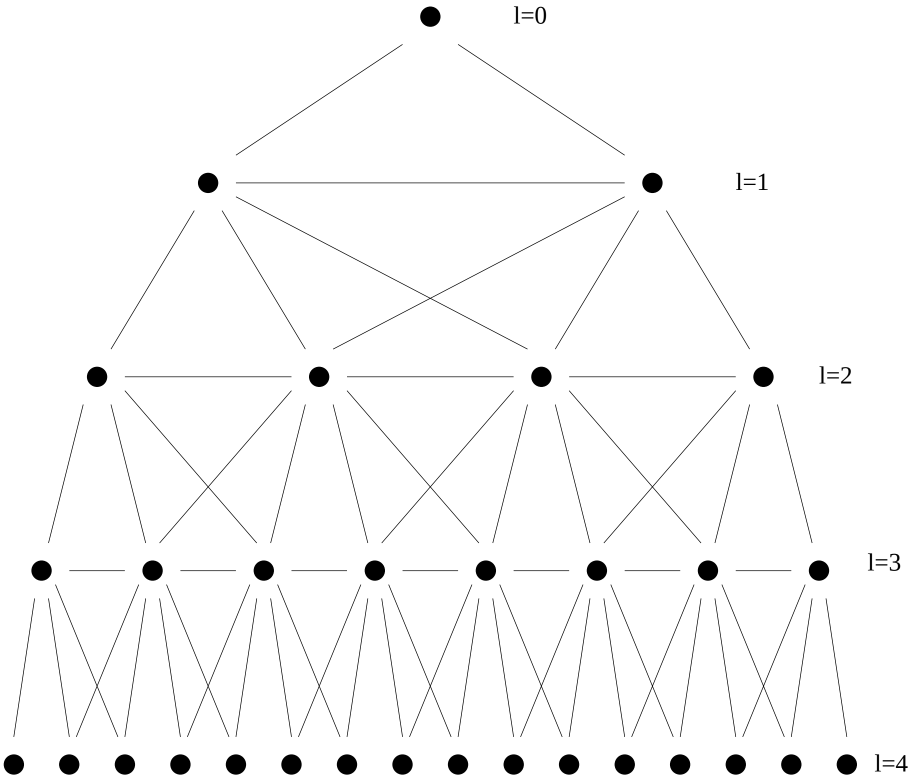

Figure 3.8: An illustration of the full conditional graph structure for the model using the design matrix in equation 3.11. The emphasis is still on the dependence upwards and downwards, but there is also the added element of direct dependence on siblings as seen in the full conditional distribution in equation 3.12.

other parents of the same facets. In this way, a “multiply connected” scheme spreads the full conditional interaction pattern somewhat, and the model graph or equivalent MRF representation for the hierarchical model structure in Figure 3.4 becomes that seen in Figure 3.8.

Parameter Choice

Consider the parameters $\{\mu, \kappa, \sigma^2\}$ in the shape portion of the model. The vector μ represent reference positions, κ is an overall scale parameter and σ^2 expresses the relative variances on the levels in the hierarchical model, respectively. The vector μ is typically taken so that the positions on each level, μ_l, form an evenly spaced square grid, giving the number of facets per level as $N_l = 2^{ld}$. This reflects our intent not to treat any parts of or positions in the images as having a priori particular significance, except possibly where we lay the facet grid down in the reference image (location of μ_0). The most appealing way to define the hierarchical conditional variances σ^2 and the overall scale parameter κ is obviously to estimate them from a set of existing facet placement data, as is demonstrated in McCulloch (1998) through use of expectation maximization (EM) methodology. Introducing a prior on κ, estimation of maximum a posteriori values is straightforward in a missing data framework (due to the large number of facets typically used, it is completely unrealistic to ever have all facet positions for clinical data available).

When such data does not exist, as is true in our current situation, they can be set according to several heuristic schemes. One way of specifying this parameter set is to choose them so that the marginal variance of a facet on the lowest level is approximately equal regardless of the number of levels. Thus, the number of levels and bottom-level density of facets can be set according to the perceived need for resolution or availability of computational resources. Another scheme is to set the parameters σ^2 empirically so that there is a balance in how much of the motion is explained by the various levels of facets. If you know that the object you want to capture moves mostly in a gross fashion, it makes sense that the lower-level facets are more constrained (lower hierarchical conditional variance). Similarly, if you don't have any previous knowledge about the scale at which the deformation occurs, it makes sense to allow approximately equal explanatory power to be given to each level in the model.

3.1.2 Feature Distribution

Now consider the image-driven part of the joint distribution $p(x, f|\theta)$, $p_I(f|\theta_f)$. As stated previously, let f be the vector of image-derived feature values associated with the set of facets. Similarly, let ϕ be the corresponding vector of template values. Then given an image match function g, the facet features are assumed to be drawn from an exponential family distribution given in equation (3.20). It is further assumed that if $\{f_j, \phi_j\}$ are the corresponding jth element or jth set of elements of the vectors $\{f, \phi\}$, respectively, then the feature distribution p_I is modeled as a product of univariate distributions with a common image match function g_I :

$$\begin{aligned} p_I(f|\tau, \phi) &\propto \exp\left\{-g(f, \phi)\right\} \\ &\propto \exp\left\{-\frac{1}{2\tau}\sum_{j=1}^{N_L} g_I(f_j, \phi_j)\right\} \ . \end{aligned} \tag{3.20}$$

Here, τ is an overall scale parameter which is similar to κ defined for the shape distribution, p_S. The sum extends over those facets that have associated feature values, which typically are the facets on the lowest level L in the hierarchy, numbering N_L. The parameters ϕ_j are taken to be some image feature value calculated at the locations μ_{Lj} in the reference image T: $\phi_j = \phi(T, \mu_{Lj})$. Note that this does not necessarily imply taking the image value at μ_{Lj} directly. The independent form of equation (3.20) means that while we may consider more than one feature value per facet, we model the observation of feature f_j at a facet j as independent of the observation of features at all other facets. This again is based on modeling the perception rather than the images, which enables us conceptually to eliminate direct spatial feature correlation modeling.

The derivation of $\{f_j, \phi_j\}$ from the image data can be specified in a number of ways. It is generally not a good idea to use the image intensity directly, as most image problems have variations in both mean intensity and intensity range, even for the imaging of an identical object. A photograph of the same person with bright sunshine for one picture and overcast the next is a perfect example. For that reason, the choice of a quantile-rescaled image intensity might make sense,

$$\phi = q(T, \mu)\ ,\ f = q(Q, x)\ , \tag{3.21}$$

where q normalizes the images T (reference) and Q to some pre-defined mean and stretches the range such that the 25th and 75th percentile of the image intensity in some region of interest match. The low-scale image Laplacian in equation (2.2),

$$\phi = L^T_{ii}(\mu; t)\ ,\ f = L^Q_{ii}(x; t)\ , \tag{3.22}$$

is another choice with several interesting features. First, it has been shown to enhance edges, so if you are interested in perception of boundaries, it may be a good candidate. Another interesting feature is that isointensies in the Laplacian form closed curves when considering an image scene not close to the edge of the image. The curves of zero-crossings are one natural set of such quantities to focus on which have often been used as a basis for region extraction in images. A point with an extremal value of the Laplacian at higher scales is usually indicative of a center of high or low intensity, for instance the eyes or mouth in the facial example.

The feature function g_I also offers several options. To choose a function appropriate for the imaging problem at hand, we once again must consider computational load against image representation. For computational reasons and ease of interpretation, one of the most commonly used functions is the squared difference, simply

$$g_I(f_j, \phi_j) = (f_j - \phi_j)^2 \ , \tag{3.23}$$

which corresponds to independent normal distributions centered at ϕ_j and with variance τ(from equation (3.20)). This feature is very local and, while sensitive, can work well in situations with good contrast betweens regions of interest and low noise. It is particularly important in this case that the images are well matched to start with, so that $f_j = \phi_j$ makes sense in the conceptual matching sense.

Another feature function g_I found to be effective for some noisier imaging problems is one based on a local intensity regression around the facet in question. A small neighborhood defined by a set of m points is placed around the facet's position in the template image T (around μ_{Lj}) and observed image Q (around x_{Lj}). Subsequently, these points are evaluated in T and Q, respectively, to form m-vectors ϕ_j and f_j, indexed by k in equation (3.24). This is motivated by the thinking that an observer may perceive not so much an isolated intensity at a point, but rather the local context when considering an image feature. A normalized regression parameter (see (McCulloch, 1998) for details) can then be calculated according to

$$\begin{aligned} g_I(f_j, \phi_j) &= 1 - a = 1 - f_j^{*\prime} \phi_j^* \phi_j^{*\prime} f_j^* \ , \\ &= 1 - \frac{(\sum_k \phi_{jk} f_{jk} - \frac{1}{m} \sum_k \phi_{jk} \sum_k f_{jk})^2}{(\sum_k \phi_{jk}^2 - \frac{1}{m}(\sum_k \phi_{jk})^2)(\sum_k f_{jk}^2 - \frac{1}{m}(\sum_k f_{jk})^2)} \ , \end{aligned} \tag{3.24}$$

which is subsequently used to define the distribution function p_I (equation (3.20)). The latter term in equation (3.24) is basically a normalized regression coefficient ($a \in [0, 1]$), reflecting the overall correspondence between the two sets of m points. It is then inverted in sign to provide a measure which gives a maximum rather than a minimum for a good match when g_I is used in equation (3.20). This feature function picks up patterns of intensity both level- and scale-independent (in image values rather than size), and is thus less sensitive to image intensity level and scaling. It is, however, much more computationally intensive.

3.1.3 Observational Constraint

We will treat the fact that we typically observe the position and feature in concert when observing a new image as a constraint, i.e. $f = Q(x)$, without modeling this explicitly. By doing this, we achieve one of our stated goals, which is to make no model assumptions about the image formation in the class. If we included a model such that $p(x, f|\theta) = p(x|\theta_x)p(f|x, \theta_f)$ we would be implying directly a certain spatial correlation structure in the image intensities. McCulloch (1998), for example, mentions a general model based on Cressie (1993) as

$$p(f|\phi, \Sigma_f(x)) \propto |\Sigma_f(x)|^{-1/2} \exp\left\{-\frac{1}{2}(f-\phi)'\Sigma_f(x)^{-1}(f-\phi)\right\} \ . \qquad (3.25)$$

This leads to a non-standard joint density $p(x, f|\theta)$ thanks to the non-linear dependence of f on x via Σ_f. This density is not at all desirable from our standpoint, however, as it makes fundamental assumptions about the underlying structure of the image data. While any model we define jointly on f and x will introduce marginal correlations, the explicit modeling of this is precisely what the facet model is not intended to do. The implementational implications of this interpretation will be further discussed in Section 3.3.

3.2 Modality-Specific Modifications

The baseline shape model defined in equations (3.4) and (3.5) captures the shape changes in the general case when little prior knowledge is available about the shape change within the image class. If we know salient facts about the expected behavior of the structure we are trying to match to the reference, it makes sense in some cases to modify the expectations in the shape portion of the model to accomodate this, given that it is computationally feasible. This corresponds to an overall adjustment to the expectation of facets' observed locations (μ).

In the cardiac gated SPECT scenario, we know a priori that the heart size changes during the heart beat cycle, and so it is sensible to build this into the model. Doing this involves a tradeoff between detailed modeling of the contraction process, something which is possible thanks to an extensive body of knowledge about the cardiac system, and computational feasibility and stability. We also should be mindful of the fact that non-linear contractions not incorporated into the gross model can still be accommodated by the baseline deformation model given our added parameters.

To define the 3-dimensional contraction model, let γ_1 consist of three coordinates for a center of contraction, and let γ_2 be a 3-dimensional set of contraction factors for orthogonal directions $\{1, 2, 3\}$. This is a reasonable representation based on the studies cited in Section 2.3, where the finding was that most of the contraction seen in the heart is directed towards a moving center of contraction. Also, we limit ourselves to these relatively few contraction parameters based on computational feasibility and the facts that the images are rather coarse and that the deformation portion of the model should be able to handle the remaining contraction patterns.

Specifying the new shape model, then, let the full vector $\gamma = \{\gamma_1, \gamma_2\}$ transform μ in the shape portion of the model in equation (3.5) to a vector of contracted means $\mu^c(t)$ as given in equation (3.26) at time t in the image time series. Given γ_2, μ^c is defined such that along each orthogonal direction, $\mu - \mu^c$ is a linear function of $\mu - \gamma_1$, increasing in magnitude the further we get from γ_1. An illustration can be seen in Figure 3.9. Thus, the facet grid will contract in expectation towards the center of contraction γ_1. When all elements of γ_2 are equal, this corresponds to a straight

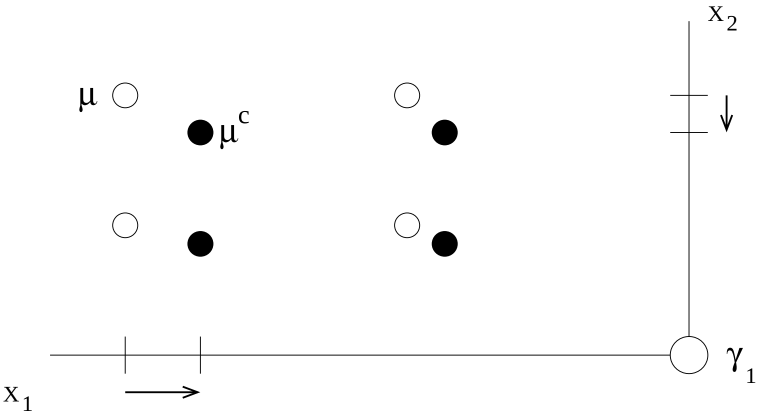

Figure 3.9: Visual illustration of how the contraction model modifies μ through γ into a set of contracted means μ^c. The open circles symbolize μ and the solids indicate μ^c. Each level in the hierarchy undergoes this same transformation.

linear radial contraction.

The shape distribution p_S, then, is modified accordingly in the general case $l \in \{1, \ldots, L\}$ to

$$p_S(x_l | x_{l-1}; \kappa, \mu, \gamma(t)) = \text{MVN}(\mu_l^c(t) + A_l \Delta x_{l-1}, \kappa \sigma_l^2 \text{I}_{N_l d}) \quad , \tag{3.26}$$

with

$$\mu_l^c(t) = \bar{\gamma_1} \mathbf{1}_{N_l d} + (\mu_l - \bar{\gamma_1} \mathbf{1}_{N_l d}) \hat{\gamma_2}(t) \ . \tag{3.27}$$

The top level $l = 0$ is similarly defined, with the simpler form

$$p_S(x_0 | \kappa, \mu, \gamma(t)) = \text{MVN}(\mu_0^c(t), \kappa \sigma_0^2 \text{I}_d) \ . \tag{3.28}$$

Throughout, $\bar{\gamma_1}$ is a stacked vector of N_l replicates of γ_1, $\hat{\gamma_2}$ is the diagonal matrix with N_l replicates of γ_2 along the diagonal, and $\mathbf{1}_n$ is the n-dimensional one-vector. The interpretation of and form for the model parameters $\{A_l, \kappa, \sigma_l^2, \mu\}$ remain unchanged. Note also that the introduction of μ^c does not change the values of ϕ_j, which are still taken as the reference image feature values at μ_{Lj}, that is, $\phi_j = \phi(T, \mu_{Lj})$. See Figure 3.10 for a schematic representation of the modified model structure.

With this added level in the hierarchy, a prior distribution can be included to capture the expected contraction pattern during the beat cycle. A natural choice for

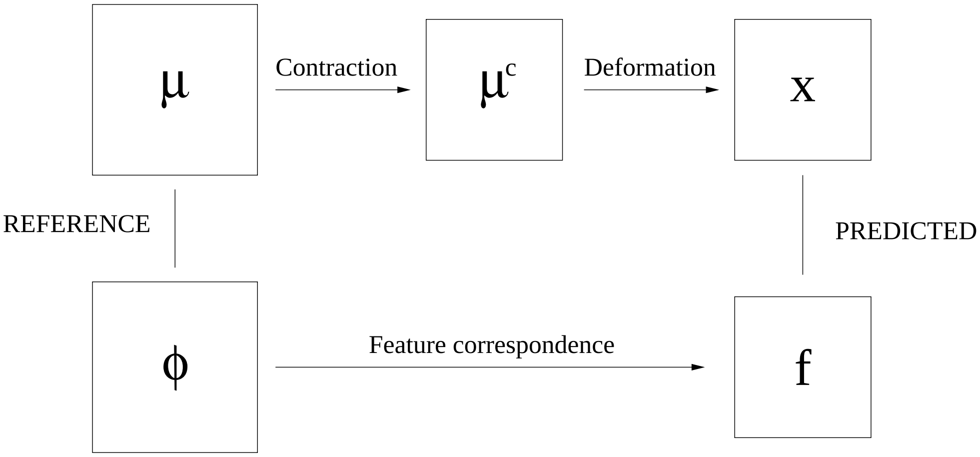

Figure 3.10: Schematic model structure with contraction modifications. The shape portion is modified in expectation according to the values of γ, whereas the image feature portion remains the same as before.

the prior on γ_2 is

$$p(\gamma_2(t)) = \mathrm{MVN}(\omega(t), \nu^{-1}\mathrm{I}_3)\ ,\ \{\gamma_{21}(t), \gamma_{22}(t), \gamma_{23}(t)\} \in < 0, \infty > \ . \tag{3.29}$$

Obviously, for a reasonable situation, γ_2 could be truncated further, particularly on the positive end. We leave the prior for γ_1 unspecified as $p(\gamma_1)$. This yields the final form for the joint shape distribution p_S as

$$p_S(x, \gamma_1, \gamma_2(t)|\theta_x) = p_S(x|\theta_x, \gamma_1, \gamma_2(t))p(\gamma_2(t))p(\gamma_1)\ \ . \tag{3.30}$$

Full Conditionals

As mentioned earlier, the most practically useful quantities in our modeling are the full conditional distributions. Here, the modality-specific corrected distributions will be considered, starting with the contraction parameter vector γ.

Consider first the center of contraction $\gamma_1 = \{\gamma_{11}, \gamma_{12}, \gamma_{13}\}$ in this model. The full conditional distribution for this set of parameters reduces to (for each direction $i \in \{1, 2, 3\}$)

$$p(\gamma_{1i}|x, \gamma_2, \theta_x) = N(\mu_{\gamma_{1i}}|\{x, \gamma_2, \theta_x\}, \sigma^2_{\gamma_{1i}}|\{x, \gamma_2, \theta_x\})\ , \tag{3.31}$$

with

$$\mu_{\gamma_{1i}} | \{x, \gamma_{2i}, \theta_x\} = \frac{x_{0i} - \gamma_2 \mu_{0i}}{1 - \gamma_{2i}} \tag{3.32}$$

and

$$\sigma^2_{\gamma_{1i}} | \{x, \gamma_2, \theta_x\}) = \sigma^2_0 (1 - \gamma_{2i})^{-1} \quad . \tag{3.33}$$

This means that its displacement only really interacts with the overall placement of the facet mean grid of positions (μ_0 only) and contraction in the image under consideration. This is an extremely vague and non-informative specification for γ_1, and so we elect to keep γ_1 fixed in the model implementation and rather introduce an empirical step, where γ_1 is estimated more directly from the evidence of the full hierarchical facet placement results for the image at hand. This will be further discussed in the section on maximization.

The full conditional distribution for the contraction factor γ_2 also is of normal form. Using standard normal methodology, write

$$- \log p(x, \gamma_2 | \theta_x) =$$

$$c(\theta_x) + \sum_l \sum_i \frac{1}{2\kappa\sigma_l^2} \left[x_{li} - \sum_{j \in P_{li}} w_{ij}^{(l)} x_{(l-1)j} - \gamma_2 \left(\mu_{li} - \sum_{j \in P_{li}} w_{ij}^{(l)} \mu_{(l-1)j} \right) \right]^2 . \tag{3.34}$$

As before, k is the index for a facet in the set P_{lj} which contribute non-zero terms to the full conditional distribution on x_{lj}. c is a constant dependent only on θ_x.

Then, re-express equation (3.34) in terms of differences defined as

$$\Delta_\mu^{lj} = \mu_{lj} - \sum_{k \in P_{lj}} w_{jk}^{(l)} \mu_{(l-1)k} \quad , \quad \Delta_x^{lj} = x_{lj} - \sum_{k \in P_{lj}} w_{jk}^{(l)} x_{(l-1)k} \tag{3.35}$$

for $l > 0$. For $l = 0$, define

$$\Delta_\mu^0 = \mu_0 - \gamma_1 \quad , \quad \Delta_x^0 = x_0 - \gamma_1 \ . \tag{3.36}$$

The sums in equation (3.35) can be interpreted as a facet's "virtual parents", that is, a non-existent facet on the level above whose motion facet lj inherits directly. From that representation, we can obtain the following for the full conditional on $\gamma_2 = \{\gamma_{21}, \gamma_{22}, \gamma_{23}\}$, for orthogonal direction $i \in \{1, 2, 3\}$,

$$p(\gamma_{2i}|x, \gamma_1, \theta_x) = N(\mu_{\gamma_{2i}}|\{x, \gamma_1, \theta_x\}, \sigma^2_{\gamma_{2i}}|\{x, \gamma_1, \theta_x\}) \ . \tag{3.37}$$

The mean and variance of this full conditional distribution $p(\gamma_2|x, \gamma_1, \theta_x)$ turns out to be a function best expressed in terms of the differences in distance just defined between facets on adjacent levels in the reference and currently considered image, namely

$$\mu_{\gamma_{2i}}|\{x, \gamma_1, \theta_x\} = \frac{\sum_{l=0}^{L} \sum_{j=1}^{N_l} \{(\Delta^{lj}_{\mu i})^{-2} (\kappa\sigma_l)^{-2} \Delta^{lj}_{xi} \Delta^{lj}_{\mu i}\}}{\sum_{l=0}^{L} \sum_{j=1}^{N_l} \{(\Delta^{lj}_{\mu i})^{-2} (\kappa\sigma_l)^{-2}\}} \quad , \tag{3.38}$$

and

$$\sigma^2_{\gamma_{2i}}|\{x, \gamma_1, \theta_x\} = \left(\sum_{l=0}^{L} \sum_{j=1}^{N_l} \{(\Delta^{lj}_{\mu i})^{-2} (\kappa\sigma_l)^{-2}\} \right)^{-1} \quad . \tag{3.39}$$

Here, $\Delta^{lj}_{\mu i}$ is the ith directional component of Δ^{lj}_{μ}.

Furthermore, if a normal distribution truncated at zero with mean ω and precision ν as a prior on $p(\gamma_2)$ is included as indicated in equation (3.29), we get a regularization of the full conditional mean and variance according to

$$\mu_{\gamma_{2i}}|\{x, \gamma_1, \theta_x, \omega, \nu\} = \frac{\sum_{l=0}^{L} \sum_{j=1}^{N_l} \{(\Delta^{lj}_{\mu i})^{-2} (\kappa\sigma_l)^{-2} \Delta^{lj}_{xi} \Delta^{lj}_{\mu i}\} + \omega\nu}{\sum_{l=0}^{L} \sum_{j=1}^{N_l} \{(\Delta^{lj}_{\mu i})^{-2} (\kappa\sigma_l)^{-2}\} + \nu} \tag{3.40}$$

and

$$\sigma^2_{\gamma_{2i}}|\{x, \gamma_1, \theta_x, \omega, \nu\} = \left(\sum_{l=0}^{L} \sum_{j=1}^{N_l} \{(\Delta^{lj}_{\mu i})^{-2} (\kappa\sigma_l)^{-2}\} + \nu \right)^{-1} \quad , \tag{3.41}$$

with $\{\Delta^{lj}_{\mu i}, \Delta^{lj}_{xi}\}$ defined as before.

Finally, the full conditional distributions for the facet placements remain very similar to the ones seen before in Section 3.1.1, (equations (3.12) through (3.19)) with the only modification of replacing the parameter vector μ with the contracted version μ^c, for example

$$\begin{aligned}\mu_{lj}|\{x_{-lj}\} = \mu^c_{lj} \quad &+ \quad \frac{\sigma^2_l}{\sigma^2_{lj}|x_{-lj}} \sum_{k\in P_{lj}} w^{(l)}_{jk}(x_{(l-1)k} - \mu^c_{(l-1)k}) \\ &+ \quad \frac{\sigma^2_{l+1}}{\sigma^2_{lj}|x_{-lj}} \sum_{k\in D_{lj}} w^{(l+1)}_{kj}\xi_{lk} \ ,\end{aligned} \tag{3.42}$$

with

$$\xi_{lk} = \left\{ (x_{(l+1)k} - \mu^c_{(l+1)k}) - \sum_{j'\in P'_{(l+1)k}} w^{(l)}_{j'k}(x_{lj'} - \mu^c_{lj'}) \right\} \ . \tag{3.43}$$

3.3 Maximization and Implementation

As described above, one goal of this methodology is to provide a computationally tractable method for estimating observed facet locations and the perceived deformation map M. This maximization can be performed relatively straightforwardly in the framework of iterated conditional modes (ICM), in which each parameter is updated by setting it to the mode of its full conditional distribution as previously described in Section 2.2.

Given the hierarchical structure of the proposed model, the natural order of maximization is fairly obvious. If we begin with maximizing the upper levels of the hierarchy, it makes sense to do the next level down next, then one further down and so on, maximizing all facet positions on one given level l before moving on to the next level $(l+1)$. When the final (bottom) level L has been reached and x_L maximized, the center of contraction and contraction factors can be updated to give a new estimate for μ^c. When seen in this light, this final portion of the model can be seen as a form of data augmentation. Here, we can think of the set $\{\mu^c\}$ as an underlying, unobservable parameter vector which corresponds to the linearly contracted version of the reference image facet representation from which the deformation proceeds. The way the model is designed, this is only inferrable after the x have been observed or estimated, making the μ^c augmentation variables. In this case, this augmentation has been reduced to $2d$ variables, where d is the dimensionality of the image class, since we treat μ^c as a deterministic function of the contraction parameters.

Since our interest lies in the facet positions x, we will treat the new image Q we are considering as a constraint on the model, thus imposing $f = f(Q, x)$, given the image Q. This follows the reasoning outlined in Section 3.1.3, and we now need to establish the implication for maximizing $p(x, f|\theta, \gamma)$ for x under this observational constraint. Following McCulloch (1998), this reduces to the distribution seen in equation (3.44) by the argument that $p(x|\theta)$ is the limit when considering a neighborhood ϵ around the line $f = f(Q, x)$ in the parameter space, as the size of that neighborhood goes to zero. Since there is no model-defined dependence between x and f, this is an uncomplicated limit to take. The resulting constrained distribution on facet locations x in an image Q is then proportional to (3.4), namely, as $||\epsilon|| \to 0$,

$$p(x|\theta) \propto p_S(x|\theta_x) p_I(f = f(Q, x)|\theta_f) \ . \tag{3.44}$$

The full hierarchical conditional distribution for facets on level L with the observational constraint taken into account can then similarly be written

$$p(x_L|x_{L-1}, \theta) \propto p_S(x_L|x_{L-1}, \theta_x) p_I(f = f(Q, x_L)|\theta_f) \ . \tag{3.45}$$

To describe the maximization in detail, let us first treat μ^c as fixed. In that case, the maximum for x under the shape distribution p_S is simply the mean μ^c. Any change in the mode for x arises from the image distribution p_I. Thus it would almost seen that one should start with the bottom level L and maximize higher levels subsequently. We have found, however, that an approximate approach which maximizes top-down works far better in terms of quickly achieving a good maximum in the global sense. To see this, revisit the concept of scale-space as described in Section 2.1.1. One of the key features of scale-space is the reduction in image intensity extrema for quantities such as the scale-space intensity or Laplacian. In other words, as the scale t increases, the number of extrema decreases. For this reason, we compose the maximization algorithm as follows. Referring to equation (3.7), maximize only a subpart of the model, with every relation below it fixed. That is, shift over all "descendants" according to the weights in A. If the simplest form in equation (3.8) is used, then the facets connected upwards through facet lj in the hierarchical scheme move over with exactly the displacement of facet lj. An illustration of this is seen in figure 3.11. One could then maximize this "fixed subtree" with contributions from p_I as

$$p_{\hat{I},lj} = \prod_{j' \in D_{lj}} p_I(f = f(Q(x;0), x_{Lj'} + \Delta x_{lj}) | \theta_f) \ , \tag{3.46}$$

where D_{lj} now represents the lowest level descendants of facet lj. This would give the following for maximization

$$\max_{\Delta x_{lj}} \{ p_{\hat{I},lj} p_S(x_{lj} | x_{(l-1)}; \theta_x) \} \tag{3.47}$$

As an approximation and to enhance computational efficiency, the maximization approach thus involves partial maximization of the upper levels with relations between lower-level facets kept fixed. It is also possible to approximate the image feature contributions based on the scale-space (Lindeberg, 1994) of the observed and reference image. To thus further reduce computational demands and also take advantage of the simpler image representation at higher scales, the scale-space at location x_{lj} is

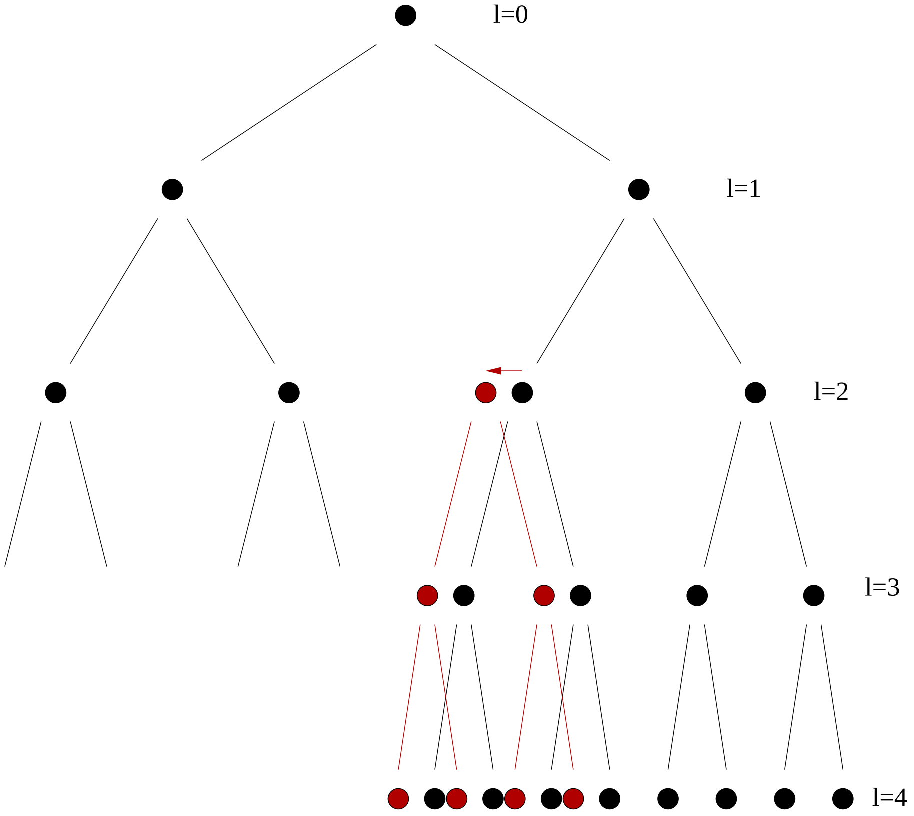

Figure 3.11: This illustrates the maximization approach used for upper-level facets. Instead of maximizing the full conditionals, maximization is based on propagating the deformation of a facet down throughout the hierarchy and thus calculating the feature contribution from all descendants.

used. In other words, we choose to maximize the facets individually according to

$$p^{mod}(x_{lj}|\theta^{mod}) = p_S(x_{lj}|x_{(l-1)};\theta_x)p_I^{mod}(f = f(Q(x;t_l), x_{lj})|\theta_I^{mod}) \ , \tag{3.48}$$

where θ_I^{mod} contains the template image-derived functions ϕ_j^{mod} on scale t_l rather than the original ϕ_j values,

$$\phi_j^{mod} = \phi(T(x;t_l), \mu_{lj}) \ . \tag{3.49}$$

Then, maximize p^{mod} directly for the x_{lj} to achieve the full conditional maximum for that level in the reduced complexity modified model. After this has been completed, we add a level and repeat the procedure, and so on. When each level is completed, we revisit higher level facets, for which we have analytical expressions as in Section 3.2 to set them to their full conditional maxima within the submodel. As we add levels, we get closer and closer to the full model, until, with level L included, we are doing ICM on the full model. From this perspective, the steps prior to the inclusion of level L can be thought of as achieving a very good starting position before applying ICM to the full model. Numerical maximization is required anytime the location of a facet is being estimated with the image contributions from p_I or p_I^{mod} taken directly into account (included in the full conditionals or approximations thereof). This is due to the fact that the p_I factor picked up in equation (3.44) introduces a non-standard distribution on x under the constraint.

When using the more isotropic A (equation (3.11)), we can no longer maximize the facets within a level l as independent due to the interaction effect described in the graphical model section. In this case, maximizing the individual facets' full conditionals correspond to an ICM step within the submodel maximization rather than an exact maximization of the reduced model. The propagation of Δx_{lj} for $p_{\hat{I},lj}$ also involves more lowest-level facets due to the wider covariance structure. See figure 3.12 for an illustration. We have found, however, that the same method using the scale-space approximation still produces a very good starting point for the full

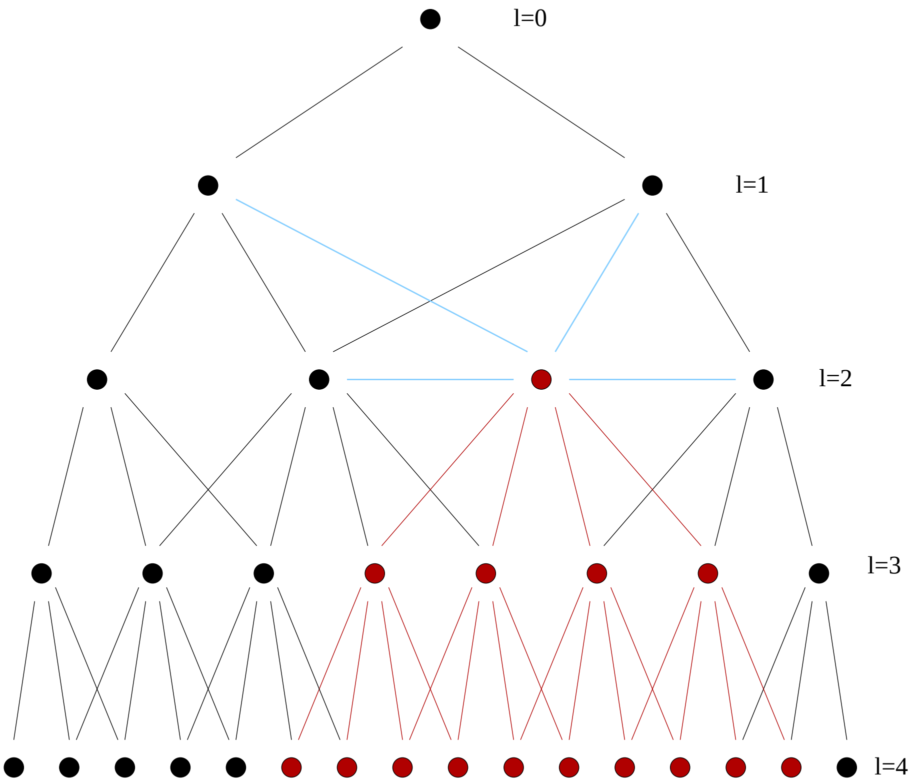

Figure 3.12: This figure illustrates the maximization of upper level facets when using the near isotropic form of the design matrix A. The red connections show facets affected when propagating the deformation of the top red facet downwards in the hierarchy through the non-zero entries w. These are the facets whose feature contributions would be considered. The blue connections indicate facets which are taken into account when looking at the full conditional shape portion of the density.

ICM. When all levels are included, the setup remains the same as before (no direct interactions have been introduced on the lowest level) except for the higher number of interactions and thus computational operations.

After each full cycle, we can set the contraction parameters γ to their full conditional mean (equal to the mode) as in equation (3.40), thus completing the full ICM cycle over all parameters to be estimated. The γ_2 can be set to their full conditional mean given in equation (3.32). The full conditional distribution on γ_1 has also been discussed, as well as some justification for using an empirical type step for its estimation. In short, the center of contraction γ_1 could be set to its conditional maximum under the model, but this is generally not very meaningful due to the overall translational near-invariance of the linear contraction modeling. Therefore, empirical estimation based on a more meaningful objective function is added as an empirical step, that is, treat the center of contraction as fixed, but estimate it under a given model outside the model framework itself. To do this, consider a perpendicular deviation $v_\perp$ from a radial contraction towards γ_1 for all lowest-level facets. First, define

$$v_x^{lj} = \{v_{x1}^{lj}, v_{x2}^{lj}, v_{x3}^{lj}\} = x_{lj} - \gamma_1 \tag{3.50}$$

and similarly

$$v_\mu^{lj} = \{v_{\mu1}^{lj}, v_{\mu2}^{lj}, v_{\mu3}^{lj}\} = \mu_{lj} - \gamma_1 \tag{3.51}$$

Then the motion of x_{lj} from μ_{lj} not explained by the contraction is

$$v_\perp^{lj} = v_x - \frac{(v_x^{lj} \cdot v_\mu^{lj}) v_\mu^{lj}}{\|v_\mu^{lj}\|^2} = v_x^{lj} - a^{lj} v_\mu^{lj} \ , \tag{3.52}$$

where

$$a^{lj} = \frac{\sum_i v_{xi}^{lj} v_{\mu i}^{lj}}{\sum_i \left(v_{\mu i}^{lj}\right)^2} \ . \tag{3.53}$$

This vector has magnitude

$$\left\| v_\perp^{lj} \right\| = \sqrt{\sum_i \left(v_{xi}^{lj} - a^{lj} v_{\mu i}^{lj} \right)^2} \ . \tag{3.54}$$

We can also constrain γ_1 to lie on a line inside the ventricle (again, as per the studies cited in Section 2)

$$\gamma_{1i}(\beta) = b_i + \beta c_i \ , \tag{3.55}$$

with b_i and c_i derived empirically from the reference image. Under this constraint, then, minimize for β the functional $b(x)$, given as

$$b(x) = \sum_{j=1}^{N_L} \left\| v_\perp^{lj} \right\|^2 \ , \tag{3.56}$$

to obtain the current estimate of γ_1 (empirically) as $\gamma_1(\hat{\beta})$, where

$$\hat{\beta} = \min_{\beta} b(x) \ . \tag{3.57}$$

For the numerical ICM maximization steps necessary when image information (p_I in some form) is included, the Nelder-Mead simplex method is applied (Nelder and Mead, 1965). This involves using an initial estimate of an $(n+1)$ dimensional vector of n-dimensional points, taken in our case to be the current estimate of the contracted reference position $P_0 = \mu_{lj}^c$ and the three point $P_i = P_0 + k_s e_i$, where e_i is the unit vector in the directions $i \in \{1, 2, 3\}$ and k_s is a constant which reflects the scale at which we are maximizing (the higher we are in the facet hierarchy, the higher k_s is set). The objective function, which is either $-p(x_{Lj}|x_{-Lj};\theta)$ or $-p^{mod}(x_{lj}|x_{-lj};\theta^{mod})$ is then evaluated at all four points, and based on the ranking of these function values, the simplex shape is altered such that the simplex is driven away from high-valued points through reflection and expansion and towards low-valued points through contraction. This method will always converge to a local minimum of the objective function. No globality of minimization (or equivalently maximization of p) can be shown, but our experience through the use of multiple starting points and sampling around the minima found, is that the method provides, at the least, a very good local extremum in the model parameter space.

3.4 Sampling

In order to learn about the stability and constraints of the models, sampling was performed based on the previously described Metropolis-Hastings algorithm (see Section 2.2.1). Referring back to the maximization scheme and the model setup, particularly the simplicity of many of the full conditional distributions, it is clear that the task of sampling is quite straightforward. The upper levels in the facet hierarchy can be sampled using Gibbs steps (the conditionals are known exactly and are possible to sample directly from). The contraction factor γ_2 is similarly straightforward to sample, with mean and variance as given in equation (3.38). The lowest level facets require a Metropolis or Metropolis-Hastings step due to the non-standard form of p_I.

Consider the full conditional distributions as stated in equation (3.45) and detailed in Section 3.1.1 and Section 3.2. Given that the shape portion of the full conditionals are normal, it makes sense to use an overly wide normal proposal density centered at the current estimate of x. We can set the proposal variance heuristically such that we get a reasonable acceptance ratio (number of accepted to proposed values for elements in x). The sampling is started at a sensible initial condition, e.g. the maximum found using the maximization technique in Section 3.3. Other alternatives considered were a fixed normal proposal centered at a reference value or a mixture of the two. These latter methods require significant Hastings corrections. Due to the high-dimensional parameter space, it is clear that we will not be able in reasonable time to probe sufficiently the entirety of that space. Rather, we use sampling to get an idea of the local characteristics of the distributions around our maximum probability estimates.

3.5 Evaluation

3.5.1 Summary Images

In order to evaluate the performance of the observer model, a few considerations must be given to which representations of the data and the estimation results should be used. For the gated image sets, summaries are often compiled to reduce the size of the data presented to the physician. One such representation is the equivalent of the standard SPECT image, that is, to sum the images voxel-wise over time. This obviously negates the entire idea of acquiring the gated dataset, but it offers a valid object to compare to when evaluating the utility of the gated method with an intelligent processing versus the standard ungated SPECT approach. The summary image which we propose for comparison is one in which the image intensities are summed back to the reference facet positions based on the estimated facet positions in the other timeframes. In other words, define the composite image S^{comp} as

$$S^{comp}(\mu_{Lj}) = \frac{1}{n}(T(\mu_{Lj}) + \sum_{t \neq t_\mu} Q_t(x_{Lj}(t))), \tag{3.58}$$

where Q_t is the t-th image in the time series of n images and t_μ is the gate used as the reference, i.e. $Q_{t_\mu} = T$.

3.5.2 Volume Comparison

Also, it is of value to compare LV chamber volumes in the different gates to that of the reference state. To do this, we apply the simple approach as follows. Assume a voxel-wise segmentation of the volume in the reference state. Then take the corresponding volume in a deformed dataset to be the sum of voxels containing facets after deformation estimation. In other words, give voxels containing facets a label $\psi_i = 1$, while all others have label $\psi_i = 0$ and calculate

$$V = \sum_i \{v\mathrm{I}(\psi_i = 1)\} \ , \tag{3.59}$$

where V is the estimated volume and v is the volume of an individual voxel element. I is the indicator function, and i is an index running over all the voxels in the image.

Chapter 4

APPLICATIONS

Three sources of data were used to evaluate the methodology for the gated cardiac SPECT modality. A clinically acquired patient dataset was used for in-depth qualitative analysis and feasibility evaluation. Two phantom data sets, one clinical and one mathematical, were then used for more accurate estimation of tissue tracking (facet placement correspondence with tissue movement), as well as for secondary qualitative evaluation.

4.1 Patient Data

The method as described was applied to a dataset from Duke University Medical Center consisting of 16 images acquired during the heart beat cycle. For each gate, an image of size 64x64x16 voxels was acquired (7.1 mm voxel size). The heart is contained entirely in a 16x16x16 voxel volume. Using a 5 level facet hierarchy, each voxel in the heart volume contains one bottom-level facet, located at the voxel centers. The entire hierarchy spans a 16^3 cube at five different resolutions ($L = 4$) and has a total of 4,681 facets. Gate 8 (filling phase, mid-diastole) was used as the reference image throughout. The resulting density on facet locations x was then maximized for each non-template gate individually. The local regression feature and near-isotropic shape model offered the best modeling performance. Typical maximization time was approximately 3 minutes per gated image on a DEC 433au workstation.

4.1.1 Qualitative Results

The results are summarized as follows. Plots of estimated facet movement from the reference image to another gated image are shown as a 3D vector field together with two-dimensional slices for further visualization. Individual facet positions are also considered and displayed on the template and observed images to demonstrate the deformation achieved under the model. We then display a composite image and compare it to the traditional SPECT image and a single gated image. Difference images for the composite versus the traditional and gated image are also shown. Subsequently, estimated changes in template size are shown for the time series, as well as estimates of the center-of-contraction location through the timeseries. Distribution characteristics, convergence and stability relative to initial condition are also briefly examined.

Facet Motion

Overall motion fields are instructive in that they give a good indication of whether the method can capture the gross motion of the heart as a relatively smooth and sensible vector field. The methods applied did indeed show such behavior, and some example fields are shown in Figures 4.1 through 4.5. Figure 4.1 shows the full estimated deformation field in the gate 3 image for the next to lowest level of facets using gate 8 as the reference. Gate 3 corresponds to the most contracted state (end-systole). The general contraction from diastole (gate 8) to systole is clearly captured by this ICM estimated mode of the joint density on facet locations. This is shown even more clearly in Figures 4.2 through 4.4, where, for clarity, facet displacement vectors are shown for representative two-dimensional slices in three orthogonal directions.

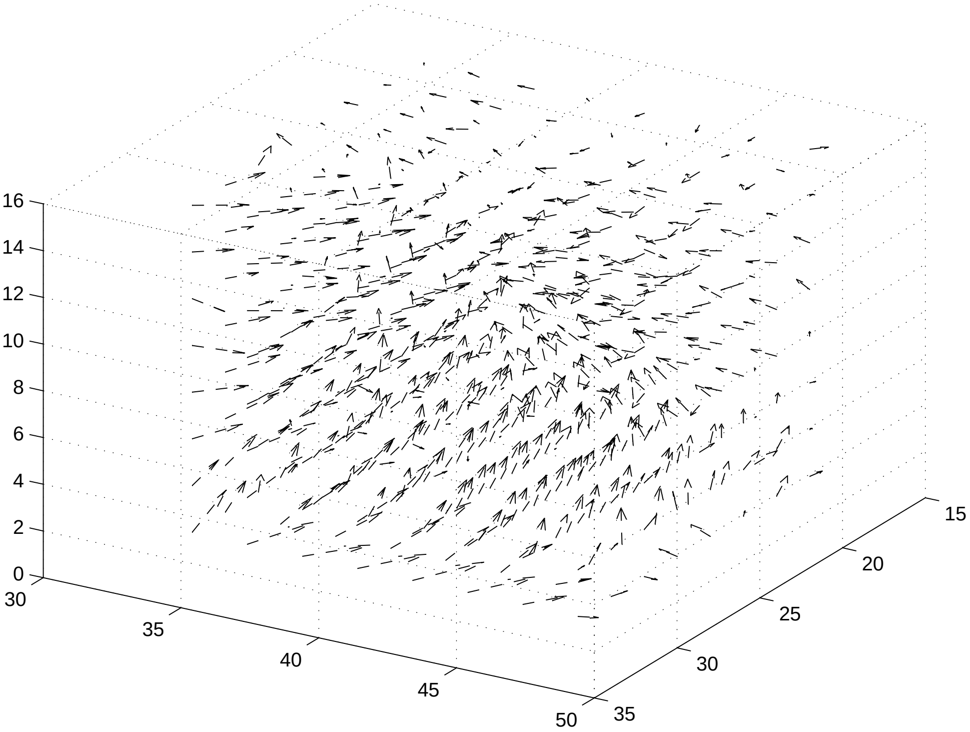

Figure 4.1: Facet motion from gate 8 (reference;mid-diastole) to gate 3 (end-systole) in three dimensions. Only level $l = 3$ (next to lowest) facets are shown to allow for a better display of overall motion. Detailed motion on the lowest level is shown in Figures 4.2 through 4.5

The grid has contracted on average, but there are also regions of the heart which have not moved significantly. This demonstrates the method's ability to capture the nonlinear deformation seen in the heart contraction as well as the overall change in expected heart size. The motion field seen is consistent with typical heart motion from mid-diastole to end-systole.

Next, the same type of figure is shown in Figure 4.5 for the deformation based on estimated observed facet positions in gate 11. This gate corresponds to the state of maximum dilation (end-diastole). This is in the period of the heart cycle when the heart is relatively quiescent, that is, oscillating slightly in size in its filled state. Again, reasonable motion estimates can be seen. Particulary, it is important to note that most parts of the heart are shown as not having moved, agreeing with visual evidence and known cardiac motion theory. Since this is a state very close to the reference, this is correct, and we can thus have confidence that the model does not overly encourage deformation. Only minor motion is detected, consistent with the quiescence of the heart at this stage in the cycle.

To further investigate the motion sensibility of the facet placements under the model, the positions of several individual facets in the reference image and gate 3 are shown in Figures 4.6 through 4.10. Points on the heart are estimated under the model to deform in a complex manner, and are consistent with the heart shapes seen. No abrupt discontinuities seem to have been introduced artificially by the model. Note also the fully 3-dimensional nature of the deformation as evidenced in the slice-jumps of several sections, particularly in Figures 4.6 and 4.8. The top part of the heart is seen to shift significantly downwards, whereas only mild upwards motion is seen for the bottom. This is consistent with known heart behavior and inspection of the image series. The same slices and points are shown for the adjacent timepoint, gate 4, to note consistency across time in facet position estimation for these two very similar

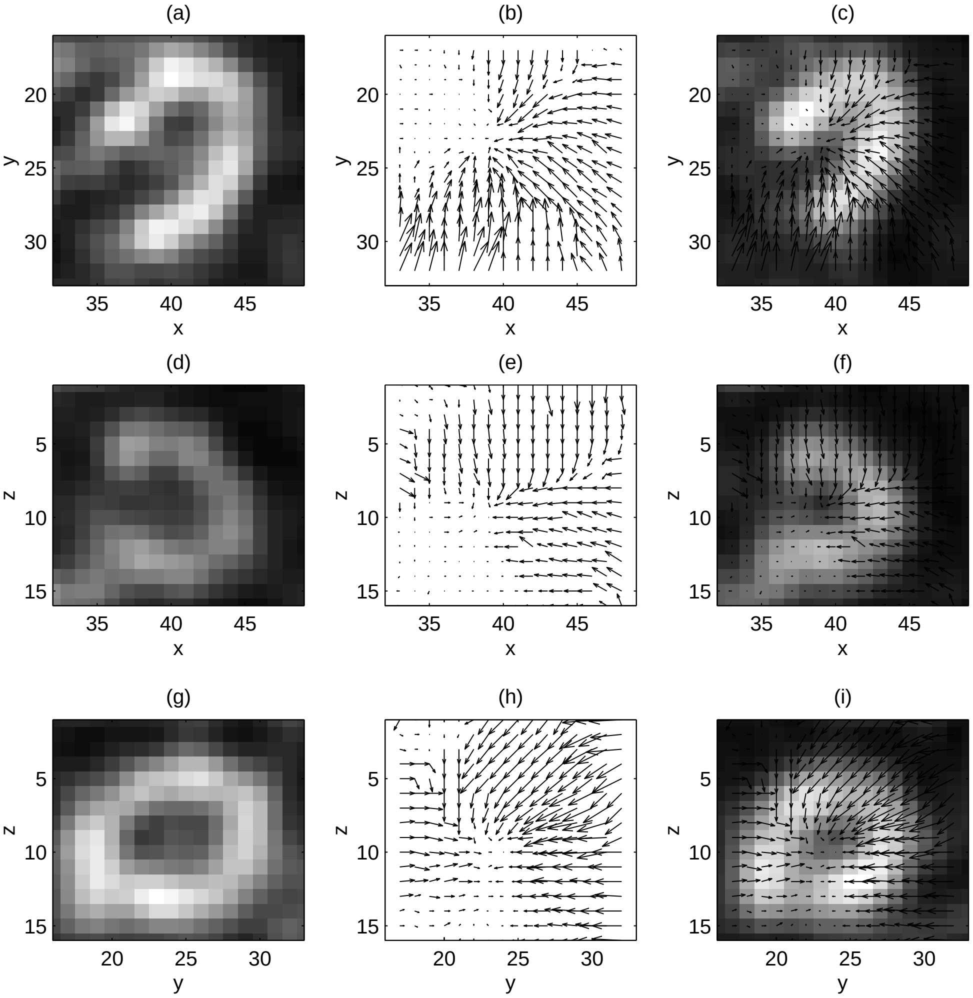

Figure 4.2: Facet motion from gate 8 (reference;mid-diastole) to gate 3 (end-systole) in perpendicular slices with intersection point at the center of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 3 image. Note the overall contraction as well as the non-uniformities, driven by the variation in image intensities to capture the actual deformation.

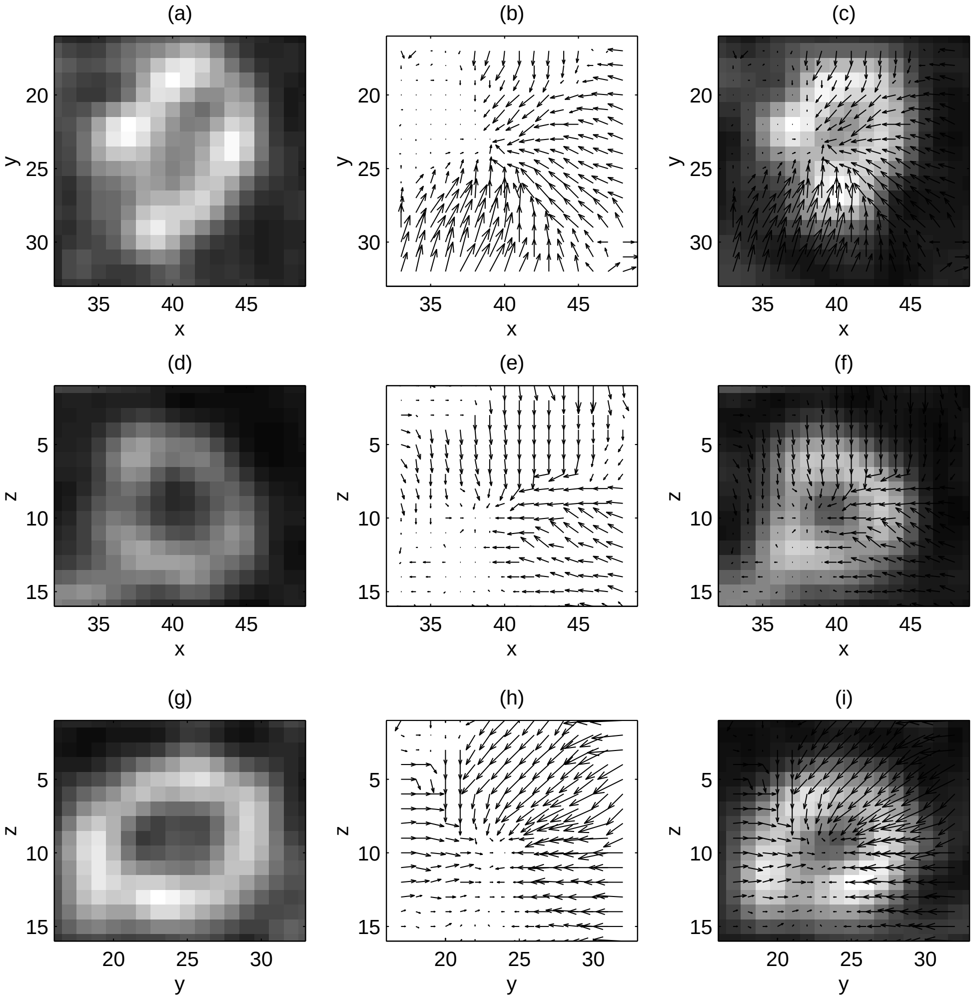

Figure 4.3: Facet motion from gate 8 (reference;mid-diastole) to gate 3 (end-systole) in perpendicular slices with intersection point towards the top of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 3 image. Note especially the similarity with Figure 4.2, showing the cohesiveness of the entire motion field, as well as the local differences.

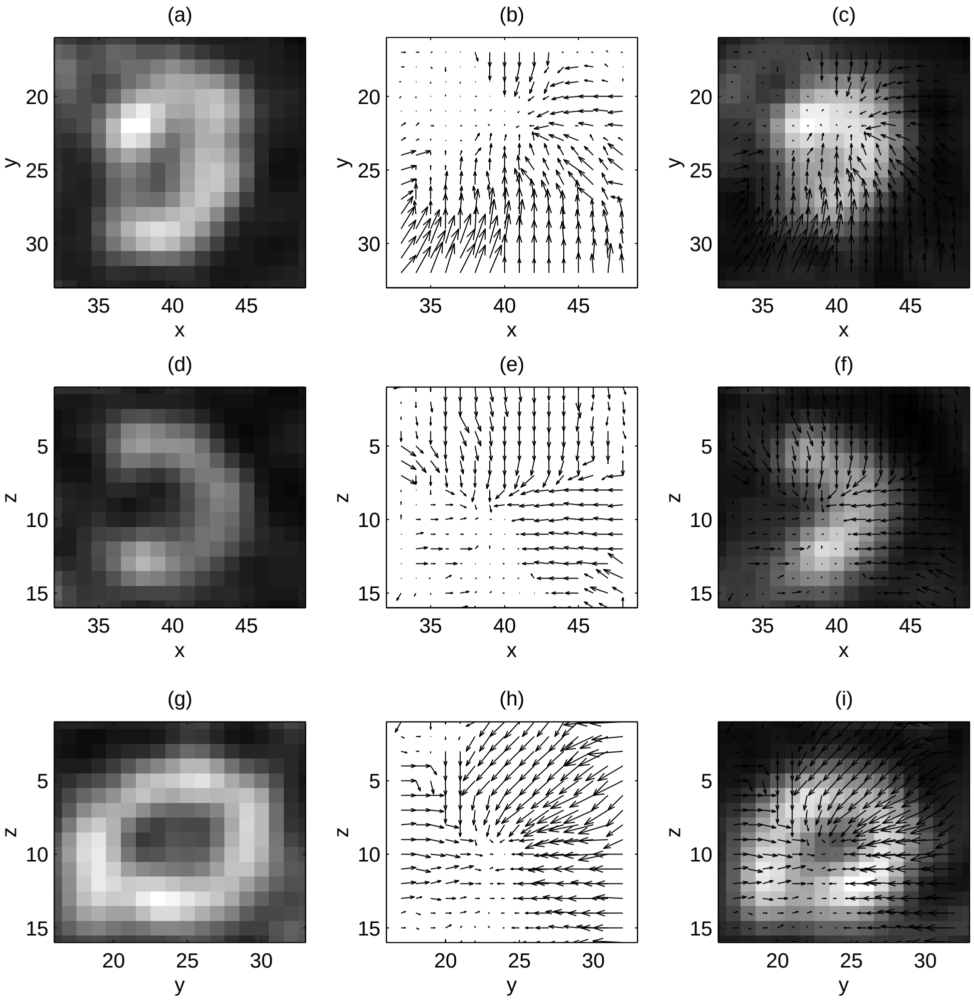

Figure 4.4: Facet motion from gate 8 (reference;mid-diastole) to gate 3 (end-systole) in slices with intersection point towards the bottom of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 3 image. Again, note the global similarities and local differences compared to Figures 4.2 and 4.3 to get an idea of the whole three-dimensional deformation volume.

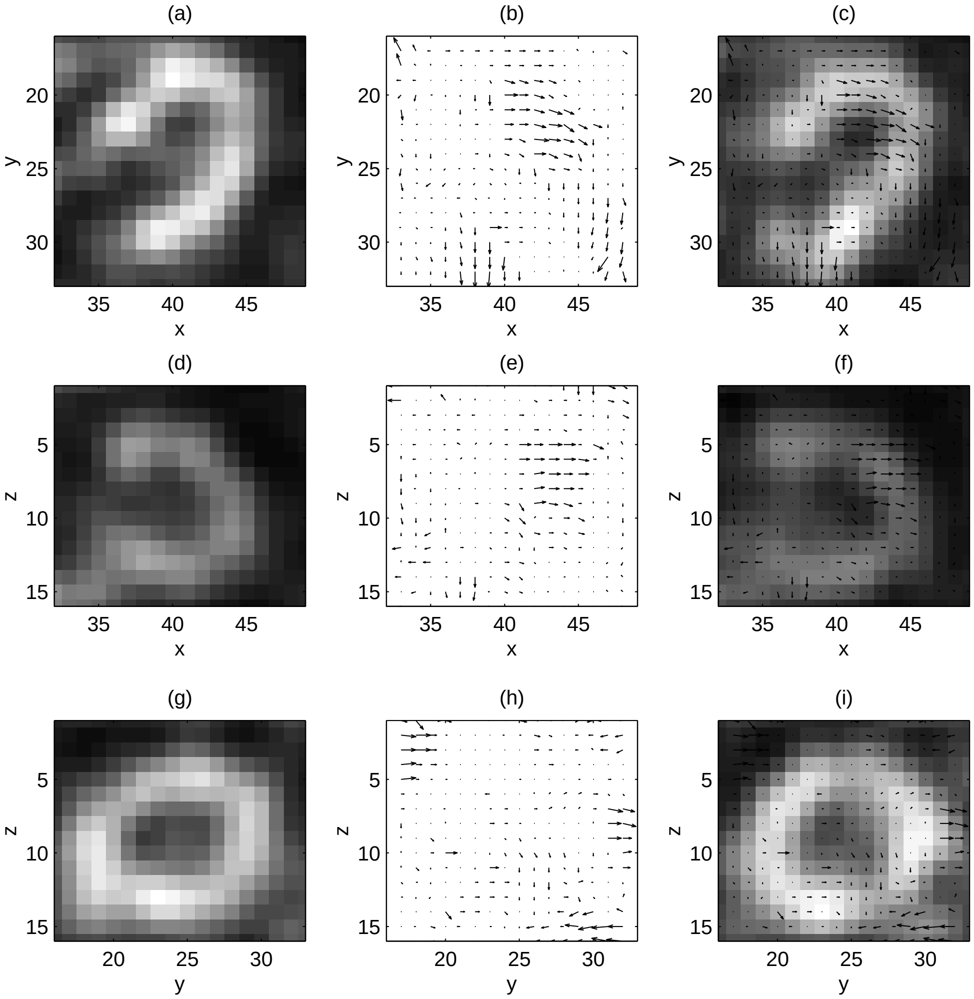

Figure 4.5: Facet motion from gate 8 (reference;mid-diastole) to gate 11 (late-diastole) in perpendicular slices with intersection point at the center of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 11 image. This illustrates that the model is stable for configurations near the template in shape, not over-encouraging motion.

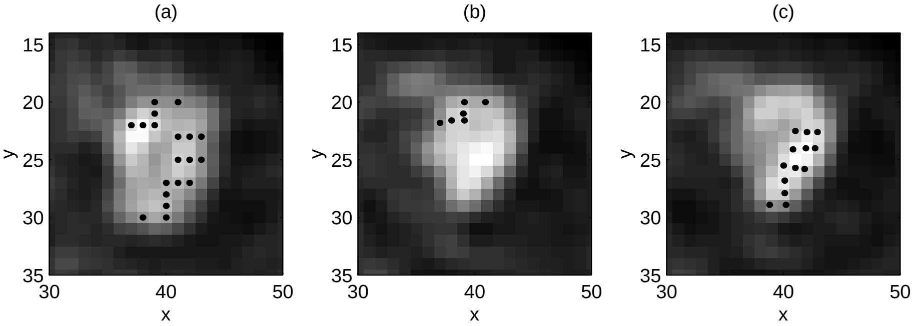

Figure 4.6: Selected facet positions in transaxial slice 5 of the reference image (a), with corresponding estimated facet positions in the gate 3 image, slice 5 (b) and 6 (c). Note the contraction and deformation in relative positioning without losing relations between neighboring facets, as well as the slice-jump, indicating fully 3D deformation.

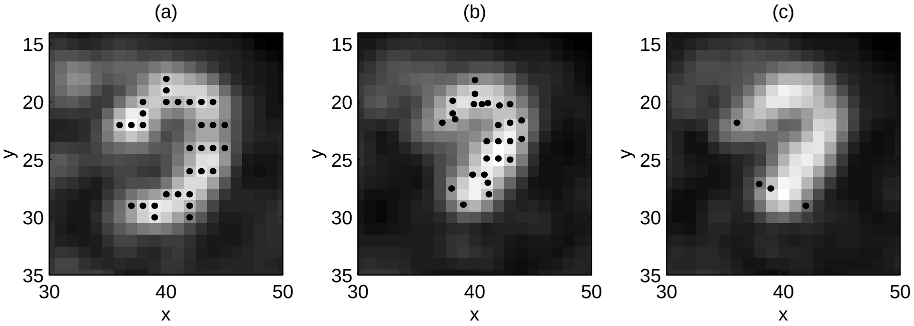

Figure 4.7: Selected facet positions in transaxial slice 7 of the reference image (a), with corresponding estimated facet positions in the gate 3 image, slice 7 (b) and 8 (c). As we move down towards the the center of contraction in the volume, we see less slice-jumping.

states, as seen in Figures 4.9 and 4.10.

Finally, results are shown to illustrate the local characteristics of the distribution being maximized. In Figure 4.11 is shown the 50 % highest probability region projected onto the maximum probability slice in the z-direction. This shows results from running the sampling scheme for 5,000 iterations, starting at the density maximum estimate. We see how the density, which is circular in the shape portion alone, has been shifted in location and deformed to fit the intensity pattern seen in the new image.

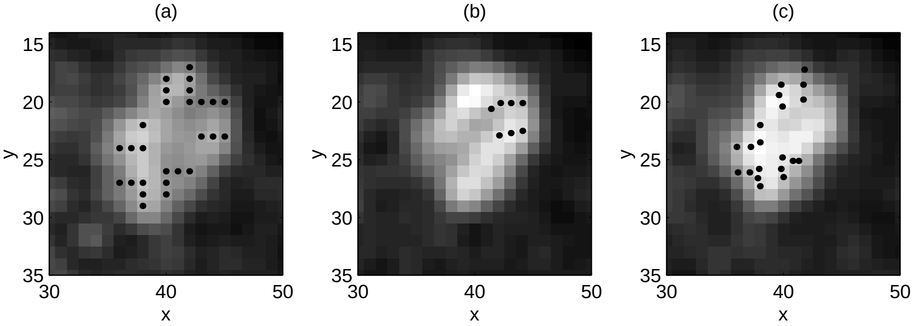

Figure 4.8: Selected facet positions in transaxial slice 11 of the reference image (a), with corresponding estimated facet positions in the gate 3 image, slice 10 (b) and 11 (c). Now we are past the mid-point, and slice-jumps (if slight) occur upwards rather than downwards.

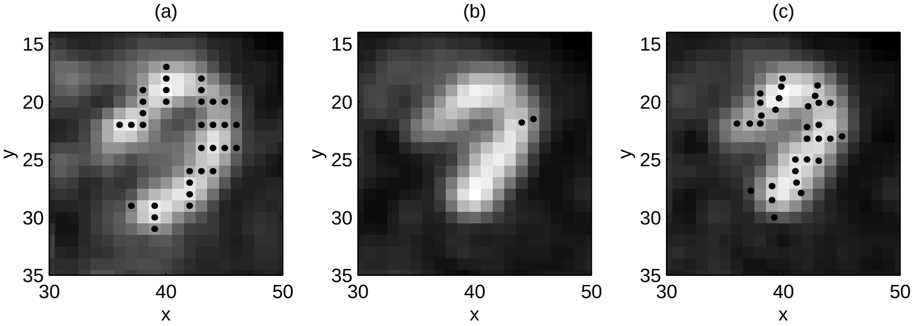

Figure 4.9: Selected facet positions in transaxial slice 9 of the reference image (a), with corresponding estimated facet positions in the gate 3 image, slice 8 (b) and 9 (c). This figure should be seen together with Figure 4.10 and shows the consitency in facet placement estimates between adjacent gates in the series.

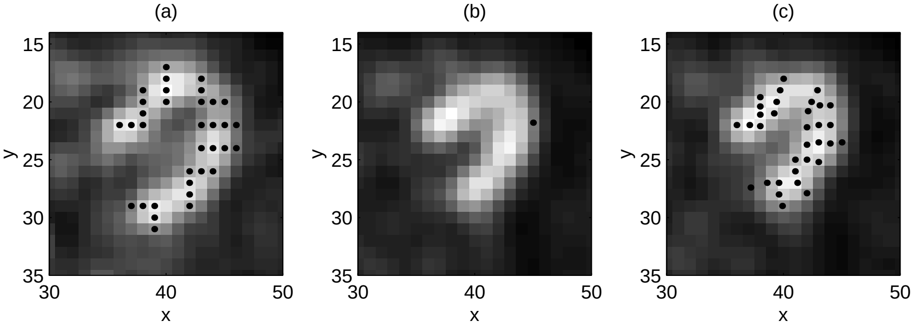

Figure 4.10: Selected facet positions in transaxial slice 9 of the reference image (a), with corresponding estimated facet positions in the gate 4 image, slice 8 (b) and 9 (c). These are the facet position estimates for the gate adjacent to that seen in the previous figures, and shows (relative to Figure 4.9) that the estimation stays consistent for close times.

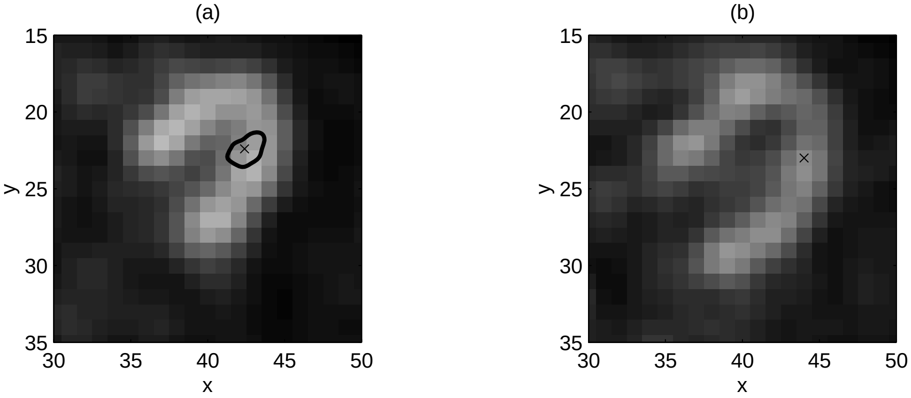

Figure 4.11: A representation of the 50 % highest probability region for predicted facet placement in the gate 3 image. The contour is formed by projecting the 3-dimensional contour onto the maximum probability slice in the z-direction. (b) shows the reference position and (a) shows the sample highest probability region with the maximum predictive probability estimate indicated

Contraction

With a relatively vague normal prior distribution on γ_2 (high variance, truncated such that $\gamma_{2i} \in < 0, 3 >$), the contraction was estimated in the three independent directions. Figure 4.12 shows these three parameters' estimated time evolutions, as well as the overall resulting contraction correction as a function of time (gates) when taken as an overall volume change (product of the three parameters). The parameters behave very sensibly through the cycle: Gates 1–7 comprise the relatively short contraction phase (systole), while the remaining images are acquired in the expanded or dilated state (diastole) of the heart. The parameter time evolution tracks this, with some variability in the individual directional contraction parameters, as is reasonable, considering the blurry nature of this image type and the complexity of the true contraction. Due to the discrete nature of the image data, a certain range of γ_{2i} is undistiguishable. That is, unless γ_{2i} causes facet means to change which voxel actually contains the facet, the data cannot help us distinguish between two close values of γ_{2i}. The more important clinically relevant quantity, overall volume, behaves exactly as expected, however. Since the contraction factors reflect an overall

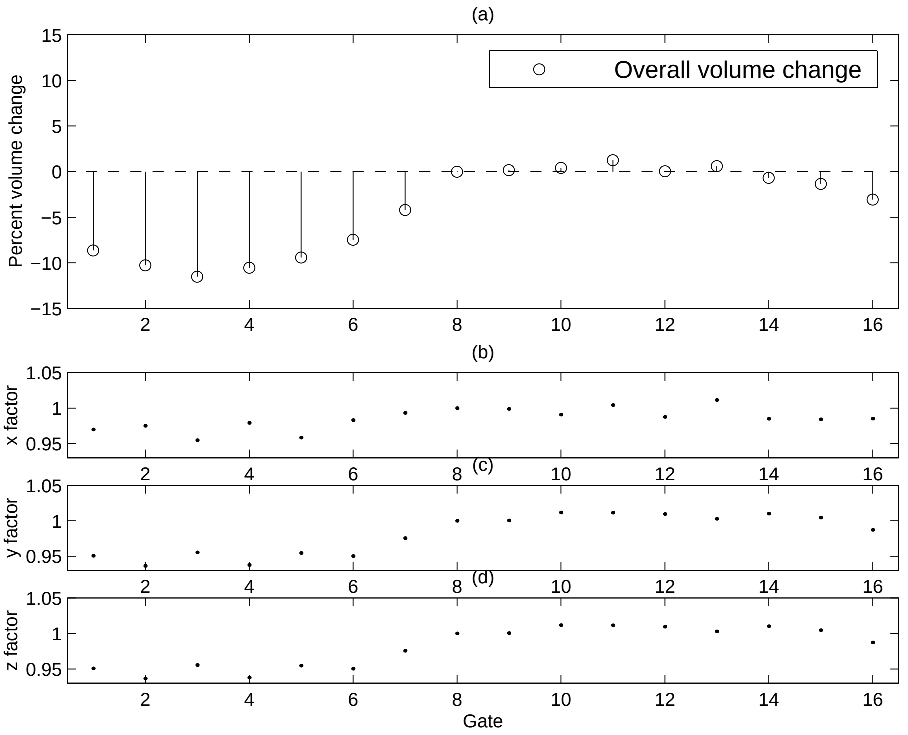

Figure 4.12: Estimated overall contraction through cycle, measured in percent of total volume spanned by μ_c relative to μ. This corresponds well to the expected and observed heart contraction through the series of images. It is not interpretable as heart chamber volume, however

correction rather than a direct estimate of, say, the heart chamber size, they can be interpreted as a rough indicator of relative heart size. They should not, however, be used as a measure of particular quantities, such as LV chamber volume. To perform this estimation, segmentation or identification data (which facet is in the chamber in the reference) is necessary. In addition to agreeing with the a priori known behavior, the general trend shown in Figure 4.12 matches well with visual inspection of changes in heart size over the image series.

Furthermore, the characteristics of the distribution being maximized were probed through sampling (jointly with the facet position estimates seen in Figure 4.11). 1,250 iterations of the sample chain and a histogram density estimate is shown in Figure 4.13. We see that, while having wide support, the maximum estimate for the

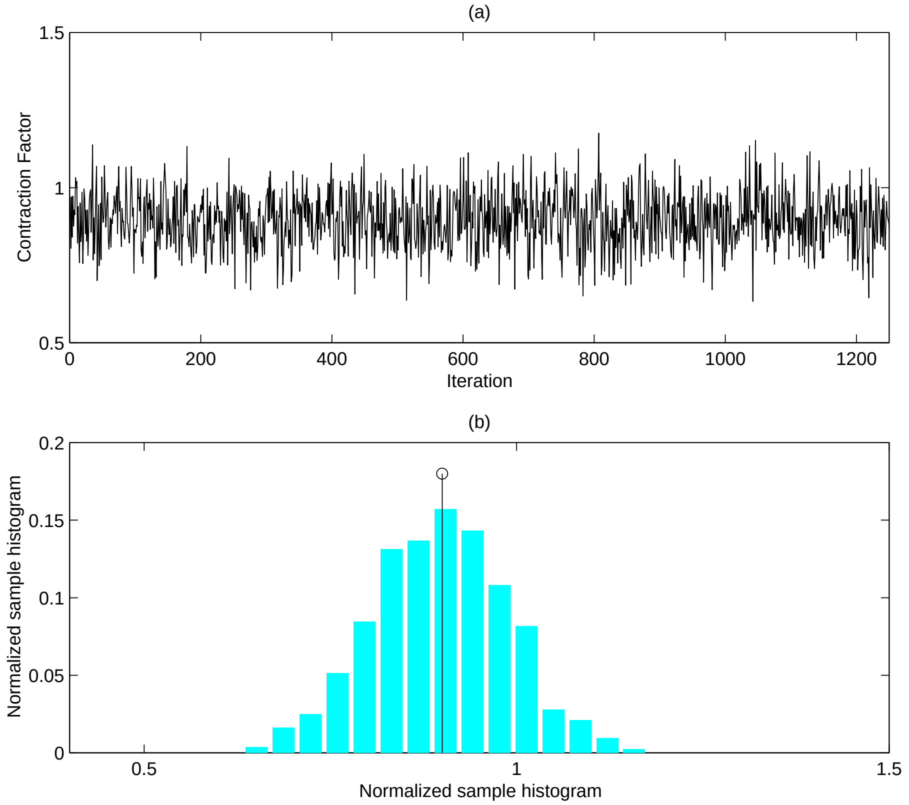

Figure 4.13: Results from the sampling chain for one of the contraction parameters from estimation of facet positions in gate 3. (a) shows a portion of the sample chain and (b) shows the resulting histogram density estimate for that parameter with the maximum found through the maximization technique indicated with a line.

contraction factor is clearly defined.

The center of contraction γ_1 similarly displays sensible behavior. In Figure 4.14 is shown the empirically estimated z-location γ_{13} (guided by $\hat{\beta}$) for the gates prior to the reference (rapid contraction and expansion). The evolution is exactly as indicated in the literature; up towards the middle of the heart in the quieter phase (end-contraction, end-expansion) and lower down in the rapidly changing phases (mid-contraction, mid-expansion).

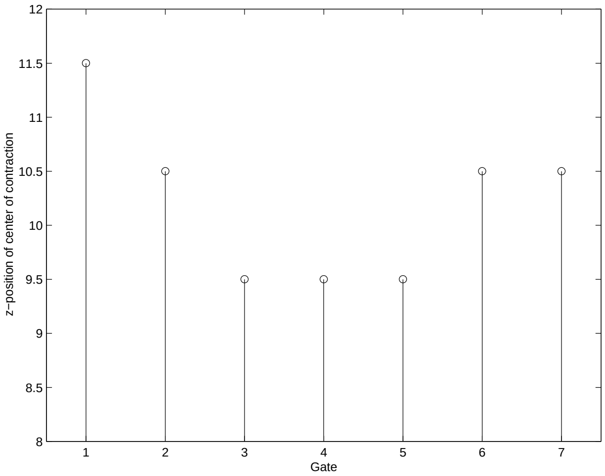

Figure 4.14: Estimated center of contraction z coordinate (γ_{13}) for the rapid contraction-expansion phase of the heart beat cycle, discretized to which slice of the dataset it resides in. The center of contraction moves down along the LV long axis in the most rapidly changing gates and resides at approximately the LV center when the size is more stable (at maximum contraction or expansion). This behavior matches that of the literature.

Summary Images

To more accurately represent the distribution of radiotracer uptake in the heart, a facet-composite (composite for short) image was calculated. As described in Section 3.5.1, the maximization for facet placement in all gated images was consolidated into this composite image by mapping the gated image intensity found at the facet position in each image to that facet's reference image position and averaging those intensities across the image sequence.

A traditional ways of comparing such representations of the data is to look at one-dimensional profiles from the volume. However, we find that this representation is relatively uninformative visually. To make comparisons about shape and intensity patterns, we therefore choose to display two-dimensional slices from the volume, which we find to offer a better tool for inspection. When looking at the results this way, the composite image (Figure 4.15(a)) compares favorably to both of the other representations of the data, the voxel-wise mean (standard SPECT equivalent as

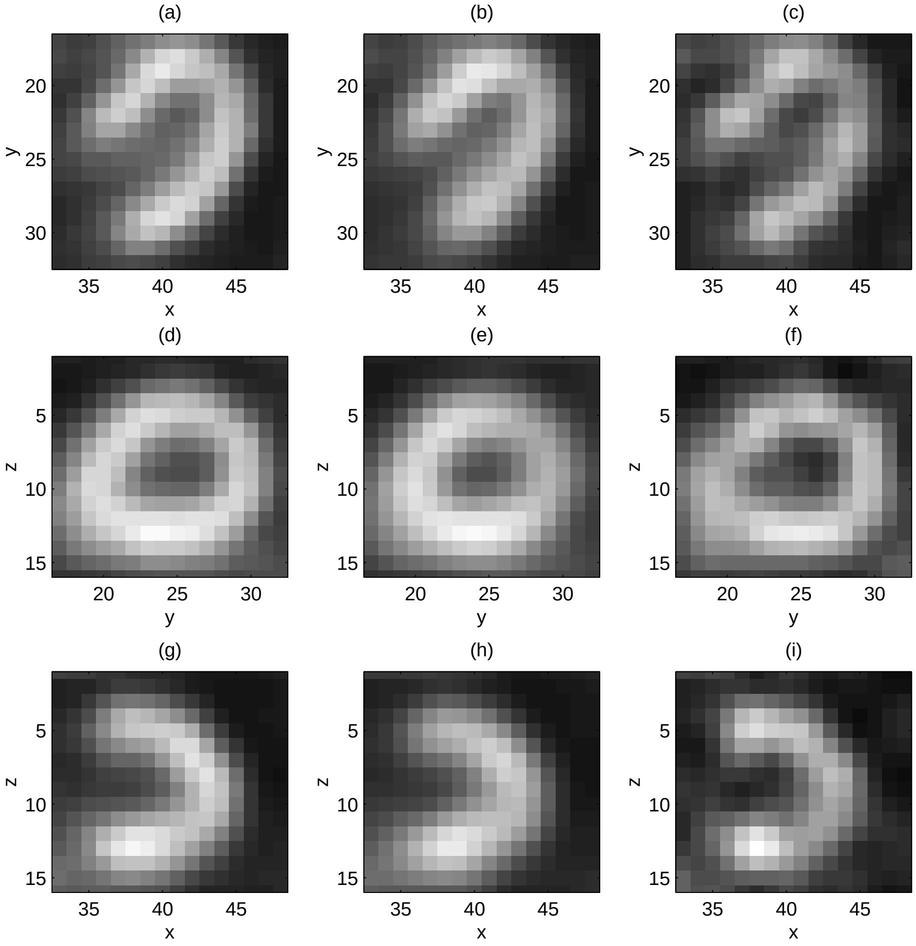

Figure 4.15: Perpendicular slices from three representations of the data. The intersection point is at the center of the LV chamber, $(40, 22, 9)$. (a)-(c) show the transaxial view, (d)-(f) show the coronal view and (g)-(i) show the sagittal view. The representations shown are : (a),(d),(g) Composite image mapped to gate 8, (b),(e),(h) mean voxel-wise image (standard SPECT), and (c),(f),(i) gate 8 image only. The composite image was formed by computing average intensity based on the facet motion through the series of images and mapping back to the reference image (gate 8). We observe similarity between gate 8 and composite, with the composite having superior smoothness in regions of activity. Furthermore, a better spatial delineation of the heart wall in the composite image relative to the standard SPECT image is seen

seen in Figure 4.15(b)) and the gated image alone (as seen in Figure 4.15(c)). Image intensity uniformity has been improved in the LV wall region relative to the gated image alone, while retaining image contrast. Comparing the facet-composite with the mean image shows a better delineation of the lateral wall of the left ventricle. The composite image thus represents a specific state of the heart (here, it maps to mid-diastole) rather than a time-averaged state which does not exist. The difference images shown in Figure 4.16 highlight the structural differences between the summary images. The mean image minus the composite image (voxel-wise difference) shows a clear pattern (dark and bright) that corresponds to the lateral wall of the left ventricle. Again, this corresponds to known heart motion. Also, when the composite image is subtracted from the gate 8 reference image, no pattern other than the known intensity difference between an average and any individual gate late in the time series is apparent. (Early gates tend to have more intensity due to an out-of-phase blurring effect which worsens towards the end of the beat cycle because in acquisition, the start of the series is more accurate than the partitioning.)

Stability

Each full ICM cycle includes an iterative maximization over facet locations x, followed by a maximization for γ_2. The model was allowed to run for 200 such cycles, and did not exhibit any significant changes in parameter estimates or facet locations from the values determined with a shorter run (5 full cycles) when initialized at a reasonable starting point. Previous work (McCulloch, 1998) has reported fast convergence of the maximization for a model which does not incorporate contraction (γ) directly. Several starting positions for the maximization routine were used without changing the final results, enforcing our belief that the ICM method applied to this problem finds, if not the global maximum, then a sensible and very good local maximum in p.

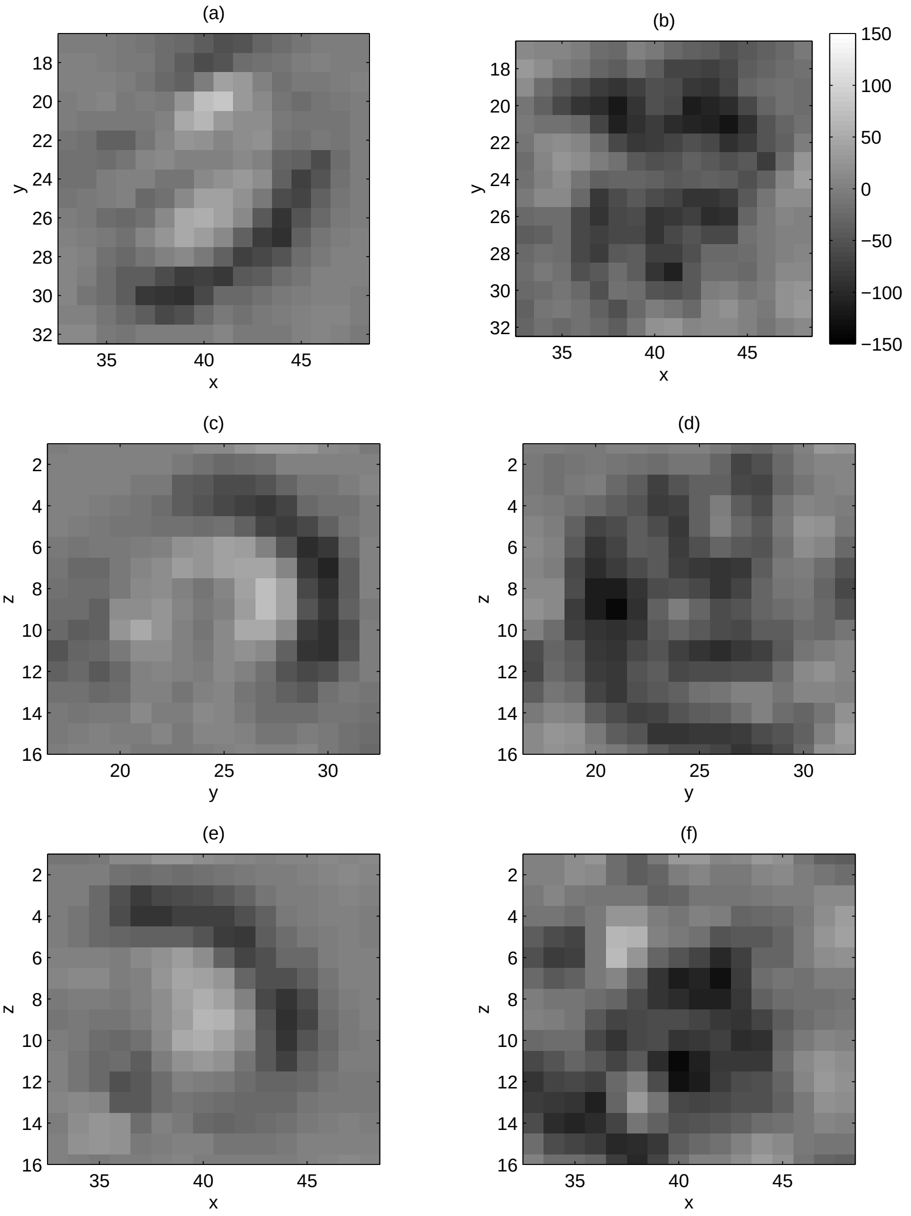

Figure 4.16: Difference images corresponding to the images in Figure 4.15; (a),(c) and (e) show mean minus composite (mapped to gate 8), and (b),(d) and (f) show gate 8 minus composite. For detailed explanation of the composite image, see section 3.5.1 . Here we see clearly the structural difference between the standard (mean) image and the composite facet-based image. The regions of dark and bright indicate that the deformation model has shifted intensity outward for the lateral wall of the left ventricle. This is in accordance with the use of gate 8 (mid-diastole) as the reference. The gated versus composite comparison shows no structural differences other than an overall intensity level difference in the heart region, which is attributable to a known intensity trend discussed in the text

4.1.2 Quantitative Results

Since this is a clinical dataset, it is impossible to ascertain the absolute motion of portions of the heart tissue, as that would involve invasiveness which is clinically unacceptable. Therefore, we can not numerically measure the accurateness of the individual facet motion in the clinical dataset. Since we are modeling the perception process, this would give us important information on how good of an observer this method represents, and thus about its possible clinical usefulness. To get some measure of performance, we performed a manual observation task on the data from both the reference and the contracted state of the heart. We then performed a standard clinical volume calculation from the model motion estimates to see if it fell within a clinically reasonable range.

Volume Estimates - Ejection Fraction

For the clinical dataset, then, we performed a manual volume segmentation of the left ventricle in the reference image (gate 8). We calculated the new volume based on the new facet position estimates, such that only those voxels containing facets were counted as being in the volume.

Using this method resulted in the volume shapes shown in Figure 4.17, which again are very reasonable based on knowledge of the heart motion. Again, we see the stationarity of the inner LV wall and the motion down and in of the outer wall. The ratio of volume change from dilated to contracted state relative to the dilated state was then calculated as

$$e = \frac{(V_{dilated} - V_{contracted})}{V_{dilated}} \; . \tag{4.1}$$

This quantity e corresponds to what is clinically known as the ejection fraction, a tool much used by clinicians as an indicator of the severity of a heart problem. Obviously,

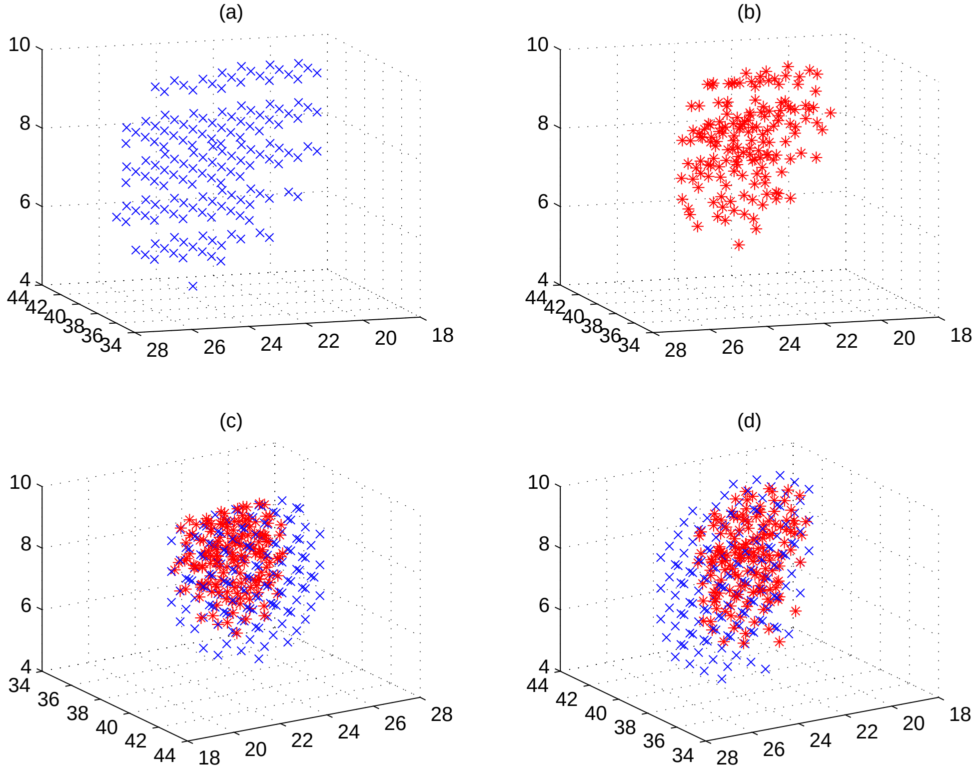

Figure 4.17: Reference and estimated positions for the facets manually segmented in the reference image as being inside the left ventricular chamber. (a) The segmented facets in the reference image. (b) The same facets' positions estimated in the gate 4 (end-systole) image from the same view. (c) and (d) The data from (a) and (b) superimposed, shown from two different and almost opposing angles to demonstrate the change in shape. We see that the chamber has been estimated to have shrunk substantially, and that this is due mostly to the motion of one side of the chamber (the outside wall).

the lower the ejection fraction, the worse the heart function. For the dataset in question, the ejection fraction from the facet motion estimates comes out to be 47 per cent, which is very reasonable for a patient with a heart condition. A “normal” ejection fraction is up around 70 per cent. This volume calculation also agrees within 3 per cent with the same calculation based on manual segmentation of the contracted volume.

4.2 Mathematical Phantom Data

The method was also applied to a dataset generated from a mathematical phantom representing the beating heart in a human thorax. Using three-dimensional hyperellipsoids and the anatomical and functional reference literature for the human and cardiac physiology and function, Peter et al (1998) have developed a realistic phantom which can be used to generate a variety of data. Each object within the thorax is controlled by its own set of shape parameters, and so a variety of anatomies and heart states can be modeled computationally. In particular, it can be used to simulate a gated cardiac SPECT scan, by imposing the geometry of the dynamic phantom at several timepoints in the cardiac cycle as the basis for a Monte Carlo simulation of SPECT emission, attenuation and detection. By generating individual photon histories, the two-dimensional projections as mentioned in Section 2.3 are formed, which can then be reconstructed into the full three-dimensional images exactly as in the patient data case. The data was generated for 32 gates as a 128x128x128 voxel representation, of which we used every other gate, a total of 16. The spatial resolution of the simulation was such that a 32 x 32 x 32 cube of voxels contained the entire LV chamber. This is approximately double the resolution (along each dimension) compared to the clinical data in the previous subsection. Note that the simulation data shown here is very high-count, making the noise characteristics superior to those typically seen for a clinical gated dataset. They offer a good test for the deformation aspects of the model, however, due to our more detailed knowledge of the underlying truth.

The methods were applied as before, but now with 6 levels in the hierarchy to once again provide one lowest-level facet per voxel in the reference image in a 32^3 cube configuration. This yields a total of 37,449 facets in the model. Arbitrarily, the reference was taken to be the first image in the series, gate 1 (expanded state). Maximization was carried out as described in Section 3.3, using the local regression feature function and the near-isotropic shape specification.

4.2.1 Qualitative Results

Facet Motion

First we investigate the overall motion pattern, displayed through the same type of results and illustrations as seen for the patient data. This is done to demonstrate consistency in behavior between the two sets of data as well as for visual inspection of the estimation. We elect to look at the motion fields for the next to lowest level of facets, since there is simply too much information on the lowest level to display effectively across the heart as a whole in single figures. Looking at these facets for the contracted state (end-systole) as seen in Figure 4.18 indicates very sensible estimates for the maximum probability placement of the facets. They correspond well to visual inspection and expected contraction behavior. As before, note the overall contraction with significant areas of non-uniformity. It is also worthwhile to mention that the fields in this case are much smoother than in the previous case, as should be expected since the contraction is modeled as hyperellipsoids changing smoothly in size from frame to frame, and also due to the better identifiability of the heart thanks to the noise characteristics of the data.

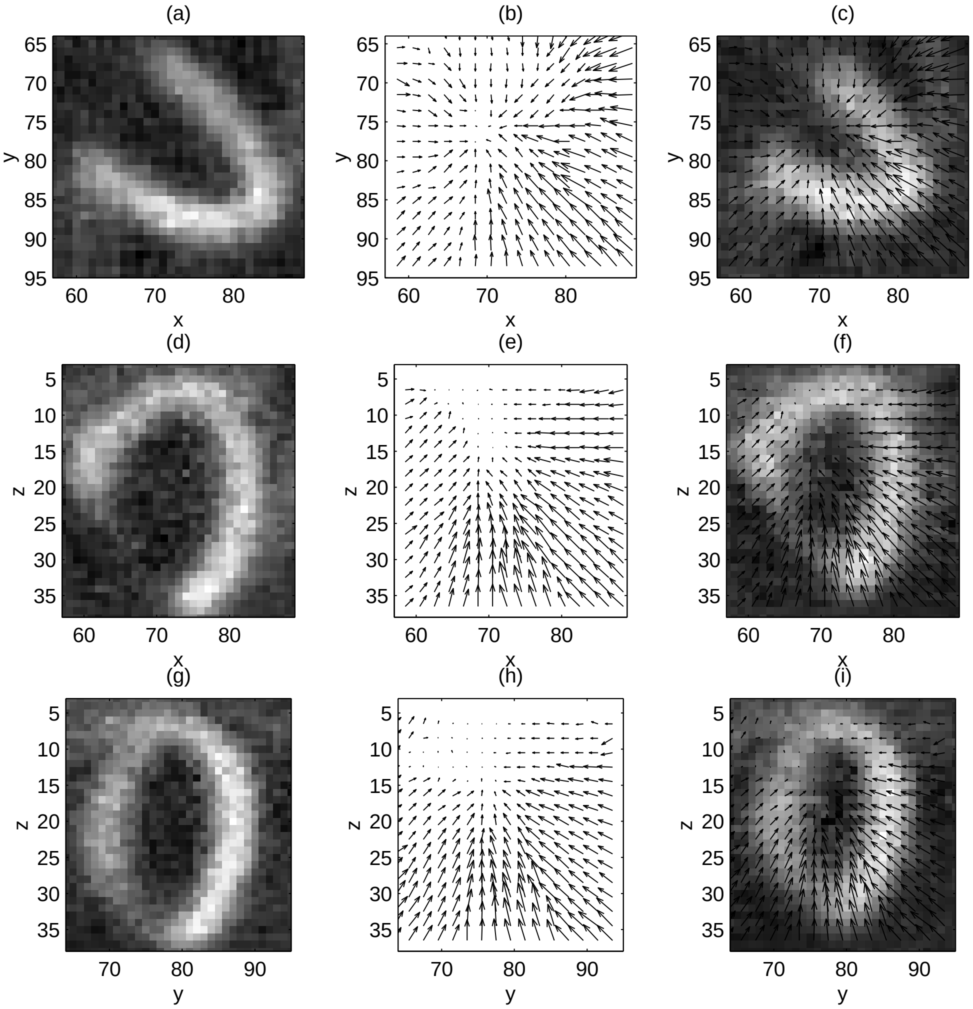

Figure 4.18: Facet motion from gate 1 (reference;expanded state) to gate 19 (end-contraction) in perpendicular slices with intersection point at the center of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 19 image. The estimated positions for the level $l = L-1 = 4$ was shown in order to be able to display estimates across the entire heart.

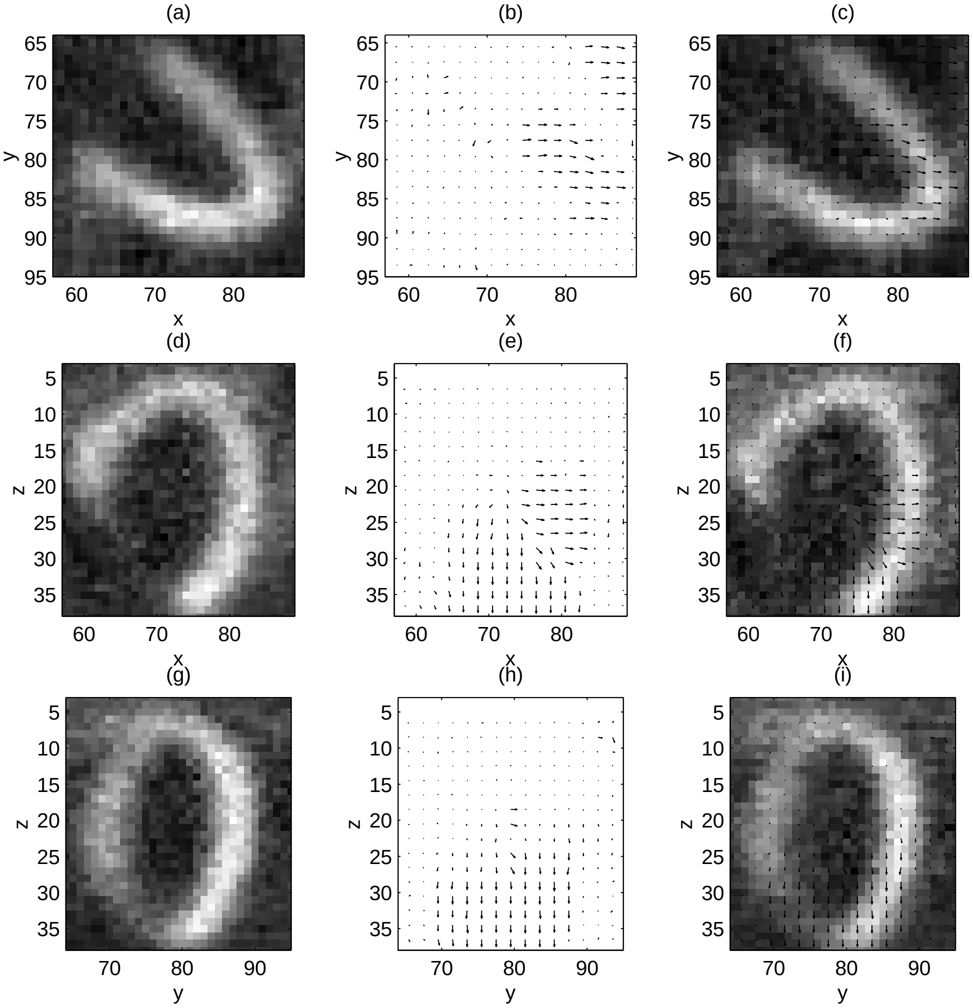

Figure 4.19: Facet motion from gate 1 (reference;expanded) to gate 11 (start-contraction) in perpendicular slices with intersection point at the center of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 11 image. We observe that the facets are estimated as having positions only slightly differently positioned from the reference.

Figure 4.19 shows the same for the gate in the series corresponding to the time-point just before contraction begins. This corresponds to a slightly expanded (dilated) state of the heart. We again see that the model estimates the slight expansion while keeping most of the facets stationary. We thus see that the method is able to capture variation of magnitude and pattern in the deformation just like in the patient data case.

An intermediate state during the expansion phase (after maximum contraction) is seen in Figure 4.20, and we see that the maximum probability estimates for the facet positions again indicate a smooth motion field of varying magnitude and direction. This also holds true across time, as we see that the overall deformation magnitude is seen to be in between the results from the two previously shown gates.

Next is shown a subarea of the estimated facet positions for this intermediary gate in detail in Figure 4.21, this time with display of all lowest-level facets corresponding to three orthogonal slices through the same intersection point as before. These are the estimated positions that will later be used to calculate the composite image for the timeseries. Again, reasonable behavior is seen.

Finally, some individual lowest level facet placements are displayed in Figure 4.22 to demonstrate that they correspond to reason and visual inspection. We see that there is cohesiveness in the facet patterns shown, while deformations vary from minor to significant. Especially note the significant slice shifting as should be expected for this fully three-dimensional estimation.

4.2.2 Quantitative Results

Since we posess the underlying truth about both heart shape generation (if not individual tissue position) and contraction pattern, the sections on summary images and contraction parameters, while still somewhat qualitative in nature, offer some

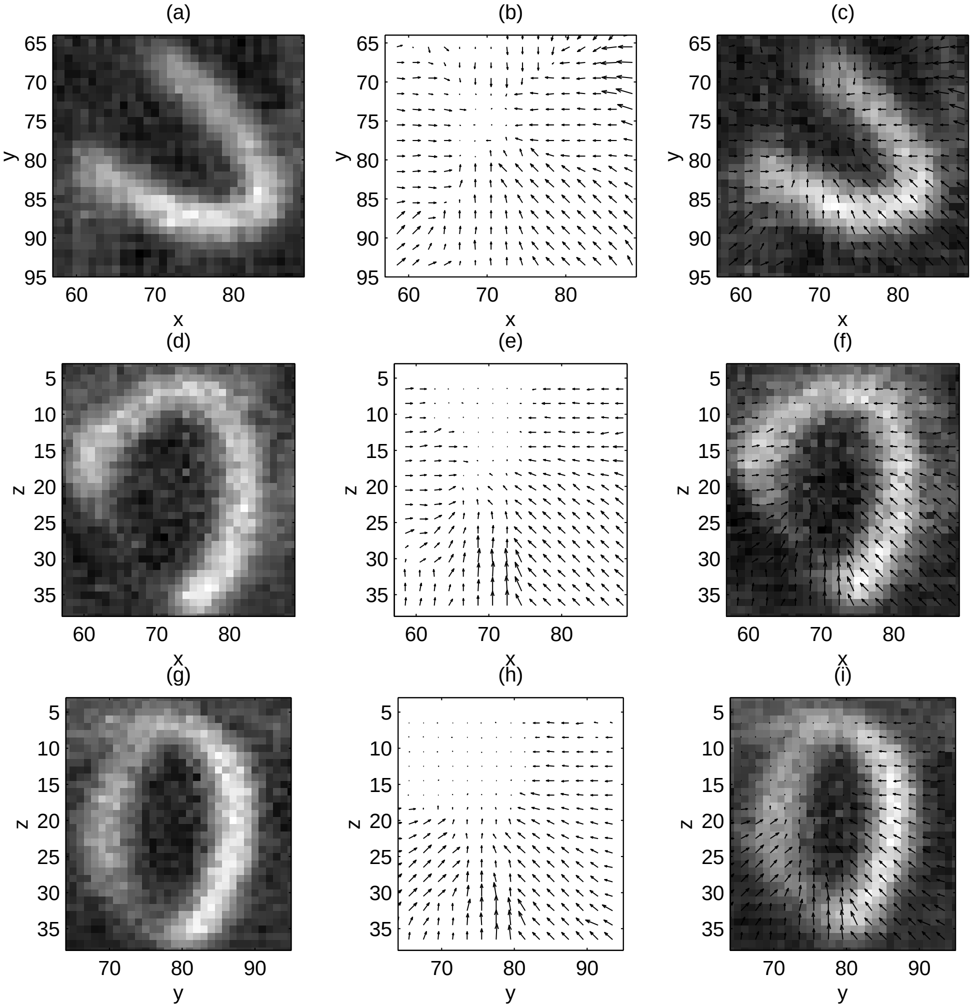

Figure 4.20: Facet motion from gate 1 (reference;expanded) to gate 25 (late-expansion) in perpendicular slices with intersection point at the center of the LV chamber. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 25 image. We see that the estimates again indicate a smoothly varying contraction and that the overall magnitude is intermediate relative to that seen in Figures 4.18 and 4.19.

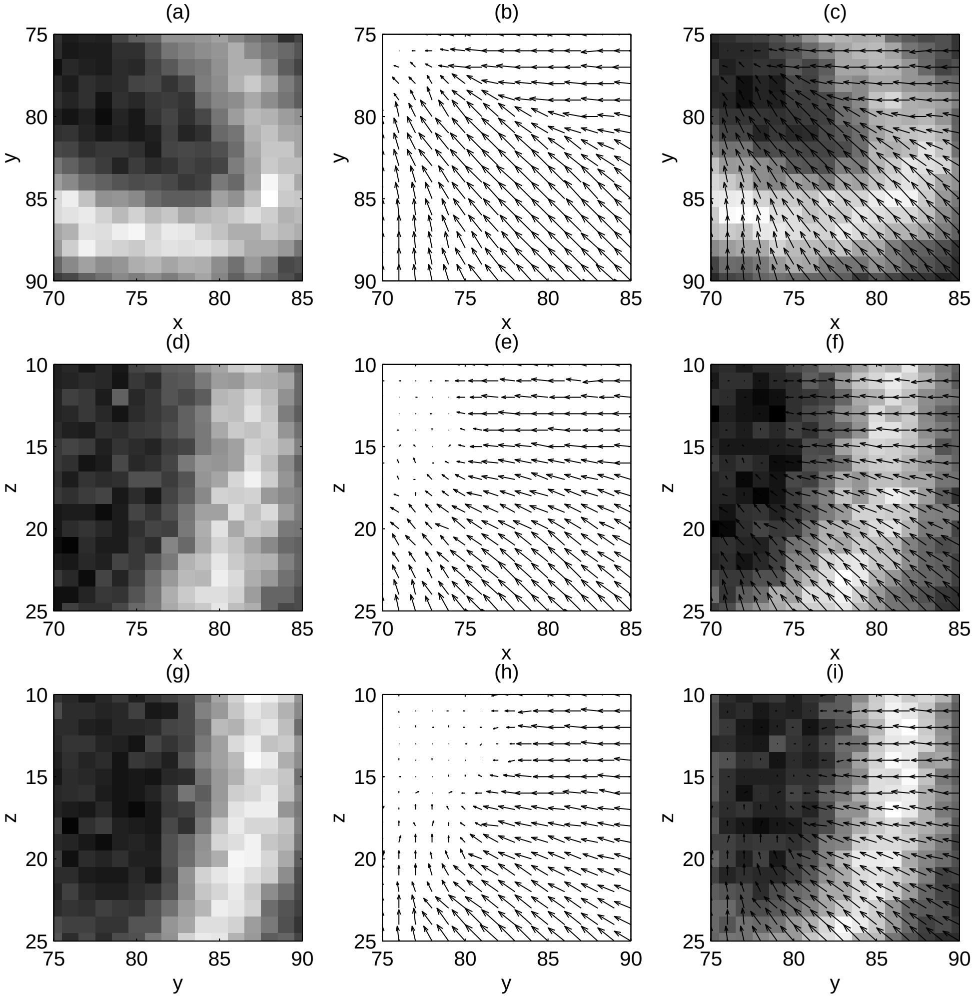

Figure 4.21: Detailed facet motion from gate 1 (reference;expanded) to gate 25 (mid-expansion) in perpendicular slices with intersection point at the center of the LV chamber as in Figure 4.20. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the gate 25 image.

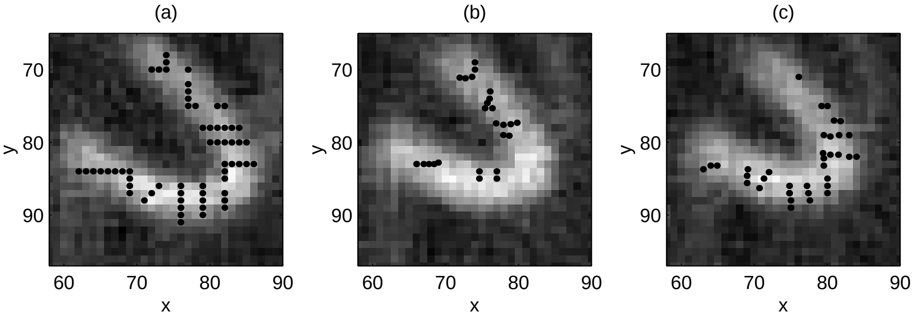

Figure 4.22: Selected facet positions in transaxial slice 22 of the reference image (a), with corresponding estimated facet positions in the gate 19 image, slice 20 (b) and 21 (c). Note the smooth contraction and deformation in relative positioning without losing relations between neighboring facets, as well as the slice-jump, indicating fully 3D deformation.

quantitative comparisons as well.

Summary Images

Since this is an image series with better noise characteristics than the earlier discussed patient data, we would expect that we can even more markedly improve the summary image explained in Section 3.5.1. This is indeed the case, as evidenced in Figures 4.23 and 4.24. In the composite image formed using the facet placement estimates we see a better delineation of the heart wall as mapped to the reference, as well as similar uniformity of intensity within the heart wall. We know that the true activity distribution was constant, and it looks like we consistently overestimate the activity in some regions based on areas that are imaged as particularly high in intensity in the reference gate we are using. The difference images clearly show, however, how much better this methodology maps the activity into a clinically meaningful shape when compared to the physiologically non-existent average state.

Contraction Parameters

The contraction parameters again behave exactly as expected, tracing the visually evident and known pattern of the contraction and expansion cycle of the heart beat.

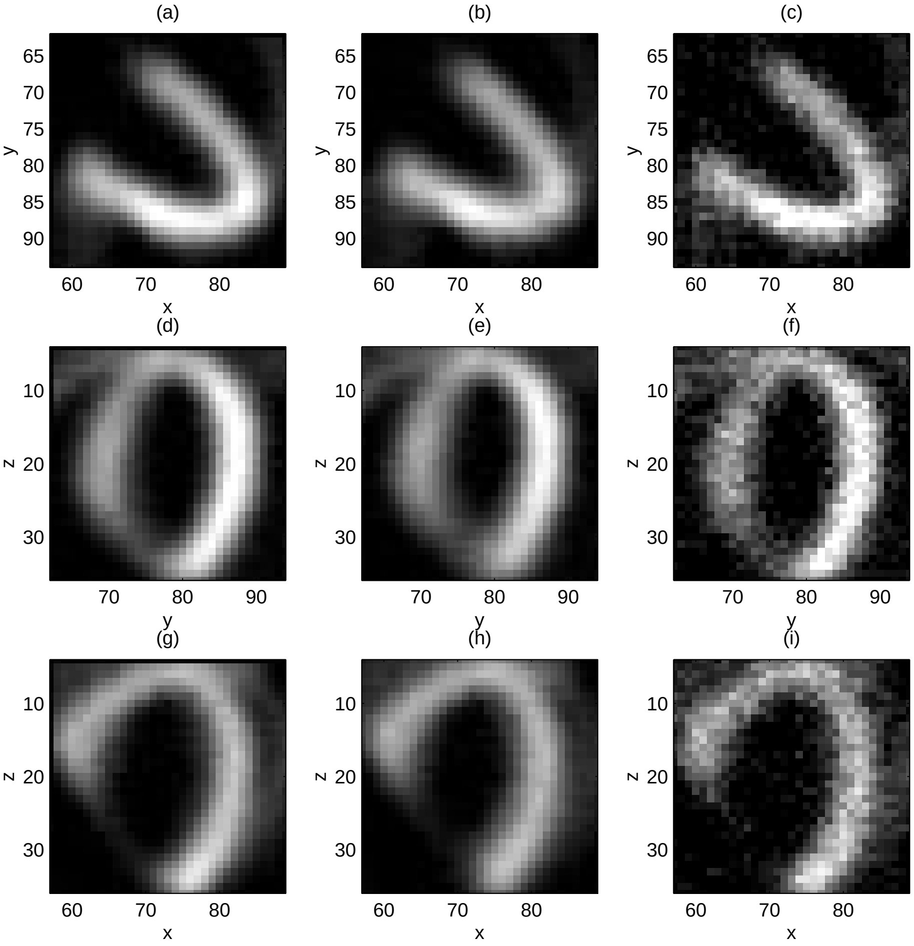

Figure 4.23: Perpendicular slices from three representations of the data. The intersection point is at the center of the LV chamber, $(22, 77, 72)$. (a)-(c) show the transaxial view, (d)-(f) show the coronal view and (g)-(i) show the sagittal view. The representations shown are : (a),(d),(g) Composite image mapped to gate 1, (b),(e),(h) mean voxel-wise image (standard SPECT), and (c),(f),(h) gate 1 image only. The composite image was formed by computing average intensity based on the facet motion through the series of images and mapping back to the reference image (gate 1). We observe similarity between gate 1 and composite, with the composite having superior smoothness in regions of activity. Furthermore, a better spatial delineation of the heart wall in the composite image relative to the standard SPECT image is seen.

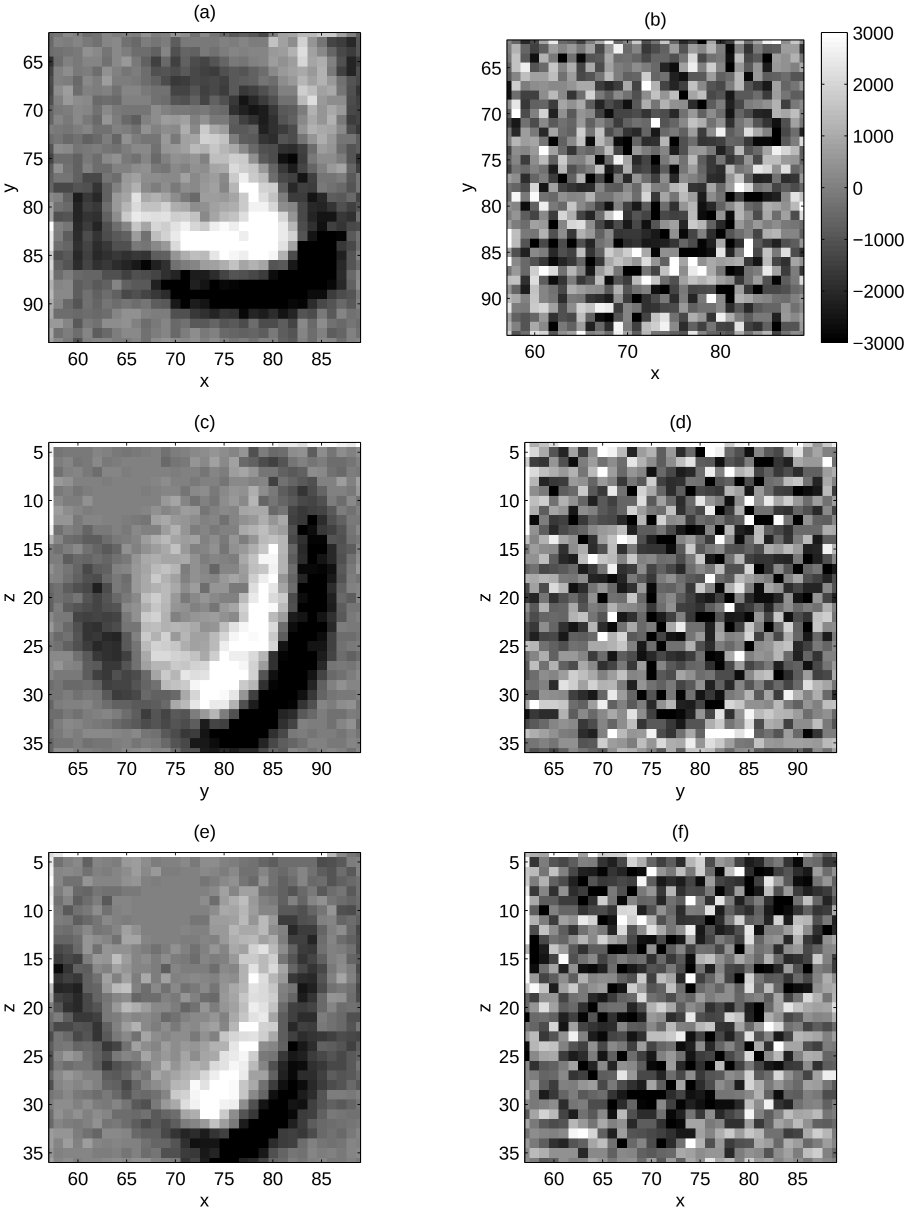

Figure 4.24: Difference images corresponding to the images in Figure 4.23; (a),(c) and (e) show mean minus composite (mapped to gate 1), and (b),(d) and (f) show gate 1 minus composite. For detailed explanation of the composite image, see Section 3.5.1. Here we see clearly the structural difference between the standard (mean) image and the composite facet-based image. The regions of dark and bright indicate that the deformation model has shifted intensity outward from the LV chamber center, especially for the lateral wall of the left ventricle. This is in accordance with the use of gate 1 (expanded) state as the reference and agrees with the moving versus stationary regions both as observed in the dataset and as known from the underlying model.

	Average	Range
Normalized residual	0.023	[0.001 0.044]
Normalized residual (gates 11-32)	.008	[0.001 0.015]

Table 4.1: Table of residuals for the regression of outer LV chamber volume on the estimated product of contraction parameters. Both the results with and without the initial phase of the heart cycle is shown. The poorer performance on the full series is probably due to a voxelization effect discussed in the text.

We also have access to the hyper-ellipsoid parameters used for the generation of the LV chamber, and so an intelligent comparison can be made.

In Figure 4.25 are shown the three contraction parameters through the beat cycle, next to a normalized version of the three ellipsoid radii used for the mathematical phantom data generation. We see good if somewhat noisy correspondence between the three sets, recalling that none of these can be said to correspond to the other, since the orientation of the heart is not the same as that used in the modeling. When viewed as a volume, which is the most meaningful clinical quantity in any case, that is, taking the product of the three parameters, we see an excellent correspondence between the three curves as seen in Figure 4.26. This indicates that the contraction parameters could be used as a good predictor for LV chamber size for this type of data. In fact, when used to predict the outer volume of the LV chamber through the rapid contraction-expansion phase of the heart beat, we get a normalized residual error of less than .025 when regressing the product of the phantom radii on the product of the γ_2. The results of this simplistic prediction is summarized in table 4.1 with the results plotted in Figure 4.27.

The only clear inconsistency in estimation is in the slight expansion phase seen in the beginning of the cycle. This is easily accounted for, however, when we consider that the change in radius must translate into actually changed voxel labels for a significant number around the heart in order to be detectable. Since the radius of the LV chamber in the reference varies from 5 to 15 voxels, this means that a size

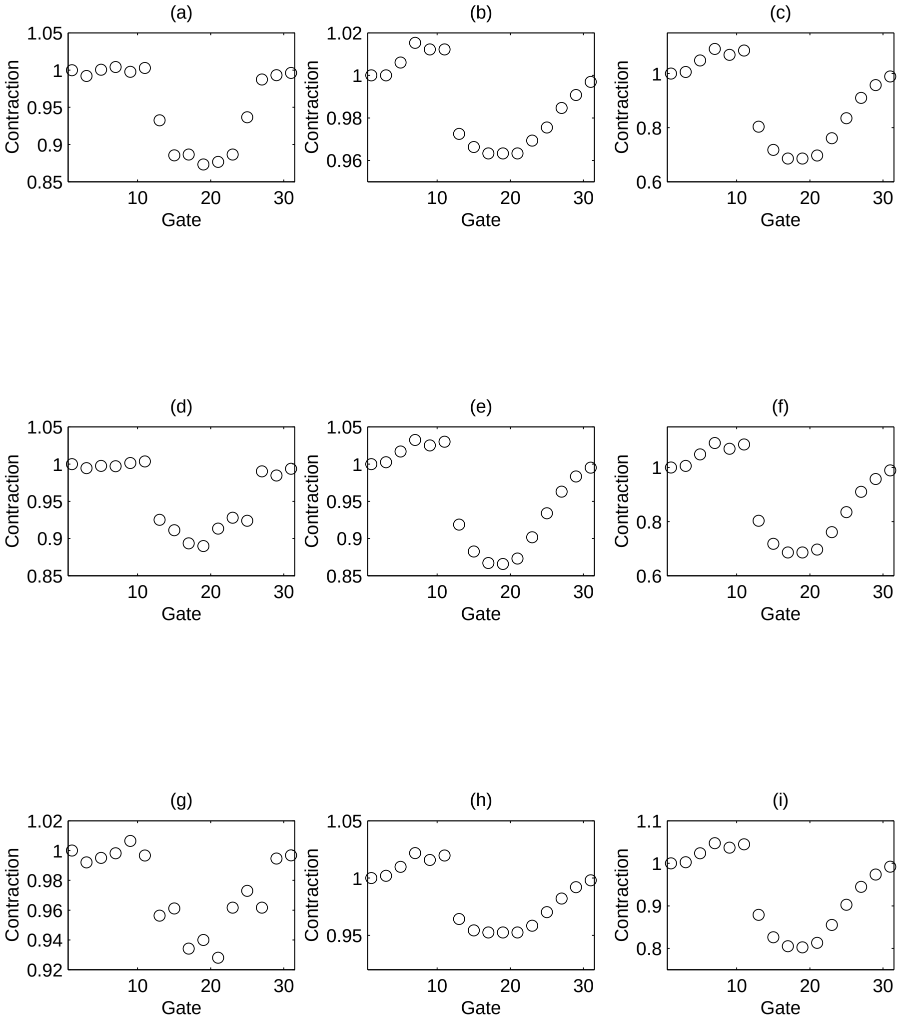

Figure 4.25: Estimated contraction parameters through cycle (a),(d),(g), contrasted to mathematical phantom parameters for the left ventricle; outer wall (b),(e),(h) and inner wall (c),(f),(i).

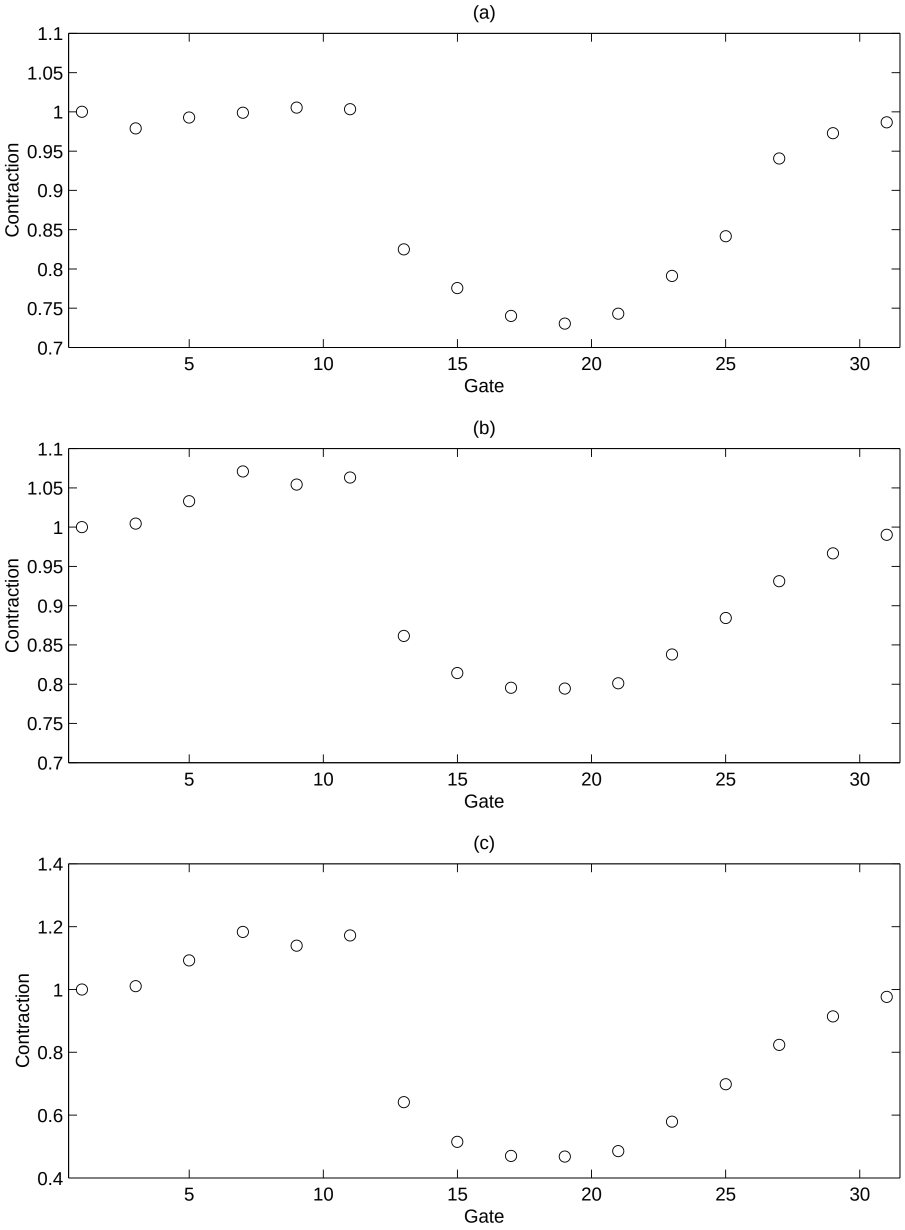

Figure 4.26: Estimated overall contraction through cycle, measured as the product of the γ_2, panel (a). This is contrasted to the product of the outer LV chamber wall (panel (b)) and the inner LV chamber wall (panel (c)).

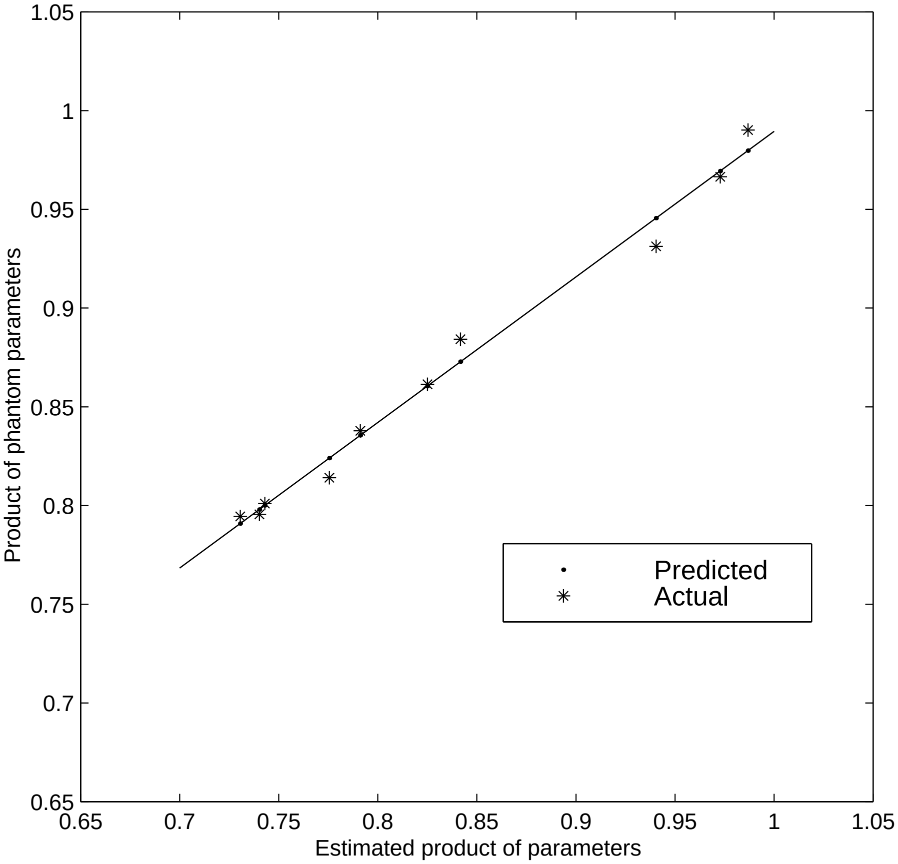

Figure 4.27: Plot of the results from using a linear regression on the product contraction parameters to predict the product of the LV chamber radii during the contraction and expansion phase (gates 11 through 31).

parameter must change by a minumum of 5 per cent uniformly for this to happen. Thus the model will not be able to pick up small changes such as that seen in the beginning of the cycle simply because they are not observable when translated into the discrete overall voxelized representation. There is also a range of values for γ_{2i} for which a change is not meaningful in real terms, since no voxel change in actual facet placement is effected. Thus, the fluctuations in a “band-like” fashion seen in Figure 4.25 are reasonable. Note that the deformation is still estimated (as seen in Figure 4.19) reasonably in the model given γ_2.

To illustrate convergence of the maximization for these parameters, the maximization was allowed to run for 100 iterations, and the beginning of the resulting timeseries for one component of γ_2 for five of the gates is shown in Figure 4.28. We see that the parameter converges for all gates within approximately 20 iterations, while those gates started closer to the estimated maximum density value converge in less than 5. This indicates that finding a heuristic starting point, maybe by automatic distance measurements on a thresholded version of the image, could speed up convergence significantly. In addition, the log probabilities of the facet estimated configuration is shown to demonstrate the maximization convergence. The slight fluctuations once the curves flatten out are due to the approximations used for estimating the upper level facet positions.

4.3 Clinical Phantom Data

The final dataset considered was formed using a hollow beating heart phantom to image three states of the heart; maximum expansion, maximum contraction and a point roughly midways between the two. These scans were performed at Duke University Medical Center with standard clinical parameters and resulted in three images consisting of 64x64x61 voxels each, covering the entire heart phantom. Note

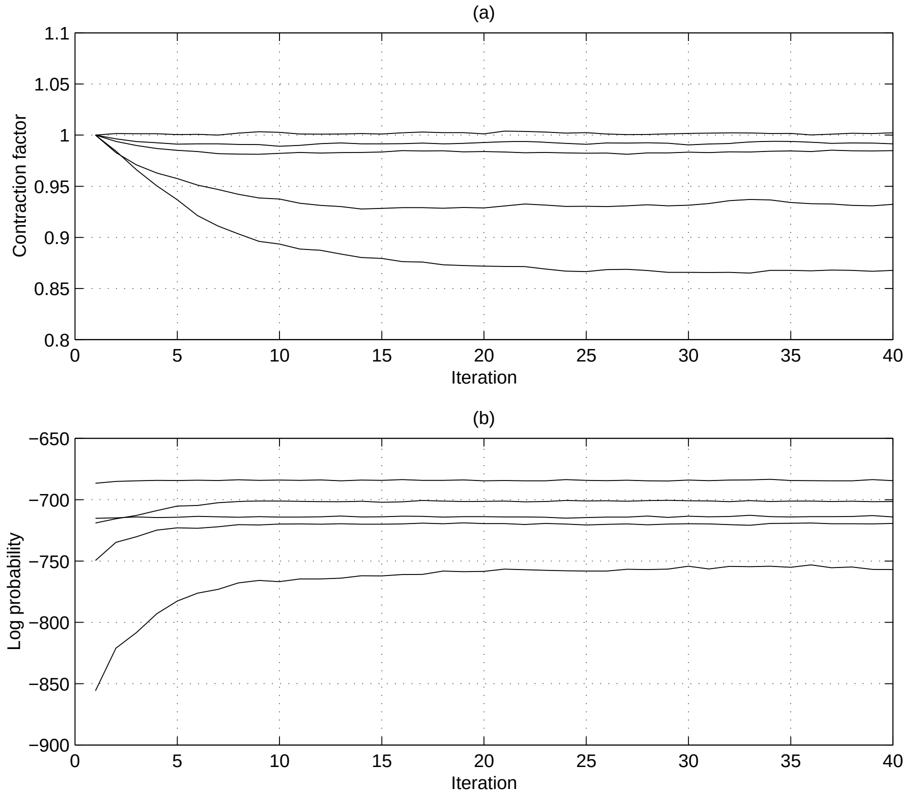

Figure 4.28: Here is shown in (a) the values for one of the contraction parameters for five of the gates in the mathematical phantom data timeseries through several maximization cycles. We see convergence on the order of 5 iterations for gates close to the reference and on the order of 20 for the gate furthest from the reference. In (b) is seen the associated log probability densities for the full parameter configuration for the same chains. The fluctuations in (b) are caused by the approximation used on the higher levels of facet.

that this dataset, unlike the previous two, has activity throughout the heart and is as such a phantom only in the volumetric and not in the functional (activity localization) sense. No attempt was made to partition out the LV chamber, as we used the set exclusively to evaluate overall motion and the tracking of markers placed on the surface of the phantom. Points on the phantom surface were labeled by imaging the phantom twice at the exact same settings, with highly radioactive sources fixed to the surface of the heart phantom in one of the scan series. Figure 4.29 shows identical slices from the two series at each heart state imaged. We can thus extract a relatively accurate motion map for a number of points on the surface and compare to the estimated facet placement under the model. In all 9 markers were placed, and of these 8 were in positions which could usefully evaluate the model estimation accuracy. The model was applied by using a 5 level facet hierarchy (4,681 facets total with a 16^3 cube on the bottom level) to cover a 20x20x25 voxel volume which contained the heart portion of the phantom. The fully expanded state was used as the reference dataset.

4.3.1 Qualitative Results

Facet Motion Estimates

As with the other datasets, the first thing considered is the overall facet motion derived from the facet maximum probability position estimates. Again, very reasonable behavior is seen. In Figure 4.30 is shown the motion field for the maximum contracted state, and once again we observe overall contraction modified non-linearly but smoothly to fit the observed image features. A small region with no movement is seen, while the rest of the heart generally contracts smoothly towards the upper portion of the heart. In Figure 4.31 is seen the same type of figure for the intermediary state imaged. This time, we see a mild contraction with significant areas of little

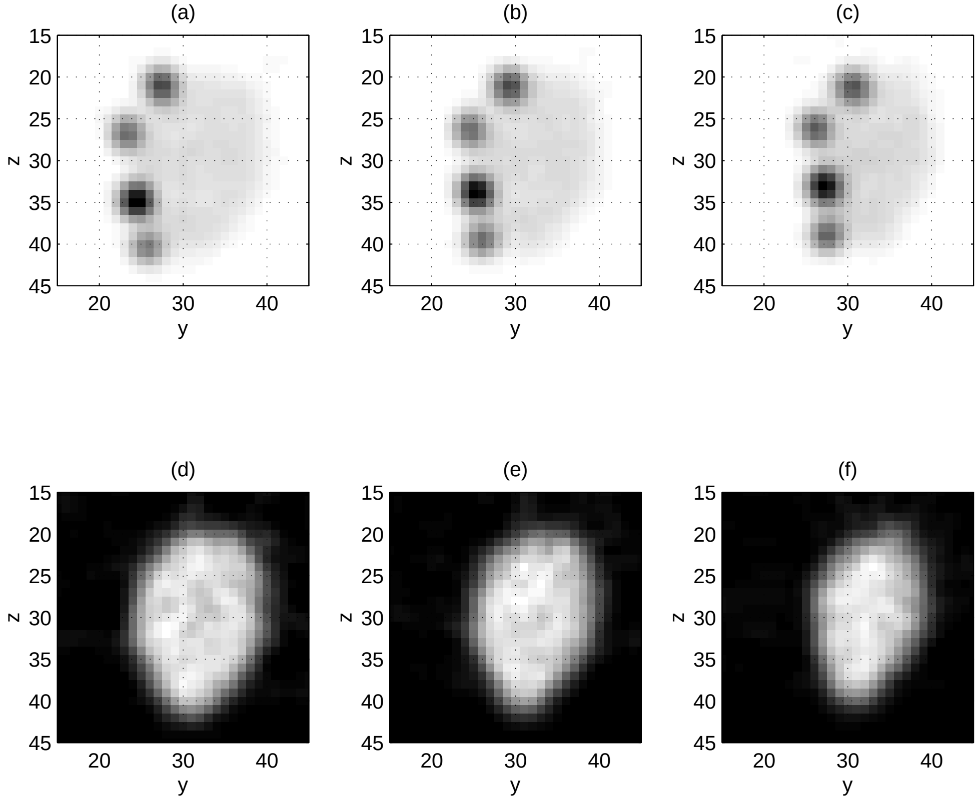

Figure 4.29: The clinical phantom data imaged with and without markers. Panels (a) through (c) show the same central sagittal plane for the three different states of the phantom that were imaged. Panels (d) through (f) show the same for the case without markers attached, which is the data used for the facet estimation. Reverse intensity was used in the marker case as it provided for better visual illustration of their location.

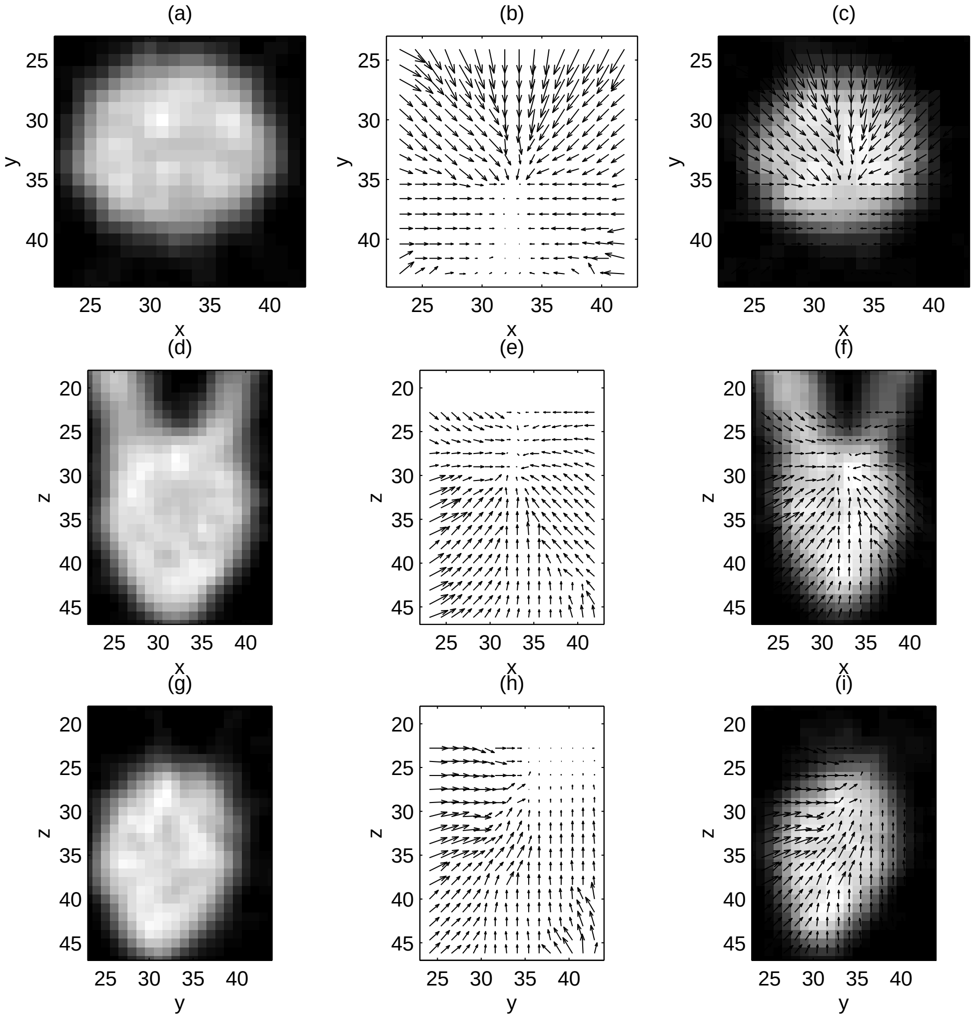

Figure 4.30: Facet motion from expanded to contracted state in perpendicular slices with intersection point at the center of the heart. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the reference image.

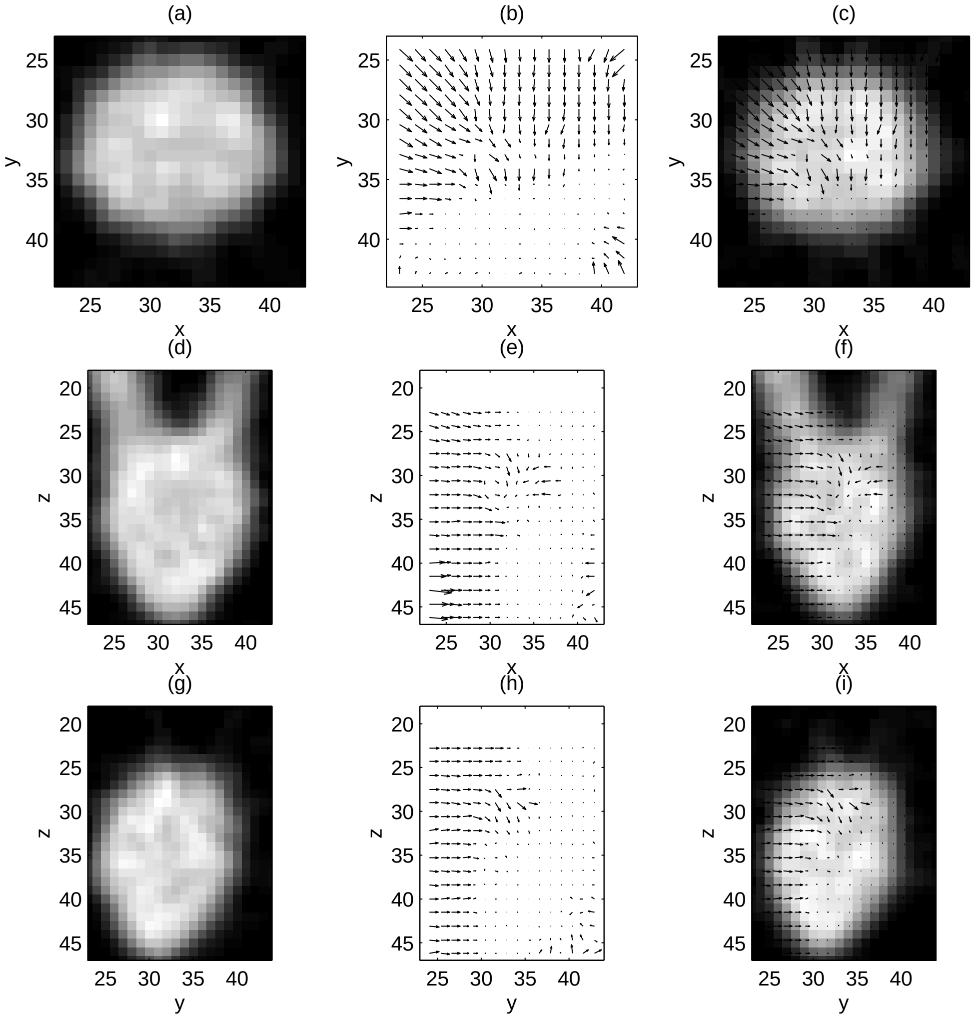

Figure 4.31: Facet motion from expanded to intermediary state in perpendicular slices with intersection point at the center of the heart. Panels (a)–(c) show a transaxial plane, (d)–(f) a coronal plane and (g)–(i) a sagittal plane. The left column shows the reference image slice, the middle column shows the facet motion estimate and the right column shows that same estimate superimposed on the image slice from the reference image.

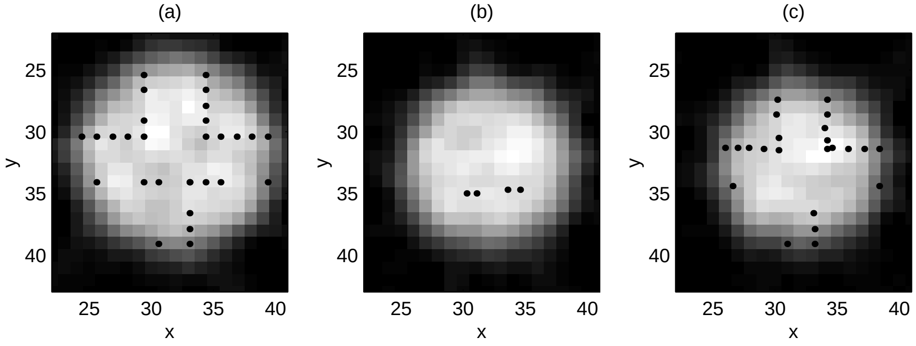

Figure 4.32: Selected facet positions in transaxial slice 33 of the reference image (a), with corresponding estimated facet positions in the maximum contracted image, slice 31 (b) and 32 (c).

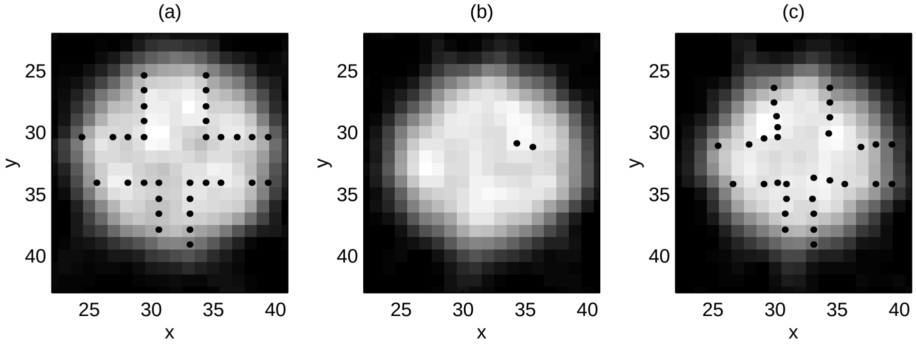

Figure 4.33: Selected facet positions in transaxial slice 33 of the reference image (a), with corresponding estimated facet positions in the intermediary contracted image, slice 32 (b) and 33 (c).

or no movement. The pattern seen is consistent both with the more expanded state in the sense of sequential motion, and also with the observed shape of the heart in this intermediary state.

Figures 4.32 and 4.33 show the same set of facets in the reference image with their estimated observed locations in the maximum contracted and intermediary state, respectively. The estimation is entirely reasonable in terms of relative facet positioning within an image (smoothly deformed). In addition, the progression from the intermediary to maximum contraction illustrates the deformation along all three dimensions as facets move from mostly being in slice 33 to mostly being in slice 32.

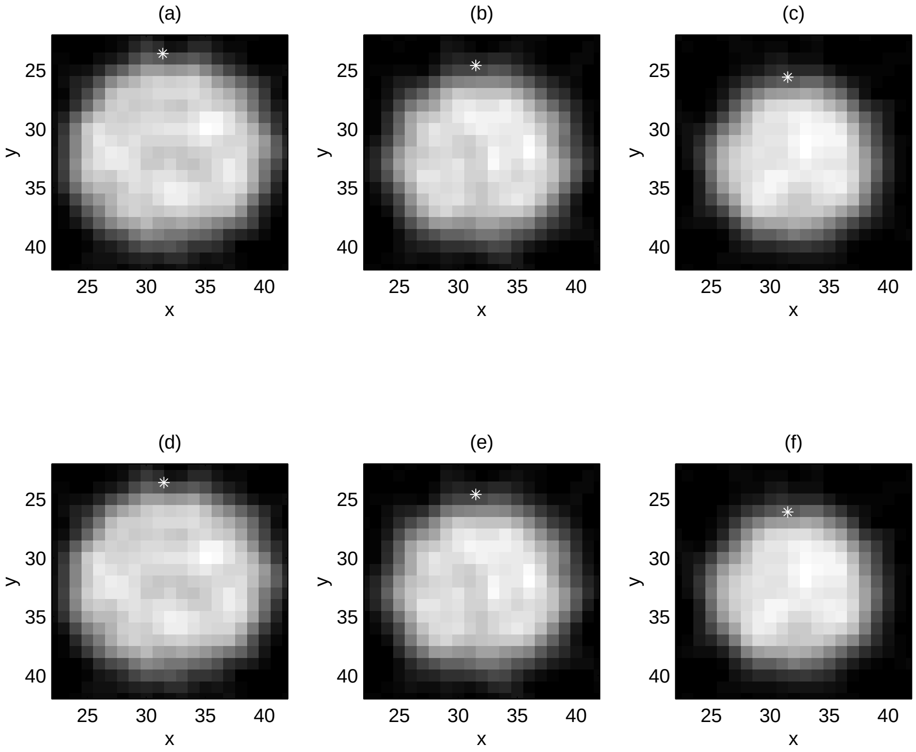

Figure 4.34: One of the markers' (situated in slice 28 in the reference image) actual and estimated position in the clinical phantom dataset. Facet position in the (a) reference (expanded) image, (b) intermediary state and (c) maximum contracted state are shown. Also, measured marker position in the (d) expanded image, (e) intermediary state and (f) maximum contracted state are indicated. A near-perfect match is seen.

4.3.2 Quantitative Results

Marker movement

Finally, the eight markers were evaluated. The facet closest to each marker in the reference image was located, and the motion estimated under the model was compared to the motion observed manually from the marker dataset. This resulted in excellent matches.

Most facets were estimated within 1 voxels of the measured position, and they all fell within a 1.5 voxel distance (see table 4.2). They also all agree well with visual inspection, as demonstrated in Figures 4.34 and 4.35. The error seen is completely

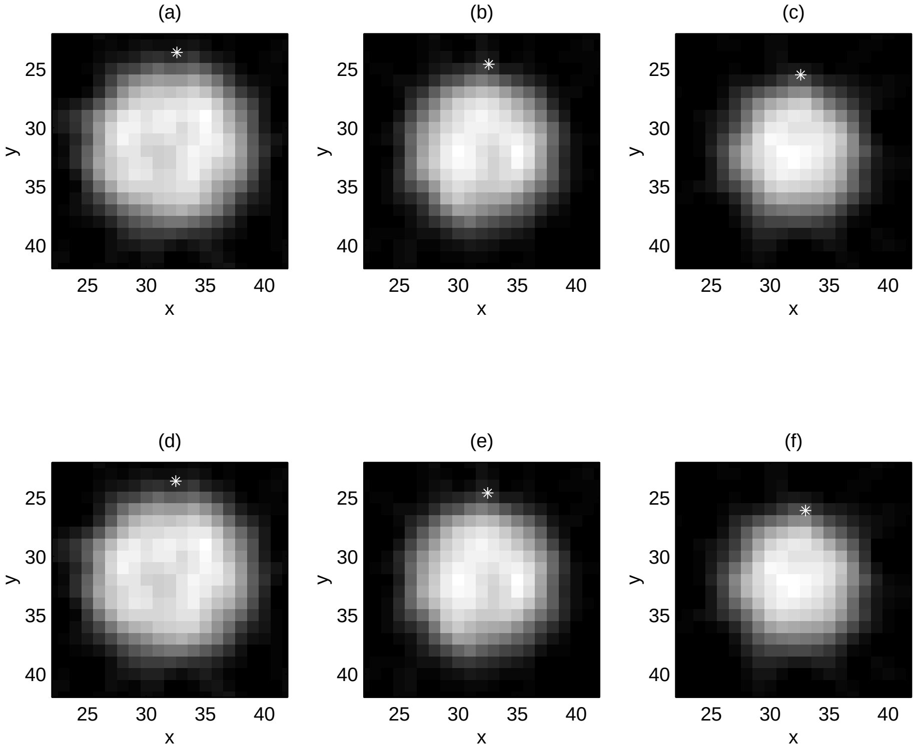

Figure 4.35: One of the markers' (situated in slice 36 in the reference image) actual and estimated position in the clinical phantom dataset. Facet position in the (a) reference (expanded) image, (b) intermediary state and (c) maximum contracted state are shown. Also, measured marker position in the (d) expanded image, (e) intermediary state and (f) maximum contracted state are indicated. A near-perfect match is again seen.

Image data set	Average Value	Range
Intermediary State	0.55	[0.10 0.74]
Fully Contracted State	0.94	[0.58 1.35]

Table 4.2: Average values and range for deviation between facet placement and observed marker location for the clinical phantom dataset. We see excellent correspondence, within the limits of error inherent in the measurement method used to determine marker locations manually and the natural scale of the image series, 1 voxel.

acceptable given that there is uncertainty associated with the location of the marker (performed by visual inspection of images such as seen in Figure 4.29 (a)-(c)). There may also not be an exact match between the two image series. While the two series agree extremely well, there may be local inaccuracies at scales as small as one voxel. Overall, the agreement seen between actual perceived motion and estimated facet positions lends great strength to the method's ability to follow individual structural points through the timeseries.

Chapter 5

DISCUSSION AND EXTENSIONS

5.1 General Discussion

The statistical model proposed in this thesis uses an image understanding approach to improve the utility and applicability of gated cardiac SPECT data. By balancing the modeling of known characteristics of heart motion against computational concerns, an estimation of perceived facet locations has proven possible on a timescale which could render the methods useful in clinical settings. The inclusion of contraction parameters encourages deformation estimates to track known cardiac behavior while remaining relatively simple in implementation. The contraction parameters themselves are meaningful in tracking the heartbeat cycle and for rough estimation of volumes. Furthermore, the structure of the facet model, that is, the use of a large number of labelled points, leads to the set of locations for every facet in every gate as a very rich representation of the image data. With further investigation of true anatomical correspondence, this representation can offer new diagnostic ways to look at heart function abnormalities via estimated deformations rather than based solely on radiotracer uptake. The facet-composite image is also a clearly improved summary of the image time series over the voxel-wise sum and offers better intensity uniformity in the heart region and from that a better SNR than the individual gated images. Furthermore, reasonable volume estimates for the patient data and corresponding ejection fraction estimates show that the methods can track relevant clinical quantities well. The successful application of the model to phantom data demonstrates that the methods provide accurate as well as qualitatively sensible results. The close tracking of known parameters as seen in the mathematical phantom case and the accurate estimation of specific point motion as seen for the clinical phantom especially support the accuracy and utility of the proposed deformation method. The maximimization technique used throughout provides good estimates, and while it does not guarantee convergence to a global maximum, it clearly provides very reasonable facet placements. This, balanced with the computational savings, justifies its use.

5.2 Modality and Data Extensions

This methodology can obviously be directly applied to other medical imaging modalities, in particular those who image the beating heart in a sequential manner. Gated magnetic resonance imaging or positron emission tomography are other techniques used to image cardiac function in this manner. The general methodology should be useful for any kind of modeling of objects where some knowledge about size evolution is available a priori or also if linear scaling in expectation of facet placement for an image scene is desireable.

The Monte Carlo computed phantom of a beating heart in a thorax as used in the evaluation can, when further refined and validated, provide a rich testing ground for this type of modeling by allowing further testing of individual facet motion and the shape deformation of regions in hearts of different shapes and sizes, and eventually, when such data becomes available, in various disease states. This is important since there is no clinical data available to evaluate the real motion of individual heart tissue elements for time series such as these, as that would require invasiveness beyond what is acceptable as clinical risk. Once the simulation data has been validated as being near-clinical, simulation data could also be used for parameter estimation based on segmentation data, inherently available from the phantom data, or individual facet positions, if these could be sensibly mathematically extracted. Currently, estimation of parameters such as κ and τ is limited by lack of data rather than lack of methodology. The approach for both types of estimation mentioned above is already outlined in McCulloch (1998).

As for cardiac SPECT itself, research is being performed in the area of including the simultaneous acquisition of a second dataset, so-called transmission computed tomography (TCT). In this modality, radiation is emitted from a radiation source mounted opposite the detector, and the transmission of the radiation that passes through the patient is measured to produce an anatomical rather than functional map of the heart. This data could then be used in a straightforward manner as additional image feature data (included in $\{f, \phi\}$) for the deformation modeling.

The application of the general model without contraction to modalities such as positron emission tomography and magnetic resonance imaging (both anatomical and functional) is already underways. One of the most challenging problems arising out of some of this data is due to these objects containing more detail than in the cardiac SPECT case. In addition, the objects being deformed (brain sulci, f.ex.) vary in more complicated fashion. Thus, the balance between gross and local deformation needs to be refined, that is, the model must accommodate the fact that there may be regions greatly deformed next to regions which are very similar to the template, something the model currently handles poorly if at all.

5.3 Other Model Extensions

In terms of the model itself, then, the most important future extension seems to be the ability to adjust linkages between facets via the corresponding weights in the design matrix A_l. This is based on the idea that structures can be the same conceptually and yet be different enough to not be captureable under the model as described. In such scenarios, the desire is for the model to be automatically adjustable. This could be achieved if a stable estimation scheme for the individual weights used to form A themselves was developed, for instance by using a Dirichlet prior $p(w)$ on the set of weights connected upwards from a facet. The model would then be specified as

$$p(x, f, w, \gamma|\theta) = p(x|\theta_x, \gamma)p(f|\theta_f)p(\gamma)p(w) \ , \tag{5.1}$$

and maximization and sampling, while more complicated than in the current model, could be performed as before based on the full conditionals,

$$p(w|x, f, \gamma, \theta), p(\gamma|x, f, w, \theta) \ , \tag{5.2}$$

and so on.

An important area which is already being pursued, is the possibility of inclusion of image features at all levels. This corresponds to extending the p_I to include terms linked to facets lj for $l \neq L$. This obviously leads to a more computationally demanding model, but is sensible in the sense of basing the deformation model itself on the scale-space of an image, rather than using the scalespace as a maximization tool as done here. The question of choice of feature also offers itself as you move up in the hierarchy. It is clear that the same features are not equally meaningful at all levels. For instance, the local correlation feature may make sense on the lowest few levels, but higher up, the scale-space intensity or scale-space Laplacian may offer a better modeling candidate for the image deformation. Finally, it is worth noting that any representation of the image can be seen as a candidate for image features, for example the wavelet representation or other image transforms. We then get away from modeling human visual perception in the direct sense presented here, but nothing in the implementation precludes such application.

Another issue is the choice of the reference values ϕ. When choosing a single gate image as the reference as done for the case discussed in this thesis, you leave yourself open to artifacts from spurious details in that image. For example, the slight over-estimation of activity in the mathematical phantom data summary images can be traced to this effect. It may be useful, therefore, either to compile an atlas of expected behavior for ϕ across the reference using more than one example. Another possibility is to add estimation of ϕ directly into the modeling, thus drawing strength from the combined observation of facet positions across the timeseries. As this would make most sense in a framework where the entire series is modeled jointly, the computational burden would be quite significant, but the case can certainly be made for at least evaluating the increase in performance such estimation would entail. In a way, this would also bring the modeling full circle, as we would then use our image understanding method to estimate not just observed image deformation, but now indirectly get back to the process underlying the image formation via a perception aproach rather than a physically based model.

Another possible extension of the model in terms of clinical applicability for SPECT, is to have several reference sequences to which the imaged series is mapped. This could for example be achieved by first estimating the contraction factors γ_2 under a reduced model (low number L of levels). Based on those estimates, one could then choose one or several of the reference series to estimate a more detailed deformation. Probabilities of having a certain heart characteristic, disease or otherwise, could then be expressed from the shape deviations from these states. The alternative would be to use tissue motion estimates from single series as direct indicators for heart malfunction. Any such approach to prove clinical significance of such derived information from the facet position estimation would require access to an extensive clinical database and close collaboration with clinical researchers and medical personnel.

Bibliography

Amit, Y and Kong, A (1994) Graphical templates for model registration. Technical Report. Department of Statistics, University of Chicago.

Amit, Yali, Grenander, Ulf and Piccione, M (1991) Structural image restoration through deformable templates. *Journal of the American Statistical Society*, **86,** 376–387.

Baydush, Alan H, Bowsher, James E, Laading, Jacob K and Floyd, Carey E Jr. (1997) A new bayesian scatter reduction algorithm for digital chest radiography. *Medical Physics*, **24,** 539–545.

Besag, Julian (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, series* B, **36,** no. 2, 192–236.

Besag, Julian (1986) On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, series* B, **48,** 259–302.

Besag, Julian, Green, Peter, Higdon, David and Mengersen, Kerrie (1995) Bayesian computation and inference. *Statistical Science*, **10,** 3–52.

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

Bookstein, Fred (1986) Size and shape spaces for landmark data in two dimensions. *Statistical Science*, **1,** 181–242.

Clarysse, Patrick, Friboulet, Denis and Magnin, Isabelle E. (1997) Tracking geometrical descriptors on 3-d deformable surfaces: Application to the left-ventricular surface of the heart. *IEEE Transactions on Medical Imaging*, **16,** no. 4, 392–404.

Collins, D., Holmes, C., Peters, T. and Evans, A. (1995) Automatic 3-d model-based neuroanatomical segmentation. *Human Brain Mapping*, **3,** 190–208.

Cootes, T F, Hill, A, Taylor, C J and Haslam, J (1994) The use of active shape models for locating structures in medical images. *Image and vision computing*, **12,** no. 6, 355–366.

Cressie, N.A.C. (1993) *Statistics for Spatial Data.* Wiley-Interscience.

Dobrushin, R L (1968) The description of a random field by means of conditional probabilities and conditions of its regularity. *Theory of Probability and its Applications*, **13,** 197–224.

Dryden, Ian and Mardia, Kanti (1996) General shape and registration analysis. Technical Report. University of Leeds.

Gelman, A, Carlin, J B, Stern, H S and Rubin, D B (1995) *Bayesian Data Analysis.* Chapman and Hall.

Geman, Stuart and Geman, Donald (1984) Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6,** 721–741.

ter Haar Romeny, Bart M, Florack, Luc, Koenderink, J and Viergever, M (1991) Scale-space: its natural operators and differential invariants. In *Lecture Notes in Computer Science*, *Berlin, Germany*, pp. 239–255. Springer-Verlag.

Hastings, W K (1970) Monte carlo sampling methods using markov chains and their application. *Biometrika*, **57,** 97–109.

Higdon, D, Bowsher, J E, Johnson, V E, Turkington, T G, Gilland, D R and Jaszczak, R J (1996) Fully bayesian estimation of gibbs hyperparameters for emission computed tomography data. Technical Report. ISDS, Duke University.

Johnson, V, Wong, W, Hu, X and Chen, C (1991) Boundary modeling, treatment of blurring and selection of hyperparameters. *IEEE Trans. PAMI*, **13,** 413–425.

Johnson, Valen (1994) A model for segmentation and analysis of images. *Journal of the American Statistical Society*, **89,** no. 425, 230–241.

Kass, M, Witkin, A and Terzopolous, D (1988) Snakes: active contour models. *International journal of computer vision*, 321–331.

Kirkpatrick, S, Gellatt, C D and Vecchi, M P (1982) Optimization by simulated annealing. Technical Report. IBM Thomas J. Watson Research Center, Yorktown Heights, NY.

Klein, G J, Reutter, B W and Huesman, R H (1997) Non-rigid summing of gated pet via optical flow. *IEEE Transactions on Nuclear Science*, **44,** no. 4, 1509–1512.

Laading, Jacob K, Floyd, Carey E, Baydush, Alan H and Bowsher, James E (1996) Statistical modeling of the chest radiograph and simulation in a bayesian framework. In *Proc. of the IEEE Engineering in Medicine and Biology Society Conference 1996, Amsterdam, the Netherlands.*

Laading, Jacob K, McColluch, C, Johnson, Valen E, Gilland, David R and Jaszczak, Ronald J (1999) A hierarchical feature based deformation model applied to 4d cardiac spect data. In *Lecture Notes in Computer Science : Information Processing in Medical Imaging, Berlin, Germany.*

Laading, Jacob K, McCulloch, Colin C and Johnson, Valen E (1997) A hierarchical object deformation model applied to the digital chest radiograph. In *The American Statistical Association Proceedings of the Section on Bayesian Statistical Science, Anaheim, California.*

Lavine, Michael (1998) Another look at conditionally gaussian markov random fields. Technical Report 98-04. ISDS, Duke University.

Lifshitz, L. and Pizer, S. (1990) A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Trans. PAMI*, **12,** no. 6, 529–640.

Lindeberg, Tony (1994) *Scale-space Theory.* Boston, MA: Kluwer Academic Publishers.

McCulloch, Colin C (1998) High-level image understanding via bayesian hierarchical models. Ph.D. Thesis. Duke University.

McCulloch, Colin C, Laading, Jacob K and Johnson, Valen E (1997) Image feature identification via bayesian hierarchical models. In *The American Statistical Association Proceedings of the Section on Bayesian Statistical Science, Anaheim, California.*

McCulloch, Colin C, Laading, Jacob K, Wilson, Alyson and Johnson, Valen E (1996) A shape-based framework for automated image segmentation. In *The American Statistical Association Proceedings of the Section on Bayesian Statistical Science, Chicago, Illinois*, pp. 1–6.

McEachen, John C and Duncan, James S (1997) Shape-based tracking of left ventricular motion. *IEEE Transactions on Medical Imaging*, **16,** no. 3, 270–283.

Metropolis, N, Rosenbluth, A W, Rosenbluth, M N, Teller, A H and Teller, E (1953) Equations of state calculations by fast computing methods. *Journal of Chemical Physics*, **21,** 1087–1091.

Nelder, J. and Mead, R. (1965) A simplex method for function minimization. *The Computer Journal*, **7,** 308–313.

Park, Jinah, Metaxas, Dimitri, Young, Alistair A and Axel, Leon (1996) Deformable models with parameter functions for cardiac motion analysis from tagged mri data. *IEEE Transactions on Medical Imaging*, **15,** no. 3, 1–13.

Peter, J, Gilland, D R, Jaszczak, R J and Coleman, R E (1998) Four-dimensional quadric-based cardiac-thorax phantom for monte carlo simulation of radiological imaging systems. Submitted to *IEEE Transactions on Nuclear Science.*

Phillips, D B and Smith, A F M (1994) Bayesian faces via hierarchical template modeling. *Journal of the American Statistical Association*, **89,** no. 428, 1151–1163.

Potel, Michael J, MacKay, Steven A, Rubin, Jonathan A, Aisen, Alex M and Sayre, Richard E (1984) Three-dimensional left ventricular wall motion in man. coordinate systems for representing wall movement direction. *Investigative Radiology*, **19,** no. 6, 499–509.

Tanner, Martin A (1994) *Tools for Statistical Inference.* New York, NY: Springer-Verlag.

Wilson, Alyson G (1995) Statistical methods for shapes and deformations. Ph.D. Thesis. Duke University.

Young, R. (1986) The gaussian derivative model for machine vision: Visual cortex stimulation. *J Opt Soc Amer*, July.

Biography

Jacob Kooter Laading was born in Bergen, Norway, on December 18,1969. He received his B.S.E.E. from Rice University in Houston, Texas, in May 1993, graduating summa cum laude, Phi Beta Kappa. He received his M.S. in Biomedical Engineering from Duke University in Durham, North Carolina, in December 1996.

His graduate studies were funded in part by funding from the National Institutes of Health and subsequently a doctoral fellowship from the Norwegian Research Council (grant 11659/410). He was a winner of a student award at the 1998 International Biometric Society - ENAR meeting in Pittsburgh, Pennsylvania. He is a member of the IEEE.

Mr. Laading has authored or co-authored the following articles:

"Improved Bayesian Image Estimation for Digital Chest Radiography"(Baydush *et al.*, 1997)

"Statistical modeling of the chest radiograph and simulation in a Bayesian framework"(Laading *et al.*, 1996)

"A Shape-based Framework for Automated Image Segmentation"(McCulloch *et al.*, 1996)

"A hierarchical object deformation model applied to the digital chest radiograph"(Laading *et al.*, 1997)

"Image Feature Identification via Bayesian Hierarchical Models"(McCulloch *et al.*, 1997)

"A Hierarchical Feature Based Deformation Model Applied to 4D Cardiac SPECT Data"(Laading *et al.*, 1999)