Robust Contour Tracking in Echocardiographic Sequences

Gary Jacob, J. Alison Noble and Andrew Blake
Department of Engineering Science, Oxford University, Oxford OX1 3PJ, England.
{jacob, noble, ab}@robots.ox.ac.uk

Abstract

In this paper we present an evaluation of a robust visual image tracker on echocardiographic image sequences. We show how the tracking framework can be customised to define an appropriate shape-space that describes heart shape deformations that can be learnt from a training data set. We also investigate an energy-based temporal boundary enhancement method to improve image feature measurement. Preliminary results are presented demonstrating tracking on real normal heart motion data sequences and synthesised and real abnormal heart motion data sequences. We conclude by discussing some of our current research efforts.

1 Introduction

There has been increasing interest in analysing left ventricular function using cardiac imaging technology. The clinical demand is for real-time analysis as most pathologies manifest themselves by abnormalities in heart dynamics. Although the ideal would be real-time analysis of temporal sequences of full volumetric data (3D+T) such as MR sequence analysis [14], ultrasound tomography [13] or free-hand probe ultrasonography [16] this is not likely to be achievable at a reasonable price in the near future. Hence, there is considerable clinical interest in developing methods to perform real-time quantification of regional heart function based on an analysis of 2D image sequences (2D+T) of echocardiograms [6, 12].

In this paper we present an experimental evaluation of a robust visual image contour tracker [1, 2] on extended echocardiographic image sequences. The potential attraction of this method relate to tracking robustness – the approach can track well in the presence of clutter (which includes distracting structures as well as large amounts of spurious sensor noise and imaging artifacts). It achieves this robustness by restricting the class of allowable motions (shape deformations) to an admissible set that has been learnt from tracking on a training data set. In particular, and unlike previous approaches [3, 6], working on extended sequences allows us to directly estimate *temporal* characteristic parameters such as periodicity and asynchronousy. Further, a robust tracker can accommodate part of a contour going out of the measurement window for a limited time. This is an attractive feature in echocardiographic image sequence analysis as due to twisting of the heart a section of the ventricle boundary wall may rotate out of the plane of the sector scan over part of the cardiac cycle. Methods based on tracking image features detected in single frame echograms [11, 3] can not do this.

The ultimate goal of this work is to develop the tracking framework as a basis for regional heart function assessment of ischemia and infarcted heart disease. This paper reports on our first studies in this area. We evaluate the limitations of the visual tracking framework for echogram analysis and then go on to describe how to adapt it to better meet the needs of echogram analysis. We are currently working on the full implementation of some of these ideas. Future work will focus on classification issues and extending the ideas to 4D (3D+T) analysis.

The outline of the paper is as follows. In section 2 we briefly review the key ideas behind the tracking algorithm. Section 2.1 explains how shape deformations are defined and can be estimated from training sequences. Section 2.2 considers the tracking model and how tracking dynamics can be estimated. Tracking experiments comparing different models of tracking dynamics are presented in Section 3. In an attempt to improve tracking performance by enhancing the measurement process, Section 4 presents results of applying energy-based filtering and temporal-based noise-reduction methods. Further results on other heart image sequences are given in Section 5. We conclude, in Section 6, with a discussion of directions of current and future work.

2 Theory

Blake's contour tracking algorithm is based on a combination of active shape modelling and stochastic methods for tracking non-rigid objects over time [1, 2]. The former encompasses the observation that the shape of an object can vary considerably over time, and between object instances. A flexible model, or deformable template, is used to allow for some degree of variability in the shape of the imaged object. The model aims to capture the natural variability within a class of shapes. The tracker can learn classes of motion (shape deformation) from a train-

ing set. To track an object, in our case a left ventricle, a flexible and robust shape model is propagated over time using stochastic differential equations, whose parameters are learnt from image training-sequences. In echocardiographic image tracking this is especially challenging because of speckle noise and artifacts of the imaging process.

2.1 Shape-space model

To begin with we need to be able to represent shape deformation of an object, which we assume is a non-rigid contour (B-spline). This is done using the concept of a shape-space. A shape-space is a linear mapping of a "shape-space vector" X to a spline-vector Q,

$$Q = WX + Q_0, \qquad (1)$$

where W is called a shape matrix. The elements of X act as weights on the columns of W. Q_0 is a constant offset, for example, a mean shape. As an example, a planar affine shape space is described by,

$$W = \begin{pmatrix} 1 & 0 & Q_0^x & 0 & 0 & Q_0^y \\ 0 & 1 & 0 & Q_0^y & Q_0^x & 0 \end{pmatrix}, \qquad (2)$$

where Q_0^x are the x template coordinates of control points, chosen with the centroid at the origin; similarly for Q_0^y. Here the first two columns of W represent horizontal and vertical translation. The third and fourth columns represent scaling (width and height respectively). The last two columns deal with rotation. Rather than using Equation 1 or Equation 2 it is possible to apply a principal component analysis [10] to the data to determine the size of space that could be used to represent the motion (or shape deformation) of the object. The advantage is that the resulting W matrix is finely tuned to the deformations of the object of interest, in our case the left ventricle. The disadvantage is that interpretation of the resulting W matrix is less clear. We return to this point in Section 3.

2.2 Tracking and training

To track an object, a shape model is propagated over time using stochastic differential equations. Tracker dynamics can be described by a second order autoregressive model which can be written in discrete form as,

$$X(t_{k+2}) - \overline{X} = A_0(X(t_k) - \overline{X}) + A_1(X(t_{k+1}) - \overline{X}) + B_0 w_k. \qquad (3)$$

A Kalman filter framework [5] is used to iteratively update the tracking algorithm using a prediction-update strategy. The prediction step updates the motion based on the model of the tracker dynamics. This prediction is then corrected in the update step using information provided by the measurement process. In the original tracker implementation measurements are made along the normals to the present estimate of the contour to save computational expense.

Features are detected by applying a one-dimensional gradient operator along the sampled normals and selecting the strongest response as the most probable feature.

Training: In Equation 3, matrices A_0, A_1 and B_0 govern the behaviour of the tracking algorithm and can either be set by specifying 'reasonable' default dynamics or learnt from extended training sequences. In practice, choosing a set of good default dynamics is time-consuming and problematic and training is necessary. Suppose that we are given a training sequence of data. We can estimate B (or equivalently B_0) by noting that the covariance of the data set is $C = BB^T$. The procedure for finding the coefficient matrices for A_0 and A_1 is a little more complicated [2].

Briefly, first a principal component analysis is applied to the data to estimate W. This is done in order to restrict the state space to a low-dimensional subspace during training to avoid overfitting. Training data is collected by tracking an ultrasound sequence using a tracker with good default dynamics. The learning exercise is then to estimate the coefficients A_0, A_1 and B_0 from this training sequence of spline contours. The discrete-time system parameters are estimated via Maximum Likelihood estimation (MLE). Assuming that the noise is Gaussian, it is straightforward to set up and maximise the likelihood function.

3 Tracking experiments

A series of experiments were performed to compare tracking performance using different models of system dynamics and training strategies. Data for these experiments was acquired using a HP SONOS 1000 ultrasound machine at the John Radcliffe Hospital, Oxford. The data was recorded on VHS video and then digitised.

3.1 Shape-space estimation:

An experiment was conducted to compute the W matrix for an echocardiographic data set. The peak of the ECG R-wave was chosen as the starting point to a cardiac cycle. The first frame of the image sequence was then manually segmented. The resulting spline, with 14 control points defined the initial template. Four non-consecutive cycles were selected in this way, with the aim of obtaining a representative sample of heart cycle variations. Table 1 summarises the results of PCA analysis. For this data set, four modes explained 95% of the variation.

We can express any shape in a training set as an initial template plus a multiple of the estimated W matrix. As we have seen in the last section we can chose W via principal component analysis, such that the $N\%$ of the variability is explained by the first k eigenvalues. (in example 1, $k = 4, N = 95$).

It is possible to take the mean shape Q_0 and add to it multiples of each mode to see what that particular mode represents. Equation 1 becomes,

$$Q = Q_0 + v_i m \sqrt{\lambda_i}$$

where, v_i is the ith eigenvector, λ_i is the ith eigenvalue which represents the sample variance of X, and m is a scalar usually varying between 1 and 3.

Mode	Eigenvalue	Variability %	Cumulative %
1	360197.5	0.712	0.712
2	60381.77	0.119	0.832
3	41323.56	0.0817	0.913
4	21986.23	0.0435	0.957

Table 1: *The results of applying a principal component analysis to 4 manually segmented cardiac cycles. 4 modes of variation explain over 95% of the variability.*

Plots of the first four modes for this example are shown in Figure 1 (top). The thicker contour is the mean shape curve. The two thinner curves represent the mean shape ± 3 standard deviations. The first mode appears to be a translation mode. The second mode appears to be a scaling mode where the scaling applies to the bottom of the left ventricle next to the mitral valve. The third and fourth modes both appear to represent a combination of scaling and translation.

An alternative way of visualising the modes of variation is depicted in Figure 1 (bottom). Here flow vectors have been used to indicate the deformation for selected points along the contour. In this figure each flow vector is centred on a point on the mean shape. The ends of the flow vectors are located at $\pm$ three standard deviations from the mean shape taken in the direction of the shape deformation. The attraction of this method of visualisation is that it can be used to highlight the degree of scaling, translation and rotation for a general shape deformation. This can be difficult to determine by simply plotting the shape modes (Figure 1 (top). For example, in Figure 1 (bottom) it is clearer now that although mode 1 is predominately a translation mode there is also a small rotation component. Mode 4 shows a strong horizontal translation component.

Recall from Equation 1 that the shape-space model is given by, $Q = WX + Q_0$. X can be recovered as, $X = W^+(Q - Q_0)$, where $W^+ = (W^T HW)^{-1} W^T H$ is the pseudo-inverse of W. Figure 2 shows plots of the shape-space vector X over time. Note in particular the period-icity of the second mode. Temporal plots of this kind are potentially of great clinical value for quantifying heart periodicity and asynchronousy. We plan to investigate this idea in future work.
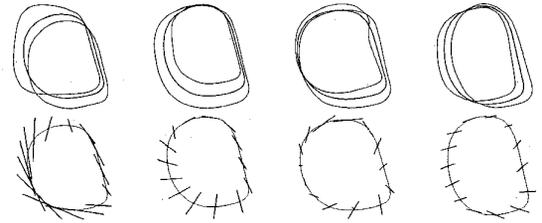

Figure 1: *Principal component analysis performed on 4 cardiac cycles of a ultrasonic image sequence. Top: The mean shape (thick curve) is plotted along with curves representing the addition of ± 3 standard deviations to the mean shape mode; From left, mode 1 (the dominant mode) to mode 4. Bottom: The mean shape (filled line) is plotted along with flow lines representing how the start of each span behaves with the addition of ± 3 standard deviations to the mean. From left, mode 1 (the dominant mode) to mode 4.*
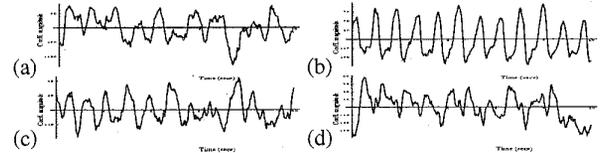

Figure 2: *Plots of the four components of X over one cardiac cycle; from (a) to (d), components 1 to 4. Each component is plotted against the time for the image sequence.*

3.2 Can we assume an affine mode of deformation?

Recall from Section 2.1 that it is possible to define the W matrix with varying degrees of freedom (dimensionality). A low dimensional space, such as an affine space, is attractive as it is easier to compute and offers an intuitive interpretation. All prior work on tracking hearts in 2D image sequences has assumed this model. On the other hand a higher dimensional space might be necessary for accurately characterising deformation and tracking. An experiment was conducted to investigate how close a W matrix estimated using PCA and training was to an affine space. The purpose of this experiment was to see whether a higher dimensional space was really necessary for characterising heart dynamics.

The residual r defined as,

$$r = \frac{||v_i - W_A W_A^+ v_i||_2}{||v_i||_2}$$

was used as the similarity metric. Here v_i is an eigenvector of the PCA W matrix, W_A is an affine shape matrix and W_A^+ is its corresponding pseudo-inverse.

Table 2 summarises the residuals computed for the first four modes of the normal heart image sequence PCA W matrix. This shows that although modes 1,2 and 4 are fairly

410

close to affine components, only 12% of mode 3 can be explained by an affine deformation. The importance of this results is that it tells us that the dynamics of the left ventricular boundary cannot be modelled well by an affine deformation.

Eigenvector	$\frac{\|v_i - W_A W_A^- v_i\|_2}{\|v_i\|_2}$	$\left(\frac{\|v_i - W_A W_A^+ v_i\|_2}{\|v_i\|_2}\right)^2$
v_1	0.3411	0.1163
v_2	0.2931	0.0859
v_3	0.9392	0.8821
v_4	0.4811	0.2315

Table 2: *Projecting a* **W** *matrix obtained using PCA into an affine space. Shown is the residual after projecting into the affine space,* $\frac{\|v_i - W_A W_A^+ v_i\|_2}{\|v_i\|_2}$.

3.3 Comparing shape models: An alternative way to compare how well different shape-models capture heart dynamics is to perform a visual inspection of tracking performance. An experiment was performed to compare tracking results using (1) a *W* matrix chosen to correspond to an affine shape matrix, (2) a *W* matrix estimated using PCA and (3) a *W* matrix estimated using PCA followed by training.

Figure 3 shows 'snapshot' views of tracking using the three approaches on three consecutive frames. The main conclusion that we could draw from this experiment was that tracking based on methods (2) and (3) gives superior results to method (1) in terms of how closely the tracker followed the observed heart chamber boundary movement. This indicates that that heart dynamics are not well modelled by a (simple) affine model. Training - method (3) - did appear to be slightly more resilient to spurious features and was less sensitive to parts of the contour fading out of the measurement window over part of the cardiac cycle. However, this approach is computationally more expensive. It was also very apparent from this study that further improvement in tracking performance could only be achieved by enhancing the image feature detection process. We consider this next.
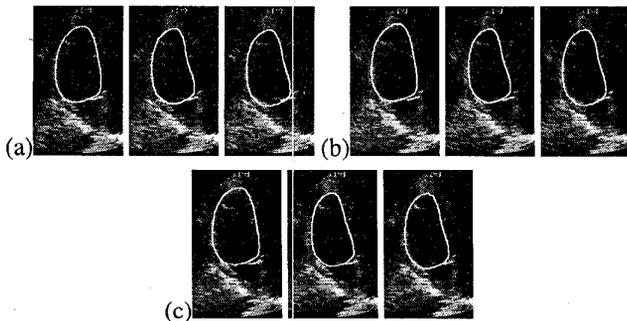

Figure 3: *Echogram tracking using affine W matrix (left), W matrix from PCA (middle), trained tracker using W matrix from PCA (right). (a) - (c) Frames 44, 45, and 46.*

4 Improving feature detection

In this section we turn our attention to improving the measurement process. We consider two ways in which this can be done; replacing the *visual edge* feature detector by an *acoustic boundary* feature enhancement operator; and using *spatio-temporal based* speckle noise reduction prior to boundary detection.

The visual image tracking algorithm uses a gradient-based operator for detecting contour points which produces many candidate responses on ultrasound images. This is due partly to the low signal-to-noise ratio and poor image contrast. There is, however, something fundamentally wrong with using such an approach. A gradient-based feature detector is designed to detect an object boundary as a *step* discontinuity in intensity. However, an ideal acoustic edge is defined as a discontinuity in acoustic impedance (an intensity ridge) or equivalently a discontinuity in acoustic energy or integrated backscatter (IBS). Formerly, integrated backscatter is defined as,

$$S = \frac{\int_{\tau-\Delta\tau}^{\tau+\Delta\tau} |V(t)|^2 dt}{\int_{\tau-\Delta\tau}^{\tau+\Delta\tau} |P(t)|^2 dt}$$

where S indicates the integrated backscatter measure, $V(t)$ and $P(t)$ are the signals received from the tissue and the reflector, respectively, τ is the position in time of the centre of the region of interest in the tissue and $\Delta\tau$ is the corresponding half-width in time. In the spatial domain, integrated backscatter can be estimated simply as the average squared greylevels over a neighbourhood. The IBS model has been used to measure properties of myocardial tissue since changes in integrated backscatter relate to changes in acoustic impedance, and hence changes in tissue relaxation and elasticity [8]. IBS boundary enhancement is also used in state-of-the-art commercial echocardiographic imaging systems for single view real-time edge detection.

We investigated tracking heart boundaries on image sequences pre-filtered by an IBS algorithm to see if this improved tracking performance. IBS filtering removed a significant amount of speckle noise within the left ventricle chamber and enhanced heart chamber boundaries but the result was blurred relative to the original. Overall there were less spurious feature responses but localisation of heart contours did not improve due to spatial blurring.

A natural question to ask is whether spatio-temporal filtering can improve ultrasound feature detection relative to a static image feature detection approach. Herlin and Ayache explored this idea for *step edge* detection in ultrasound images [6]. They assumed a step model and Gaussian noise.

We implemented a spatio-temporal *acoustic boundary* detection scheme based on a combination of spatio-

411

temporal speckle reduction filtering [4] and acoustic energy discontinuity detection. Figure 4 shows one frame together with IBS enhanced and spatio-temporal speckle-reduced/IBS enhanced versions of the same images. A frame-by-frame visual comparison of tracking using the three approaches showed that the reliability of detection of boundaries was best for the spatio-temporal boundary detection algorithm. Quantifying the degree of improvement is difficult because we do not have any ground truth by which to compare the algorithms. However, since we can estimate the state vectors we can look for consistency of state vector component trajectories over a number of cycles as a measure of algorithm robustness to measurement noise. In figure 5 we show coefficient plots based on tracking on the original data (light curve) and the data pre-filtered by the spatio-temporal boundary enhancement method (dark curve). Observe that the plots are more consistent for the algorithm which used the spatio-temporal filtering approach.

The current approach involving the sequential application of a temporal-based noise-reduction filter followed by static image feature detection gives improved tracking performance. However, further improvement could be achieved if temporal information was utilised in the detection step. We are currently exploring methods to extend the idea of spatio-temporal acoustic boundary detection to a truly 3D (2D+T) filtering process.

Figure 4: *Noise reduction. The original ultrasound image (left), the image after the two-dimensional least mean square (TDLMS) filter is applied (middle), TDLMS filter followed by IBS (right).*
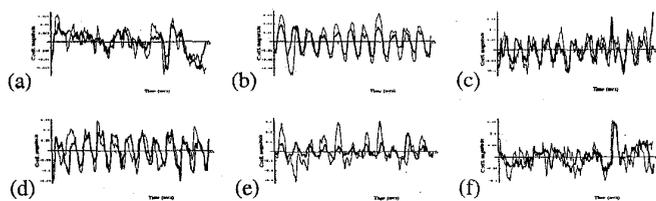

Figure 5: *Comparison of component plots for tracking on ultrasound data — original data (light curve) and enhanced data (dark line) using an affine W matrix. Each component is plotted against the time for the image sequence.*

5 Towards classification

The ultimate goal of this work is to demonstrate that automated image analysis can be used for to perform temporal-based quantification of regional heart function. As a step towards this goal, in this section we present some results of applying the training and tracking procedures outlined in Section 3 to some further real heart image sequences.

5.1 A synthesised abnormal heart: Abnormal heart motion was simulated by editing images corresponding to the diastole section of the cardiac cycle for a normal heart image sequence (the example used in Section 3) so that the posterior (left) wall appeared sluggish. This was done by shifting an image block containing the posterior wall from each of the diastole images by 10 pixels to the right. The image block was then blended in with the data using an exponential weight function. Finally a PCA was performed as before. Figure 6 summarises the results. The key thing to observe is that the nature of the principal modes remain unchanged although the magnitude is affected (compare with Figure 1). In particular the second mode (middle plot) shows that the posterior wall scales outwards to a lesser degree which is consistent with the imposed abnormality.

Figure 6: *Principal component analysis performed on 4 cardiac cycles of a ultrasonic image sequence. The actual data (light curve) is plotted along with the simulated data (dark curve). The mean shape (left) is plotted. The modes represent the addition of ±3 standard deviations to the mean shape; From second left mode 1 to mode 4.*

5.2 A real abnormal heart: A PCA was performed on four manually segmented non-consecutive cycles of real data for a patient diagnosed with a disease which manifests as a loss in elasticity of the heart. Figure 7 summarises the results of the PCA. In this case seven modes of variation express 95% of the variability as compared to four modes with the normal heart. The first mode appears to be a scaling of the anterior wall, the second mode a translation mode, the third mode a scaling, the fourth mode a mixture of a translation and scaling and the fifth mode is translation.

Figure 7: *First five principal modes for an abnormal heart. The mean shape (filled line) is plotted along with flow lines representing how the start of each span behaves with the addition of ±3 standard deviations to the mean. From left, mode 1 to mode 5.*

Mode	Eigenvalue	Variability %	Cumulative %
1	669.51	0.321	0.321
2	544.00	0.294	0.615
3	284.00	0.139	0.755
4	169.4	0.115	0.871
5	85.29	0.0546	0.925
6	45.54	0.0185	0.944
7	31.35	0.0139	0.958

Table 3: *The results of applying a principal component analysis to 4 manually segmented cardiac cycles from an abnormal heart. 7 modes of variation explain over 95% of the variability.*

6 Discussion and future work

In this paper we have presented the results of an evaluation of a robust visual image tracker on echocardiographic image sequences. Our preliminary results are encouraging. We now plan to further investigate how this tracking framework can be developed into a clinical tool for automated regional heart function analysis.

Our current efforts are directed in three key areas: developing alternative training strategies and generalising the class of motions that the tracking algorithm can handle; improving the detection of image features; and developing further insight into the clinical interpretation of the deformation parameters.

In Section 2 we outlined how the tracking algorithm is trained using a single-step estimation of the system matrices A_0, A_1 and B_0. We are currently investigating how we could use a related idea to build a generic model of heart motion from training data based on an average model. We plan to investigate how well this type of model can represent the dynamics of a normal heart and different heart conditions. Clearly the general model is not going to track a specific heart as well as a tracking algorithm tuned specifically to an individual heart. However, the goal here is to provide a general enough description of heart dynamics that can be used in conjunction with a robust feature detector to provide robust tracking results.

It is clear that the main way in which tracking performance can be further improved is through the development of new methodology for robust acoustic boundary feature measurement. In Section 4 we found that spatio-temporal noise reduction improved image feature detection. We plan to investigate methods, possibly based on energy filters [9] (wavelets) and anisotropic diffusion [15, 17], to extend the idea of spatio-temporal acoustic boundary detection to a truly 3D (2D+T) filtering process.

Finally, the ultimate measure of the success of this work will be to demonstrate that it is possible to relate the tracking parameters to clinical meaningful descriptors of the cardiac performance. We plan to evaluate the clinical po-

tential of our algorithms using the objective quantification of ischemic heart disease and stress testing as example cardiac application domains.

References

[1] Blake, A., Curwen, R. and Zisserman, A., A framework for spatio-temporal control in the tracking of visual contours. *Int. J. Comp. Vision*, 11, pp. 127-145, 1993.

[2] Blake, A., Isard, M. and Reynard, D. Learning to track the visual motion of contours. *Art. Intell.*, 78, pp. 179-212, 1995.

[3] Cootes, T. F., Taylor, C. J., Cooper, D. H and Graham, J., Active shape models - their training and application. *CVGIP*, 61, pp. 38-59, 1995.

[4] Evans, A. N. and Nixon. M. S., Biased motion-adaptive temporal filtering for speckle reduction in echocardiography. *IEEE* TMI., 15:1, pp. 39-50, 1996.

[5] Gelb, A., *Applied Optimal Estimation*. MIT Press, Cambridge MA, 1974.

[6] Herlin, I. and Ayache, N., Feature extraction and analysis methods for sequences of ultrasound images, in *Proceedings ECCV*, pp. 43-57, 1992.

[7] Horn, B. K. P. and Schunck, B. G., Determining optic flow. *Art. Intell.*, 17, pp. 185-203, 1981.

[8] Lange,A. et. al, The variation in integrated backscatter in human hearts in differing ultrasonic transthoracic views. *J. Am. Soc. Echo.*, pp. 830-838, 1995.

[9] Kovesi, P., *Invariant measures of image features from phase information*. PhD thesis, U. West. Australia, May 1996.

[10] Krzanowski, W. J., *Principles of multivariate analysis*. OUP, 1988.

[11] Mailloux, G. E., Bleau, A., Bertrand, M. and Petitclerc, R., Computer analysis of heart motion from two-dimensional echocardiograms. *IEEE Trans. Biomed. Eng.*, 34:5, pp. 356-364, 1987.

[12] McEachen, J. C. II and Duncan, J. S., Shape-based tracking of the left ventricular wall motion. *IEEE Trans. Medical Imaging*, 16:3, pp.270-283, 1997.

[13] Ofili, E. O. and Nanda, N. C., Three-dimensional and four-dimensional echocardiography. *Ultrasound in Med. and Biol.*, 20:8, pp. 669-675, 1994.

[14] Park, J., Metaxas, D. and Young, A., Analysis of left ventricular wall motion based on deformable models and MRI-SPAMM. *MedIA*, 1:1, pp. 53-71, 1996.

[15] Perona. P and Malik J., Scale space and edge detection using anisotropic diffusion. *IEEE Pattern Anal. Machine Intell.*, 12:7, pp. 629-639, 1990.

[16] Rohling, R. N, Gee, A. H. and Berman, L., 3-D spatial compounding of ultrasound images. *MedIA*, 1:3, pp. 177-193, 1997.

[17] Sanchez-Ortiz, G. I., Rueckert, D. and Burger, P., Knowledge-based anisotropic diffusion of vector valued 4-D cardiac MR images. *In Proc. BMVC*, pp. 605-614, 1996.