Medical Image Analysis (1999) volume 3, number 1, pp 63–75

Evaluating a robust contour tracker on echocardiographic sequences

Gary Jacob*, J. Alison Noble, Miguel Mulet-Parada and Andrew Blake

Robotics Research Group, Department of Engineering Science, Parks Road, University of Oxford, Oxford OX1 3PJ, 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 customized to define an appropriate shape space that describes heart shape deformations that can be learnt from a training data set. We also investigate energy-based temporal boundary enhancement methods to improve image feature measurement. Results are presented demonstrating real-time tracking on real normal heart motion data sequences and abnormal synthesized and real heart motion data sequences. We conclude by discussing some of our current research efforts.

Keywords: 2-D+T ultrasound, echocardiography, energy filters, heart wall motion, real-time tracking

Received October 9, 1997; revised February 19, 1998; accepted May 18, 1998

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 (3-D+T) such as magnetic-resonance (MR) sequence analysis (Bardinet *et al.*, 1996; Park *et al.*, 1996; Declerk *et al.*, 1998), ultrasound tomography (Ofili and Nanda, 1994) or free-hand probe ultrasonography (Salustri and Roelandt, 1995; Rohling *et al.*, 1997) 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 two-dimensional (2-D) image sequences (2-D+T) of echocardiograms (Mailloux *et al.*, 1987; Chu *et al.*, 1988; Han *et al.*, 1991; Herlin and Ayache, 1992; Herlin *et al.*, 1994; Chalana *et al.*, 1996; Dias and Leitão, 1996; McEachen and Duncan, 1997).

In this paper we present an experimental evaluation of a robust visual image contour tracker (Blake *et al.*, 1993, 1995) on extended echocardiographic image sequences. The potential attraction of this method relates 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 artefacts). It achieves this robustness by restricting the class of allowable motions (shape deformations) to an admissible set that has been learnt from tracking a training data set. In particular, and unlike previous approaches (Herlin and Ayache, 1992; Cootes *et al.*, 1995), working on extended sequences allows us to directly estimate temporal characteristic parameters such as periodicity and asynchronicity. This is a particularly challenging task for a sonographer to perform by visual methods. 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 (Mailloux *et al.*, 1987; Cootes *et al.*, 1995) cannot do this. Finally, we use a method of feature measurement which is invariant to intensity amplitude. This property is attractive for echogram segmentation as contour intensities vary widely depending on acquisition settings and over the cardiac cycle.

The ultimate goal of this work is to develop a tracking framework as a basis for regional heart function assessment of ischemia and infarcted heart disease. In this paper we evaluate the limitations of the visual tracking framework for

*Corresponding author
(e-mail: jacob@robots.ox.ac.uk)

echogram analysis and then proceed to describe how to adapt it to better meet the needs of echogram analysis. Future work will focus on classification issues and extending the ideas to four-dimensional (3-D+T) analysis.

The outline of the paper is as follows. In Section 2 we briefly review the key ideas behind the tracking algorithm. Subsection 2.1 explains how shape deformations are defined and can be estimated from training sequences. Subsection 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. Section 4 shows how to improve tracking performance by enhancing the measurement process via energy-based filtering and temporal-based noise-reduction methods. In particular, this leads us to develop a novel approach to feature localization based on local-phase information (Kovesi, 1996) in Subsection 4.1, which we show in Subsection 4.2, gives significantly better results than the classical derivative-based approach. 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. Part of this research was first presented by Jacob and Noble (1997).

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 (Blake *et al.*, 1993, 1995). 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 training set. To track an object, in our case a left ventricle (LV), 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 artefacts 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' $\mathbf{X}$ to a spline-vector $\mathbf{Q}$,

$$\mathbf{Q} = \mathbf{W}\mathbf{X} + \mathbf{Q}_0, \tag{1}$$

where $\mathbf{W}$ is called a shape matrix. The elements of $\mathbf{X}$ act as weights on the columns of $\mathbf{W}$. $\mathbf{Q}_0$ is a constant offset, for example, a mean shape.

As an example, a planar Affine shape space is described by

$$\mathbf{W} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{Q}_0^x & \mathbf{0} & \mathbf{0} & \mathbf{Q}_0^y \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{Q}_0^y & \mathbf{Q}_0^x & \mathbf{0} \end{pmatrix}, \tag{2}$$

where $\mathbf{Q}_0^x$ are the x coordinates of the template control points; similarly for $\mathbf{Q}_0^y$.

In Equation (2) the first two columns of $\mathbf{W}$ 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 (PCA) (Krzanowski, 1988) 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 $\mathbf{W}$ matrix is finely tuned to the deformations of the object of interest, in our case the LV. The disadvantage is that interpretation of the resulting $\mathbf{W}$ 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

$$\mathbf{X}(t_{k+2}) - \overline{\mathbf{X}} = \mathbf{A}_0(\mathbf{X}(t_k) - \overline{\mathbf{X}}) + \mathbf{A}_1(\mathbf{X}(t_{k+1}) - \overline{\mathbf{X}}) + \mathbf{B}_0\mathbf{w}_k \tag{3}$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{I} \\ \mathbf{A}_0 & \mathbf{A}_1 \end{pmatrix}, \qquad \mathbf{B} = (0, \mathbf{B}_0)^{\mathrm{T}}, \tag{4}$$

and $\mathbf{w}_k$ is Gaussian noise.

A Kalman filter framework (Gelb, 1974) 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. Image features are detected by applying a one-dimensional (1-D) gradient operator along the sampled normals and selecting the strongest response as the most probable feature.

In Equation (3), matrices $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{B}_0$ 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 $\mathbf{B}$ (or equivalently $\mathbf{B}_0$) by noting that the covariance of the data set is $\mathbf{C} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$. The procedure for finding the coefficient matrices for $\mathbf{A}_0$ and $\mathbf{A}_1$ is a little more complicated (Blake *et al.*, 1995). Briefly, first a PCA is applied to the data to estimate $\mathbf{W}$. This is done in order to restrict the state space to a low-dimensional subspace. 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 $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{B}_0$ 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 maximize 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 digitized. Data from one normal patient and one abnormal patient have been used throughout this paper.

3.1. Shape-space estimation

An experiment was conducted to compute the $\mathbf{W}$ matrix for an echocardiographic data set. The peak of the ECG R-wave was chosen as the starting point to a cardiac cycle. To provide a representative sample of heart-cycle variations, four non-consecutive cardiac cycles were manually segmented using a B-spline with 14 control points. Thus the total data set was a matrix of 4×21 rows and 28 columns. Note that the training sets have not been normalized to take out the global transformation. This is beneficial for tracking purposes but makes the interpretation of the PCA in terms of physical motions more complicated. Table 1 summarizes the results of PCA. 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 $\mathbf{W}$ matrix. As we have seen in the last section we can choose $\mathbf{W}$ via PCA, 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 $\mathbf{Q}_0$ and add to it multiples of each mode to see what that particular mode represents. Equation (1) becomes,

$$\mathbf{Q} = \mathbf{Q}_0 + \mathbf{v}_i m\sqrt{\lambda_i}$$

where $\mathbf{v}_i$ is the ith eigenvector, λ_i is the ith eigenvalue which represents the sample variance of $\mathbf{X}$, and, m is a scalar, usually a fixed value between 1 and 3.

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 (sd). 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 LV next to the mitral valve. The third and fourth modes both appear to represent a combination of scaling and translation.

An alternative way of visualizing the modes of variation is depicted in Figure 1 (bottom). Here flow vectors have been

Table 1. The results of applying a PCA to four manually segmented cardiac cycles. Four modes of variation explain >95% of the variability.

Mode	Eigenvalue	Variability (%)	Cumulative (%)
1	360 197.5	0.712	0.712
2	60 381.77	0.119	0.832
3	41 323.56	0.0817	0.913
4	21 986.23	0.0435	0.957

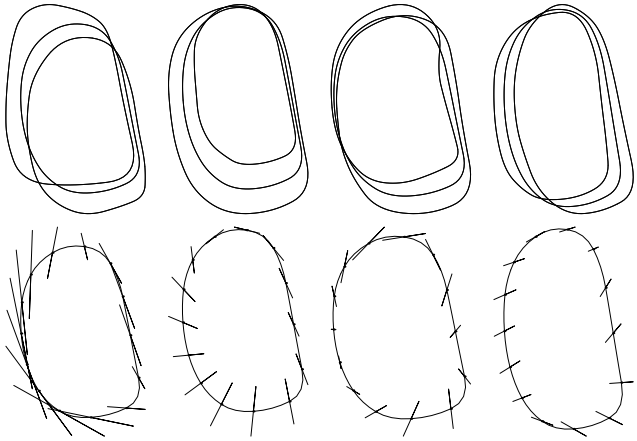


Figure 1. PCA performed on four cardiac cycles of an ultrasonic image sequence. Top, the mean shape (thick curve) is plotted along with curves representing the addition of ± 3 sd to the mean shape mode; from the left, mode 1 (the dominant mode) to mode 4. Bottom, the mean shape (filled line) is plotted along with flow lines representing how the start of each span behaves with the addition of ± 3 sd to the mean. From left, mode 1 (the dominant mode) to mode 4.

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 ± 3 sd from the mean shape taken in the direction of the shape deformation. The attraction of this method of visualization 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

$$\mathbf{Q} = \mathbf{W}\mathbf{X} + \mathbf{Q}_0.$$

$\mathbf{X}$ can be recovered as

$$\mathbf{X} = \mathbf{W}^{+}(\mathbf{Q} - \mathbf{Q}_0),$$

where $\mathbf{W}^{+} = (\mathbf{W}^{\mathrm{T}}\mathbf{H}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{H}$ is the pseudo-inverse of $\mathbf{W}$. Figure 2 shows plots of the shape-space vector $\mathbf{X}$ over time. In particular, note the periodicity of the second mode. Temporal plots of this kind are potentially of great clinical value for quantifying heart periodicity and asynchrony. We plan to investigate this idea in future work.

3.2. Can we assume an Affine mode of deformation?

Recall from Subsection 2.1 that it is possible to define the $\mathbf{W}$ 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 2-D image sequences has assumed this model. On the other hand, a higher-dimensional space might be necessary for accurately characterizing deformation and tracking. An experiment was conducted to investigate how close a $\mathbf{W}$ 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 characterizing heart dynamics.

Let the residual r be defined as

$$r = \frac{\|\mathbf{v}_i - \mathbf{W}_{\mathrm{A}}\mathbf{W}_{\mathrm{A}}^{+}\mathbf{v}_i\|_2}{\|\mathbf{v}_i\|_2}.$$

Here $\mathbf{v}_i$ is an eigenvector of the PCA $\mathbf{W}$ matrix, $\mathbf{W}_{\mathrm{A}}$ is an Affine shape matrix and $\mathbf{W}_{\mathrm{A}}^{+}$ is its corresponding pseudo-inverse.

Table 2 summarizes the residuals computed for the first four modes of the normal heart image sequence based on a PCA $\mathbf{W}$ matrix. This shows that although modes 1, 2 and 4 are fairly close to Affine components, only 12% of mode 3 can be explained by an Affine deformation. The importance of this result is that it tells us that the dynamics of the LV boundary cannot be modelled well by an Affine deformation.

Table 2. Projecting a $\mathbf{W}$ matrix obtained using PCA into an Affine space. Shown is the residual after projecting into the Affine space, $\|\mathbf{v}_i - \mathbf{W}_{\mathrm{A}}\mathbf{W}_{\mathrm{A}}^{+}\mathbf{v}_i\|_2/\|\mathbf{v}_i\|_2$.

Eigenvector	$\frac{\|\mathbf{v}_i - \mathbf{W}_{\mathrm{A}}\mathbf{W}_{\mathrm{A}}^{+}\mathbf{v}_i\|_2}{\|\mathbf{v}_i\|_2}$	$\left(\frac{\|\mathbf{v}_i - \mathbf{W}_{\mathrm{A}}\mathbf{W}_{\mathrm{A}}^{+}\mathbf{v}_i\|_2}{\|\mathbf{v}_i\|_2}\right)^2$
$\mathbf{v}_1$	0.3411	0.1163
$\mathbf{v}_2$	0.2931	0.0859
$\mathbf{v}_3$	0.9392	0.8821
$\mathbf{v}_4$	0.4811	0.2315

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:

Click on relevant button or figure for video

(i) a $\mathbf{W}$ matrix chosen to correspond to an Affine shape matrix, Play Video

(ii) a $\mathbf{W}$ matrix estimated using PCA, and, Play Video

(iii) a $\mathbf{W}$ matrix estimated using PCA followed by training, performed on one patient. Play Video

Figure 7 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 (ii) and (iii) gives superior results to method (i) in terms of how closely the tracker followed the observed heart chamber boundary movement. This indicates that heart dynamics are not well modelled by a (simple) Affine model. Training [method (iii)] 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.

4. IMPROVING FEATURE DETECTION

In this section we turn our attention to improving the feature measurement process. We develop an alternative to gradient-based feature detection using a combination of spatio-temporal noise reduction filtering (Evans and Nixon,

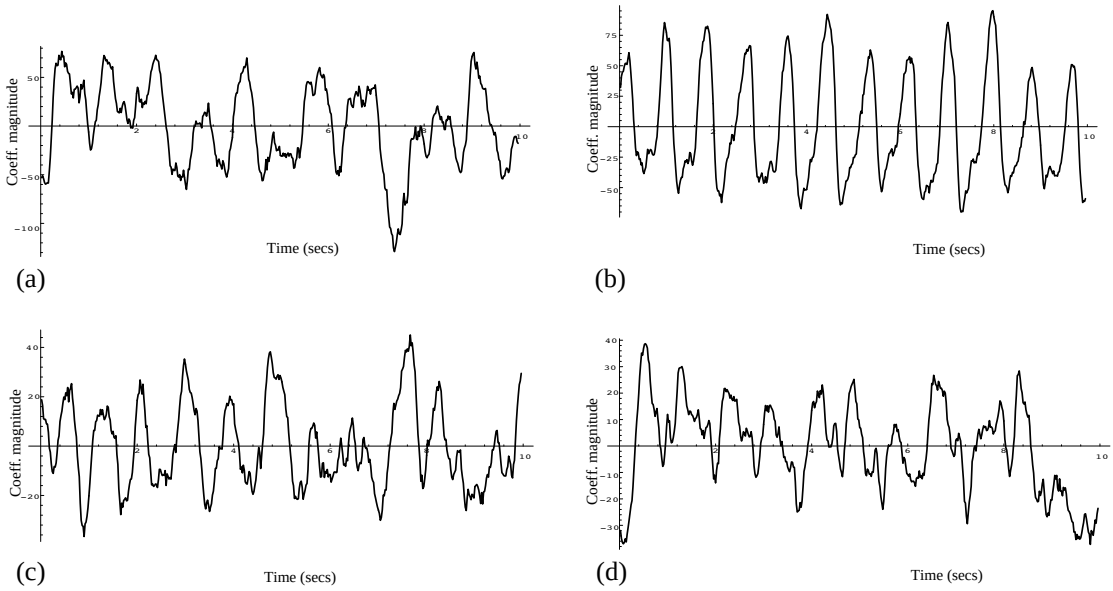


Figure 2. Plots of the four components of **X** over one cardiac cycle; from (**a**) to (**d**), components 1 to 4. Each component is plotted against the time for the image sequence.

1996) and phase-congruency ideas (Kovesi, 1996) and show how this leads to a more stable and robust measurement process in the tracking algorithm.

The original visual image tracking algorithm uses a gradient-based operator for detecting contour points. This approach produces many candidate responses on ultrasound images. This is due partly to the low signal-to-noise ratio and also to poor image contrast. A further problem relates to choosing a global edge threshold that works well at finding features in different regions of an image as well as across images captured at different points in the cardiac cycle. This is an especially challenging problem as edge intensities along the endocardial contour in an echocardiographic image can vary considerably depending on the orientation of the tissue with respect to the incident ultrasonic beam and the effect of occlusion artefacts (for example, in Figure 8a, which shows the output of a gradient-based detector on an echogram frame, a large threshold was set to eliminate noise responses on the left wall of the cavity. As a result, low-intensity features are missed on the right-hand wall). Below, we show how to significantly improve the quality of feature detection by

(i) taking into consideration temporal continuity to reduce noise effects, and,

(ii) replacing the visual edge feature detector by,

  (a) applying an acoustic boundary feature enhancement operator, and,

  (b) modifying the localization process to make it invariant to intensity.

To achieve (i) we apply the 2-D least-mean-square (TDLMS) filter proposed by Evans and Nixon (1996). Briefly, the TDLMS filter considers a given image in a sequence at time t. Images close in time to this image, for example $t-2$, $t-1$, $t+1$, $t+2$ are combined using a linear weighting scheme. The weights are chosen to minimize the least-mean-square error between the resulting output image and the corresponding image at time t.

Next, we enhance the boundaries using integrated backscatter (IBS) boundary enhancement. Recall that in visual image analysis, object boundaries are associated with step edges in intensity. Therefore boundaries can be found by localizing extrema in the output of a gradient-based operator. This is the feature measurement operator used in the original visual image tracker. However, an ideal acoustic edge is defined as a discontinuity in acoustic impedance (an intensity ridge) or equivalently a discontinuity in acoustic energy or IBS. This model has been used to measure properties of myocardial tissue since changes in IBS relate to changes in acoustic impedance which in turn relate to changes in relaxation and elasticity (Lange *et al.*, 1995). IBS boundary enhancement is also available in state-of-the-art commercial echocardiographic imaging systems for single-view real-time edge detection.

Click on relevant button below for video

Play Video Play Video Play Video

(a) (b) (c)

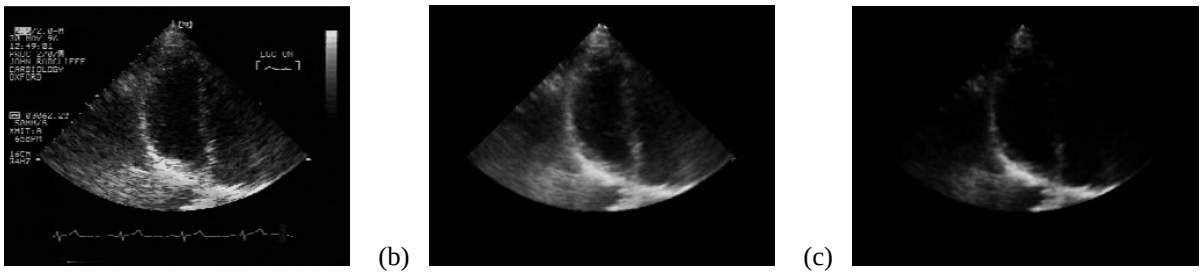


Figure 3. Noise reduction: (**a**) the original ultrasound image; (**b**) the image after the TDLMS filter is applied and (**c**) TDLMS filter followed by IBS filtering.

Figure 3a shows one frame of an echogram together with (b) an IBS-enhanced and (c) spatio-temporal speckle-reduced/IBS-enhanced versions of the same image. From Figure 3c it is clear how speckle noise is significantly reduced and endocardial borders enhanced by this two-step process.

Finally, acoustic boundaries are localized by applying an intensity-invariant localization process based on phase information. This is discussed in the next section.

4.1. Localization based on local phase information

This section outlines how feature measurements are localized using a novel approach which is invariant to intensity amplitude. This is a particularly attractive feature for echogram segmentation where finding global intensity thresholds that work well is problematic. Further details of our approach are described in Mulet-Parada (1997).

After post-processing an image using TDLMS/IBS filtering the endocardial border resembles a large-scale smooth step edge corrupted by small-scale noise (Figure 3c). A step edge can be defined in terms of the phase values of its frequency components. Venkatesh and Owens (1990) show that different 1-D image features possess different local phase signatures, where local phase is defined in terms of the local Fourier expansion of the signal $F(x)$ and its Hilbert transform $H(x)$, where

$$F(x) = \sum (a_n \sin(nwx + \phi)),$$

$$H(x) = \sum (a_n \cos(nwx + \phi)).$$

Here ω is the signal frequency, a_n the frequency component amplitude, and ϕ a phase angle. The local phase of the signal $E(x) = F(x) - \mathrm{j}H(x)$ is then defined as

$$\arg(E(x)) = \tan^{-1}(F(x)/H(x)).$$

In practice, $E(x)$ can be estimated by computing an alternative form $\hat{E}(x) = f(x) - \mathrm{j}h(x)$ where $f(x)$ and $h(x)$ are the result of convolving the original signal with

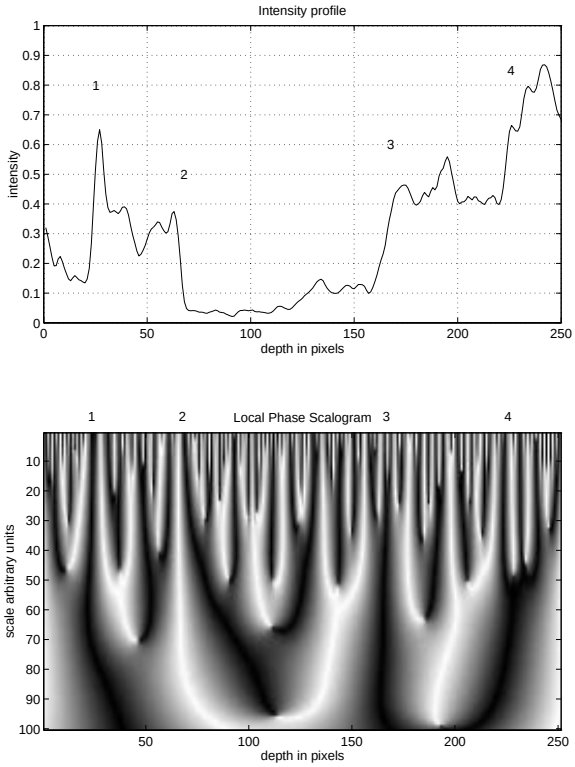


Figure 4. A 1-D profile across the LV (top); local phase scalogram (bottom). Observe the stability of phase information over scale.

a quadrature filter pair formed by a symmetric (even) and an antisymmetric (odd) filter. In their work, Venkatesh and Owens use Gabor wavelets as quadrature filters. They show that the output of these filters suffices to uniquely discriminate between different features, such as positive and negative edges, ridges and valleys, all of which possess different phase signatures. For example, a positive edge is uniquely defined as a zero crossing of the output of the even filter and a maxima in the output of the odd filter. This gives

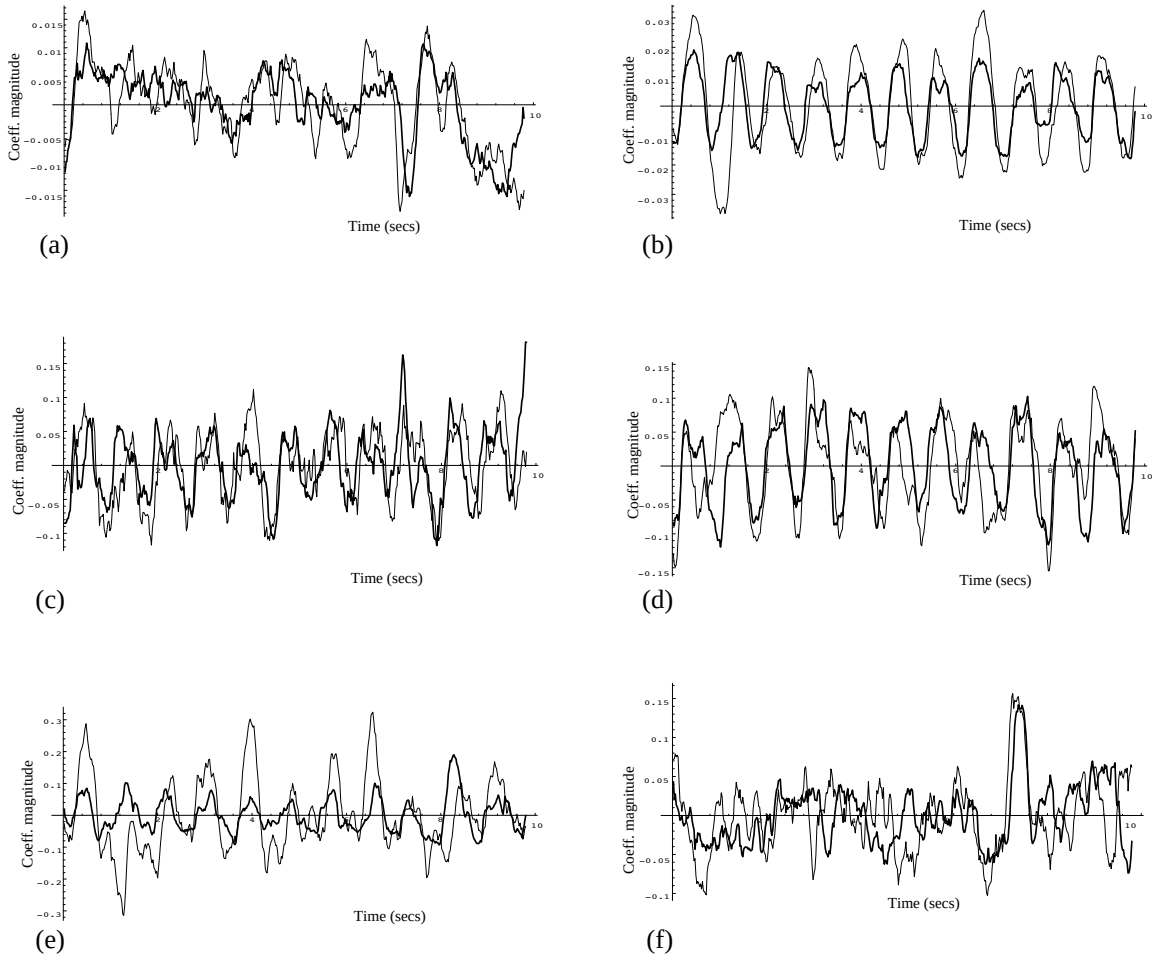


Figure 5. Comparison of component plots for tracking on ultrasound data—original data (light curve) and enhanced data (dark line) using an Affine **W** matrix. Each component is plotted against the time for the image sequence.

a local phase signature of 0 rad. Similarly, a negative edge gives a maximum in local phase (π). Thus, in our application, consider taking 1-D normals traced along the cavity boundary, starting from the interior of the ventricle outwards. A (positive-going) edge would be located as a point of zero phase in the local phase representation.

Venkatesh and Owen's analysis applies to idealized image feature characterization in the absence of image noise. Studying the properties of local phase measurements over scales allows us to adapt this scheme to localize edge features in a noisy profile (Kovesi, 1996). The wavelets we use are based on the log–Gabor transfer function, $G(\omega)$, which is defined in the frequency domain by

$$G(\omega) = \exp -\frac{(\log(\omega/\omega_0))^2}{2(\log(\kappa/\omega_0))^2}. \qquad (5)$$

Here κ is related to the bandwidth of the filter and ω_0 the centre frequency of the filter. For the two octave filters used in this work $\kappa = 0.55$. The advantage of using this representation is that the log–Gabor even function has zero power for negative frequencies. In contrast even Gabor functions to not have this property (Kovesi, 1996).

Consider taking a 1-D profile across an echogram of an LV (Figure 4, top). A scalogram can be derived from the output of a sequence of log–Gabor filters of increasing spread (scale) (Figure 4, bottom). In this figure, the absolute value of the local phase is plotted so that black corresponds to zero local phase, white corresponds to π. Observe that at small scales many features corresponding to these extreme values are found but that as the filter scale increases the number of features reduces to a few straight bands that correspond to the epicardial (1), to the endocardial (2, 3) and to the pericardial

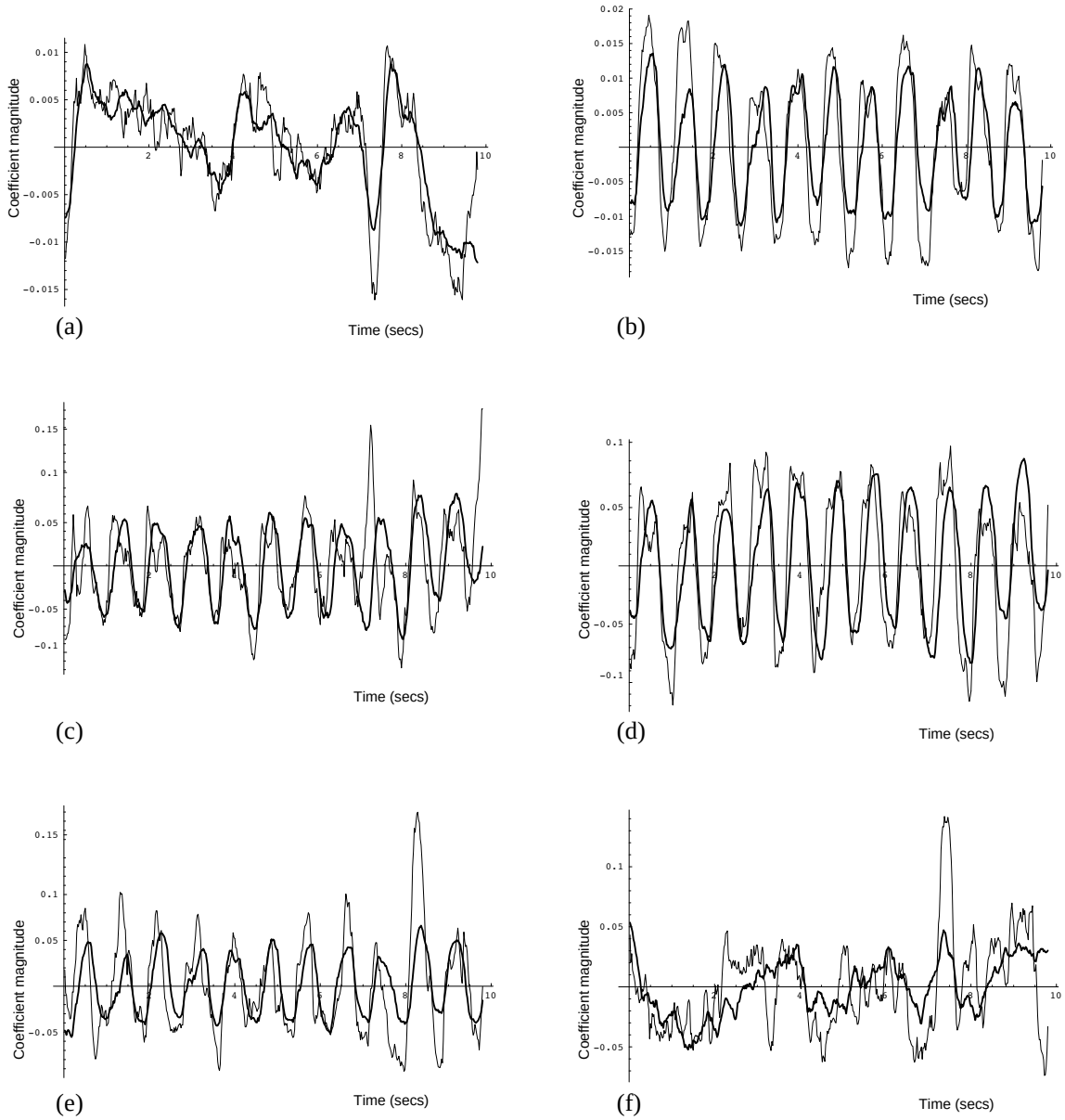


Figure 6. Comparison of component plots for tracking on ultrasound data—enhanced data with gradient-based localization (light curve) and enhanced data with local-phase based localization (dark line) using an Affine **W** matrix. Each component is plotted against the time for the image sequence.

(4) edges. It is important to stress that these features give a localized response over a wide range of scales. Thus, unlike other features, the centre of a step does not move (i.e. is stable) over a wide range of scales as long as it is 'significant' in the region of filter support. This contrasts with the poor localization performance observed for a gradient-based operator at large scales. Obviously if the filter width is very large, two steps will fuse into a roof and the localization of the phase signature will be lost. Nevertheless, provided the scale of the filter is reasonable, localization is preserved. This is shown in Figure 4 (bottom) by the straight vertical lines corresponding to the endocardial borders (2, 3).

From the scalogram it is clear that noise and signal structure occupy distinct regions of scale space. The two regions

Click on relevant button for video

Play Video

(a)

Play Video

(b)

Play Video

(c)

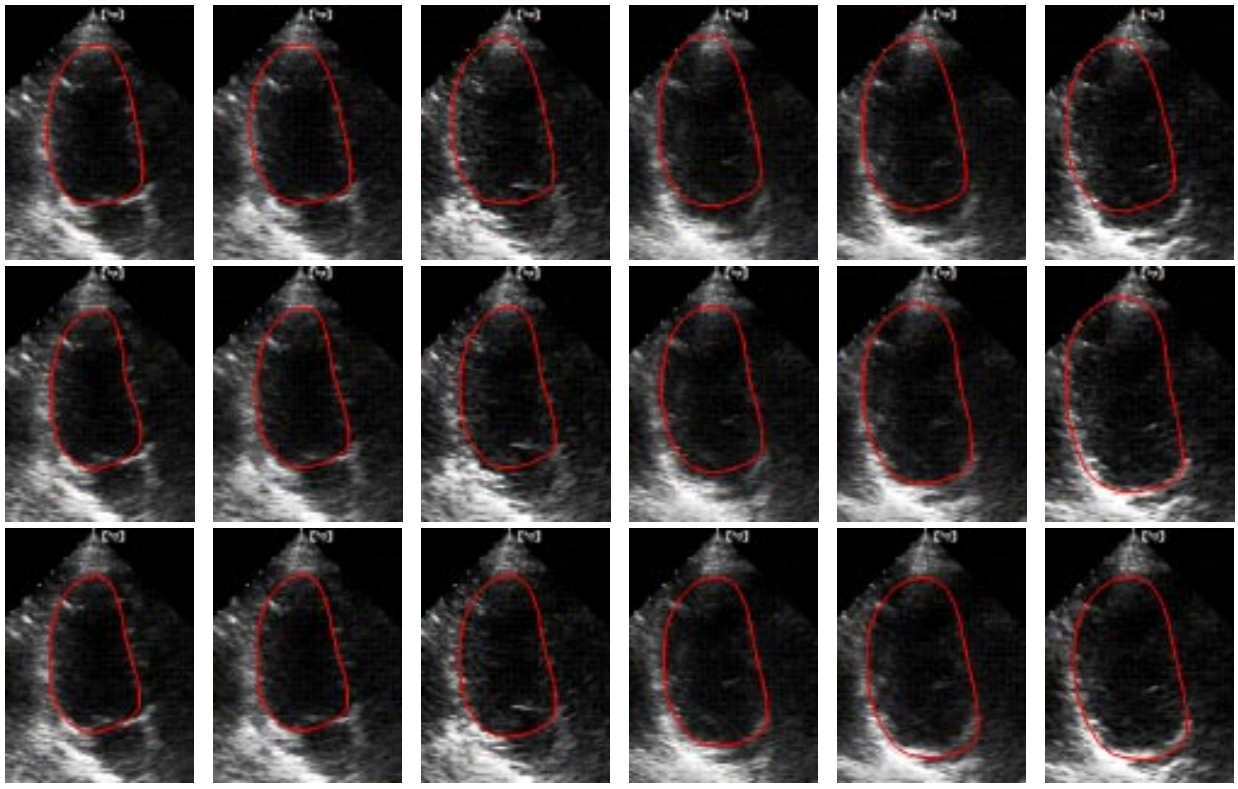
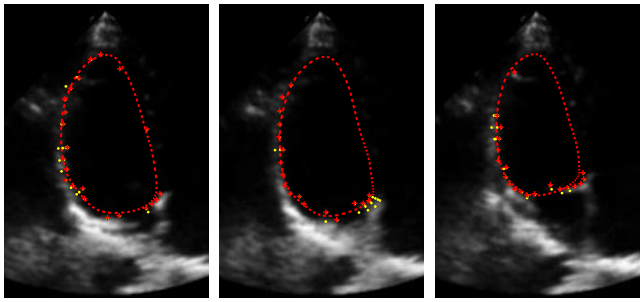


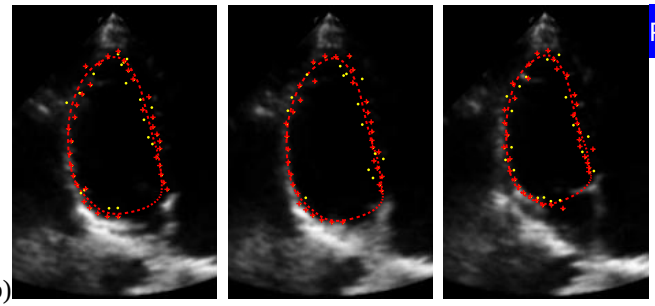
Figure 7. Echogram tracking using Affine **W** matrix (top), **W** matrix from PCA (middle), trained tracker using **W** matrix from PCA (bottom).

Play Video

(a)



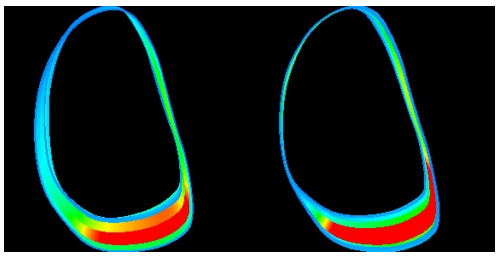
(b)



Play Video

Figure 8. Image features detected using (**a**) gradient-based information and (**b**) based on local phase information. The best candidate points along a normal are shown in red, rejected candidate points in yellow.

(a)



(b)

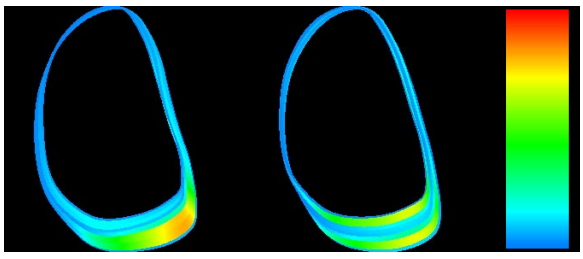


Figure 9. Velocity plots for one cardiac cycle. (**a**) The motion end-systole to end-diastole and (**b**) end-diastole to end-systole. In each case, the left-hand figure represents a normal heart sequence and the right-hand figure the synthesized abnormal heart. Colour scale: red = large movement; blue = small movement.

can be separated by a noise threshold set at the location where most small-scale features due to noise disappear. This allows us to select a region in scale space where the main features are detected but where noise effects are insignificant. This noise floor can be set experimentally with respect to the significance of the features that we want to detect.

In our application, we assume that only one target is present along each normal and use the average of the local phase results over three scales above the noise threshold to localize a candidate boundary point. Edge points are then localized as minima in the averaged local phase measure under a pre-defined threshold (points near 0°). In our implementation we apply 1-D log–Gabor filters (Kovesi, 1996) because these have no DC component and allow for large scales to be used without spreading into negative frequencies.

Figure 8a and b show results of the two feature extraction schemes for identical frames at different points in a cardiac cycle. Note that measurements have been deliberately ignored in the region of the mitral valve to avoid the local-phase tracker following the valve during heart contraction in Figure 8b. In the case of the gradient-based tracker it was necessary to use these (bad) measurements as they were the only measurements that provided any information about right-hand wall motion.

Frames in Figure 8a show how a gradient-based tracker finds very few features along the septum (right-hand wall) and in regions of low image intensity. In contrast, using local phase measurements, a large number of features are detected along the contour (Figure 8b). These features give a larger measurement set on which the tracker can base a good estimate of the update to the prediction.

4.2. Comparing feature detection schemes

In this section we compare the performance of the tracking algorithm using (i) gradient-based feature detection on the original data (the original visual image tracker); (ii) gradient-based feature detection on TDLMS/IBS-enhanced images; and (iii) local-phase-based feature detection on TDLMS/IBS-enhanced images.

A frame-by-frame visual comparison of tracking using the three approaches shows that the reliability of detection of boundaries was best using method (iii).

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. Figure 5 shows coefficient plots based on tracking on the original data (light curve) and the data pre-filtered by the spatio-temporal boundary enhancement method with gradient-based localization (dark curve). Observe that the plots are more consistent for the algorithm which uses the spatio-temporal filtering approach. Figure 6 compares the state vectors for the two spatio-temporal tracking algorithms which use different localization methods. Observe that the plots for the local-phase localization are more consistent. This shows that the local-phase based localization scheme gives the most stable tracking performance.

Table 3. The results of applying a PCA to four manually segmented cardiac cycles from an abnormal heart. It is clear that seven modes of variation explain over 95% of the variability.

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

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 utilized in the detection step. This approach has been investigated by Herlin and Ayache (1992) although their approach is quite different as they aim to find step edges in raw (displayed) ultrasound images and assume a Gaussian noise model. We are currently exploring methods to extend our approach to spatio-temporal acoustic boundary detection to a truly 3-D (2-D+T) filtering process.

5. TOWARDS CLASSIFICATION

The ultimate goal of this work is to demonstrate that automated image analysis can be used 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 heart image sequences.

5.1. A synthesized 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-hand) wall appeared sluggish. This was achieved by shifting an image block containing the posterior wall from each of the diastole images

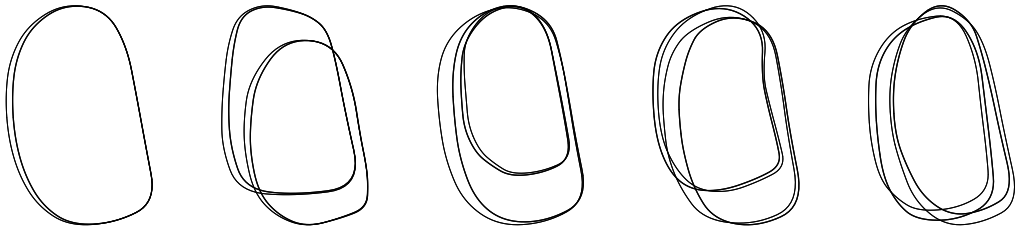


Figure 10. PCA performed on four cardiac cycles of an ultrasonic image sequence. The actual data (light curve) is plotted along with the simulated data (dark curve). The mean shape (left) is plotted. The modes represent the addition of ± 3 sd to the mean shape; from second left mode 1 (the dominant mode) to mode 4.

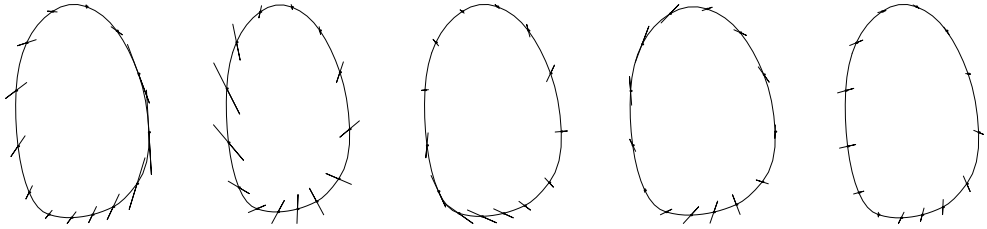


Figure 11. First five principal modes for an abnormal heart. The mean shape (filled line) is plotted along with flow lines representing how the start of each span behaves with the addition of ± 3 sd to the mean. From the left, mode 1 (the dominant mode) to mode 5.

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 10 summarizes the results. The key factor to observe is that the nature of the principal modes remain unchanged although the magnitude is affected (cf. 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 9 shows plots of wall motion from (a) end-systole to end-diastole and (b) end-diastole to end-systole for the original heart data (left) and synthesized abnormal heart motion (right). Wall motion has been computed between consecutive image frames at equally spaced points along the contour and displayed using a colour coding scheme (red = large movement, blue = small movement). The reduced motion in the abnormal heart case is clearly evident. Displays of this kind are potentially useful to clinicians to highlight regional wall motion abnormalities.

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 11 summarizes the results of the PCA. In this case seven modes of variation express 95% of the variability as compared with 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 and the fourth mode a mixture of a translation and scaling. The fifth mode is translation.

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 generalizing 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.

One of the constraints of our current work is that the appearance of the echogram depends on which plane of the ventricle is cut, and our model dynamics cannot currently take this into account. To date, we have also only applied the methods to training and tracking on the same patient. In Section 2 we outlined how the tracking algorithm is trained using a single-step estimation of the system matrices $\mathbf{A}_0$, $\mathbf{A}_1$ and $\mathbf{B}_0$. 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 a new methodology for robust acoustic boundary feature measurement. In Section 4 we found that spatio-temporal noise reduction and local-phase-based localization improved image feature detection. We are considering how to extend our current approach to a truly 3-D (2-D+T) filtering process as well as considering ideas based on anisotropic diffusion (Perona and Malik, 1990; Sanchez-Ortiz *et al.*, 1996).

Finally, the ultimate measure of the success of this work will be to demonstrate that it is possible to relate the tracking parameters to clinically meaningful descriptors of the cardiac performance. To achieve this we are presently working on a regional heart decomposition model to interpret tracking sequences. We plan to evaluate the clinical potential of our algorithms using the objective quantification of ischemic heart disease, stress testing and dobutamine stress echocardiography as example cardiac application domains.

ACKNOWLEDGEMENTS

We are grateful to George Hart and George Sitafidis for useful clinical discussions and for providing clinical data. We thank Mike Isard and members of the Visual Dynamics Group at Oxford for providing the implementation of their tracking algorithm and advice on its use. We would also like to thank the anonymous referees for their comments. GJ and MMP gratefully acknowledge the support of the EPSRC.

REFERENCES

Bardinet, E., Cohen, L. D. and Ayache, N. (1996) Tracking and motion analysis of the left ventricle with deformable superquadrics. *Med. Imag. Anal.*, 1, 129–149.

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

Blake, A., Isard, M. and Reynard, D. (1995) Learning to track the visual motion of contours. *Artific. Intell.*, 78, 179–212.

Chalana, V., Linker, D. T., Haynor, D. R. and Kim, Y. (1996) A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Trans. Med. Imag.*, 15, 290–298.

Chu, C. H., Delp, E. J. and Buda, A. J. (1988) Detecting left ventricular endocardial and epicardial boundaries by digital two-dimensional echocardiography. *IEEE Trans. Med. Imag.*, 7, 81–90.

Cootes, T. F., Taylor, C. J., Cooper, D. H and Graham, J. (1995) Active shape models—their training and application. *Computer Vision Image Understanding*, 61, 38–59.

Declerk, J., Feldmar, J. and Ayache, N. (1998) Definition of a four-dimensional continuous planispheric transformation for the tracking and the analysis of left-ventricle motion. *Med. Imag. Anal.*, 2, 197–213.

Dias, J. M. B. and Leitão, J. M. N. (1996) Wall position and thickness estimation from sequences of echocardiographic images. *IEEE Trans. Med. Imag.*, 15, 25–38.

Evans, A. N. and Nixon, M. S. (1996) Biased motion-adaptive temporal filtering for speckle reduction in echocardiography. *IEEE Trans. Med. Imag.*, 15, 39–50.

Gelb, A. (1974) *Applied Optimal Estimation*. MIT Press, Cambridge, MA.

Han, C. Y., Kwun, N. L., Wee, W. G., Minitz, R. M. and Porembka, D. T. (1991) Knowledge based image analysis for automated boundary enhancement of transesophageal echocardiographic left ventricular images. *IEEE Trans. Med. Imag.*, 10, 602–610.

Herlin, I. and Ayache, N. (1992) Feature extraction and analysis methods for sequences of ultrasound images. In *Proc. ECCV*, pp. 43–57.

Herlin, I., Bereziat, D., Giraudon, G., Nguyen, C. and Graffigne, C. (1994) Segmentation of echocardiographic images with markov random fields. In *Proc. ECCV*, pp. 201–206.

Jacob, G. and Noble, J. A. (1997) Evaluating a robust contour tracker on echocardiographic sequences. In *Proc. MIUA*, pp. 81–84.

Kovesi, P. (1996) *Invariant Measures of Image Features from Phase Information*. *Ph.D. Thesis*, Department of Psychology, University of Western Australia.

Krzanowski, W. J. (1988) *Principles of Multivariate Analysis*. Oxford University Press, Oxford.

Lange, A., Moran, C. M., Palka, P., Fenn, L., Sutherland, G. and McDicken, W. N. (1995) The variation in integrated backscatter in human hearts in differing ultrasonic transthoracic views. *J. Am. Soc. Echocardiogr.*, 8, 830–838.

Mailloux, G. E., Bleau, A., Bertrand, M. and Petitclerc, R. (1987) Computer analysis of heart motion from two-dimensional echocardiograms. *IEEE Trans. Biomed. Eng.*, 34, 356–364.

McEachen, J. C. II and Duncan, J. S. (1997) Shape-based tracking of the left ventricular wall motion. *IEEE Trans. Med. Imag.*, 16, 270–283.

Mulet-Parada, M. (1997) *Feature Extraction in Echocardiographic Images—Towards 3D+T Reconstruction*. PRS Transfer Report, Department of Engineering Science, University of Oxford.

Ofili, E. O. and Nanda, N. C. (1994) Three-dimensional and four-dimensional echocardiography. *Ultrasound Med. Biol.*, 20, 669–675.

Park, J., Metaxas, D. and Young, A. (1996) Analysis of left ventricular wall motion based on deformable models and MRI-SPAMM. *Med. Image Anal.*, 1, 53–71.

Perona, P. and Malik, J. (1990) Scale space and edge detection using anisotropic diffusion. *IEEE Trans. PAMI*, 12, 629–639.

Rohling, R. N, Gee, A. H. and Berman, L. (1997) 3-D spatial compounding of ultrasound images. *Med. Image Anal.*, 1, 177–193.

Salustri, A. and Roelandt, J. R. T. C. (1995) Ultrasonic three-dimensional reconstruction of the heart. *Ultrasound Med. Biol.*, 21, 281–293.

Sanchez-Ortiz, G. I., Rueckert, D. and Burger, P. (1996) Knowledge-based anisotropic diffusion of vector valued 4-D cardiac MR images. In *Proc. British Machine Vision Conf.*, pp. 605–614.

Venkatesh, S. and Owens, R. (1990) On the classification of image features. *Pattern Recogn. Lett.*, 11, 339–349.