Medical Image Tracking

Nicholas Ayache, Isaac Cohen, Isabelle Herlin

17 Medical Image Tracking

Nicholas Ayache, Isaac Cohen and Isabelle Herlin

17.1 The Analysis of Ultrasound Image

There is a continuously increasing demand in the automated analysis of 2D and 3D medical images at the hospital (Ayache *et al.*, 1989; Ayache *et al.*, 1990). Among these images, ultrasound images play a crucial role, because they can be produced at video-rate and therefore allow a dynamic analysis of moving structures. Moreover, the acquisition of these images is non-invasive (it does not hurt the patient!) and the cost of acquisition is relatively low compared to other medical imaging techniques.

On the other hand, the automated analysis of ultrasound images is a real challenge for active vision, because it combines most of the difficult problems encountered in computer vision in addition to some specific ones related to the acquisition mode:

- images are degraded by a very high level of corrupting noise,
- observed objects usually correspond to non-static, non-polyhedral and non-rigid structures,
- images are usually produced in polar geometry instead of cartesian geometry.

We show in the first part of this chapter how a new sonar-space filtering method combined with a snake-based tracker (see also chapters 1, 3) can be used to provide a *global* tracking of some anatomical structures in time sequences of noisy echocardiographic images.

We then show how to introduce additional shape constraints to obtain, if required, a *local* tracking of individual points along the globally tracked moving structure.

Additional details can be found by the interested readers in (Herlin & Ayache, 1992a; Herlin & Ayache, 1992b) for the first part, and in (Cohen *et al.*, 1992a; Cohen *et al.*, 1992b), for the second part. Also, this work should be compared to some previous attempts to provide an automated analysis of ultrasound images, for instance (Taxt *et al.*, 1990; Faure *et al.*, 1988; Buda *et al.*, 1983; Jenkins *et al.*, 1981; Zhang & Geiser, 1982; Unser *et al.*, 1989).

17.2 Ultrasound Image Reconstruction and Filtering

17.2.1 Geometry of Image Acquisition

The fact that ultrasound scan lines are acquired in polar coordinates creates an important anisotropy in spatial resolution. The geometric transformation which transforms the data from a polar representation to the correct cartesian representation is called scan correction.

Let us suppose that M different orientations are used to obtain an echocardiographic image, and that each return signal is digitized to L points. Fig. 17.1 shows an echographic image, with M rows and L columns, obtained using a commercial echographic machine, providing an image represented in polar coordinates. Fig. 17.2 shows the cartesian image corresponding to the same data.

Figure 17.1
Ultrasound image on raw data.

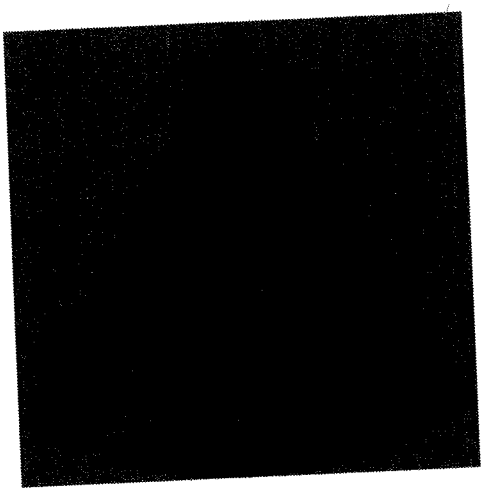

Figure 17.2
Cartesian image after conversion.

This scan conversion requires the knowledge of the following set of parameters (see Fig. 17.3):

- the angular extent of data acquisition wedge α,
- minimal distance d for data acquisition,
- total distance D for data acquisition (these distances being measured from the skin), and
- the number of rows, N, desired in the output cartesian image (The number of columns will be related to α, and we will assume square pixels).

Several methods may be used for the conversion process. Usually, the video image on the echograph is obtained by assigning to a cartesian point the grey level value of the nearest available point in polar coordinates, or the

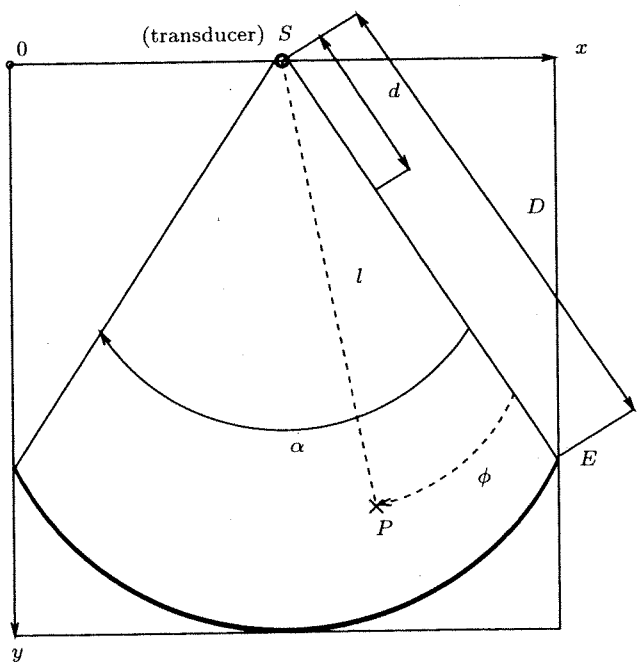

Figure 17.3
Parameters of the conversion process.

value of the bilinear interpolation of its four nearest points. An alternative method consists in using Bresenham's method, well-known in the Computer Graphics field.

In fact we found that these methods do not make optimal use of the available original data, and we introduced, a new method, sonar-space filtering, which can be used to adapt locally the computation of the convolution product in cartesian space with respect to the available resolution of the polar data. In this sense, this provides an optimal reconstruction for a chosen level of spatial filtering.

17.2.2 Sonar-Space Filtering

We assume that it is desired that the continuous input cartesian image $I(x, y)$ be filtered by the impulse response $f(x, y)$. The resulting image $R(x, y)$, in continuous space, is given by the convolution product:

$$R(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-u, y-v) \cdot I(u,v)\, du\, dv. \tag{17.1}$$

However, the input is only available in the polar coordinate space. We thus apply the following change of variables:

$$\begin{cases} \theta = \left[\arctan\left(\dfrac{y}{x-x_0}\right) - \left(\dfrac{\pi-\alpha}{2}\right)\right] / \Delta_\alpha \\ \rho = \left[\sqrt{(x-x_0)^2+y^2}\right] e - \Delta_N \end{cases}$$

where

- $\Delta_N = d\frac{(L-1)}{(D-d)}$ is the distance (from the surface of the skin) where the acquisition process begins, measured in pixel units along a scan line of the raw data,
- $\Delta_\alpha = \frac{\alpha}{(M-1)}$ is the angular difference between two successive angular positions of the probe,
- $e = \frac{D}{(D-d)}\frac{(L-1)}{(N-1)}$ adjusts the pixel sampling rates along the axial direction of the beam, according to the desired height N of the cartesian image.

This yields the following:

$$R(x,y) = \int_0^{2\pi}\int_0^{\infty} f\big(x-u(\rho,\theta), y-v(\rho,\theta)\big)\cdot I(\rho,\theta)\cdot |\mathbf{J}(\rho,\theta)|\, d\rho\, d\theta. \quad (17.2)$$

Here $|\mathbf{J}(\rho,\theta)|$ is the determinant of the Jacobian matrix corresponding to the inverse transformation of variables:

$$\begin{cases} x &= \dfrac{\rho+\Delta_N}{e}\cos\left(\theta\Delta_\alpha + \dfrac{\pi-\alpha}{2}\right) + x_0 \\ y &= \dfrac{\rho+\Delta_N}{e}\sin\left(\theta\Delta_\alpha + \dfrac{\pi-\alpha}{2}\right). \end{cases}$$

It is easily seen that

$$|\mathbf{J}(\rho,\theta)| = \frac{(\rho+\Delta_N)*\Delta_\alpha}{e^2}.$$

We have transformed the convolution in the cartesian coordinates to an infinite integral in polar coordinates, corresponding to the domain of the raw data. But the computer implementation requires a finite discrete summation.

Once the two-dimensional convolution filter f is chosen, we define its rectangle of essential support: say a rectangular window of width $2X$ and $2Y$. Outside this region of support the absolute value of the impulse response must be lower than a pre-selected threshold s, i.e.,:

$|f(u,v)| < s$ if $((|u| \geq X)$ OR $(|v| \geq Y))$.

Therefore the integral is approximated by a finite integral over the domain

$(x-X \leq u \leq x+X)$ AND $(y-Y \leq v \leq y+Y)$.

The filter is also sampled in this domain in order to approximate the integral by a discrete summation. Filtered numerical outputs are evaluated at original data point locations within the continuous domain.

We thus obtain the following equation:

$$R(x,y) = C\sum_k f\big(x - u(\rho_k,\theta_k), y - v(\rho_k,\theta_k)\big)\cdot I(\rho_k,\theta_k)\cdot|\mathbf{J}(\rho_k,\theta_k)|\,, \quad (17.3)$$

where the summation is over the discrete collection of point (ρ_k, θ_k) in polar coordinates and where C is used to normalize the data, and

$$|\mathbf{J}(\rho_k,\theta_k)| = \frac{(\rho_k + \Delta_N) * \Delta_\alpha}{e^2}.$$

We have thus transformed the computation of $R(x,y)$ into a discrete convolution on a window of size $2X$ by $2Y$, making use of image data only at points where it is defined in the polar coordinate domain. Note that for the raw data, or more generally for any sonar-like data, the sampling of the filter is not regular along the x and y axes but it rather conforms to the sampling density of raw data. In this formula, the $|\mathbf{J}(\rho_k,\theta_k)|$ value represents the surface area of the polar pixel patch in the cartesian domain. We will present in the following subsections the different values that we have chosen for the function f for different processing of the echocardiographic raw data.

17.2.3 Image Visualization

In practice, the convolution filter $f(x,y)$ is typically separable and is denoted by $f(x)g(y)$. Then, classical 1-D smoothing and derivation filters can be used.

For visualization applications of sonar-space filtering, we used the Deriche's smoothing function (Deriche, 1987), $f(x) = g(x) = L(x)$:

$$L(x) = k_2\big(\alpha\sin(\omega|x|) + \omega\cos(\omega|x|)\big)^{-\alpha|x|} \quad (17.4)$$

The conversion algorithm performs simultaneously a conversion to cartesian coordinates and a smoothing of the data (whose amplitude can be adjusted with α) thus producing a cartesian image with a reduced speckle. Other smoothing functions could be used instead.

17.2.4 Edge Detection

For automatic boundary tracking, we investigated the use of spatio-temporal approaches (Monga & Deriche, 1988). A time-varying edge may be represented as a surface in 3-D space, in which x and y are two spatial dimensions and t is the temporal dimension. We modify Deriche's edge detector for this goal.

We denote by α_x, α_y and α_t the filtering parameters of the Deriche filter (cf. Eq. 17.4) for the x, y and t axes respectively.

Since the 2D space is homogeneous, we can choose $\alpha_x = \alpha_y$. On the other hand, the value of α_t is independent and must be chosen according to the temporal resolution. (Another approach could be to generalize Deriche's detector with spatio-temporal functions as in (Hwang & Clark, 1990).)

We denote by G_x and G_y the two spatial components of the gradient vector and $I(x, y, t)$ the 3-D image function. Let D be the Deriche differentiation filter and L the associated smoothing filter. The two components of the gradient vector have the following expression:

- $G_x = (D_x L_y L_t) \otimes I(x, y, t)$
- $G_y = (L_x D_y L_t) \otimes I(x, y, t)$

where the subscripts are used to indicate along which axis the corresponding filter is applied and $\otimes$ denotes the convolution product.

Each component is obtained by differentiation in the associated direction and filtering in the other spatial direction and in the temporal direction. The squared norm of the gradient is defined by: $N^2(x, y, t) = G_x^2 + G_y^2$.

Finally, edges are obtained as local maxima of the gradient norm in the direction of the 2D gradient vector. The temporal dimension is only used to smooth the 3D intensity function. This produces a significant image enhancement in regions which are not moving too fast.

17.3 Global Temporal Tracking

17.3.1 Crude Initialization of Structure Edges

To obtain a crude estimation of the boundaries of anatomical structures, we use techniques from mathematical morphology. The model of a cardiac cavity is very simple. This is an ovoid region with low intensity. These regions cannot be obtained by simple thresholding because of the speckle noise. But the fine structures of the speckle may be easily suppressed by the following morphological operations (Serra, 1982):

- A first order opening eliminates the small bright structures on dark background.
- The dual operation (first order closing) suppresses the small dark structures.

After these operations, a simple thresholding gives an image C where all the cardiac cavities are represented in white. This detection can be refined by the use of higher level information. The specialist points out, using a computer mouse, the chosen cavity on the image. The whole cavity is then obtained by a conditional dilatation which begins at this point.

17.3.2 Use of an Active Deformable Model

The previous operations usually provide a gross approximation of the structure boundaries. In order to improve this crude segmentation to an accurate determination of the boundaries, we use the deformable models of (Cohen & Cohen, 1990), in the spirit of (Kass *et al.*, 1987a) (see also chapters 1 and 3).

The deformable model is initialized in the first image by the gross approximation of the structure boundaries, and evolves under the action of image forces which are counter-balanced by its own internal forces, which preserve its regularity.

Image forces are computed as the derivative of an attraction potential related to the previously computed spatio-temporal edges. Typically, the potential is inversely proportional to the distance of the nearest edge point.

Once the model has converged in the first frame, its final position is used as the initial one in the next frame, and the process is repeated. This is equivalent to a zero-mass tracking snake.

17.4 Global Tracking Results

Fig. 17.4 is a manual sketch of the image which shows where cardiac cavity is located, and can be considered the "ground-truth" of the echographic data in Fig. 17.2.

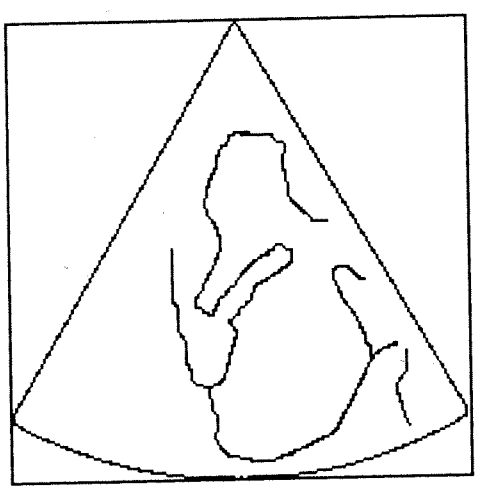

Figure 17.4
Manual sketch of the image.

Cartesian reconstruction of the polar image of Fig. 17.2 is shown in Fig. 17.5. The image has been reconstructed using direct convolution with the low-pass filter of equation 17.4 with $\alpha = 1$. Then, edges have been

extracted without temporal smoothing by applying a spatial Deriche Filter (Deriche, 1987) to the cartesian image, with a resolution parameter $\alpha = 1$.

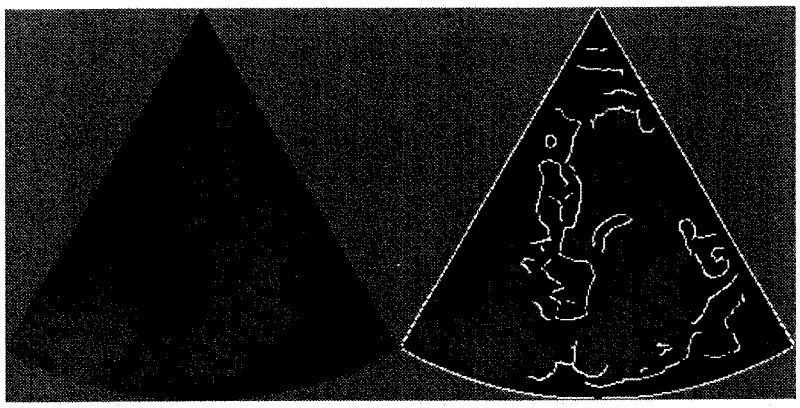

Figure 17.5
Reconstructed cartesian image and edges obtained by sonar-space filtering.

The effect of temporal smoothing can be seen on figure 17.6. There is a reduction of some local distortions on the deeper edges of the left auricle, as can been seen when comparing the bottom right cavity of figure 17.4 with figure 17.6. Simultaneously, temporal smoothing can make the mitral valve vanish (middle thin structure of the same figures), which is due to the fast motion of this structure with respect to the temporal resolution.

The chosen strategy is thus to use temporal smoothing only to study cavities (slow motion) and not when studying fast moving structures like the valves.

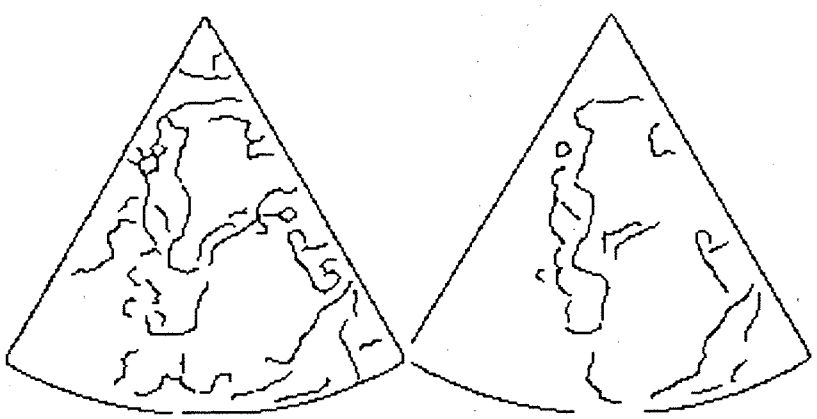

Figure 17.6
Left: no temporal smoothing of edges, right: temporal smoothing of edges.

Our methodology was applied to four different sequences obtained from

two different echographs. We only present here the results obtained from a time sequence acquired in polar coordinates from a VIGMED echograph at the Hospital Henri Mondor in Creteil, France. This sequence contains 38 images from a cardiac cycle.

Fig. 17.7 shows a cartesian representation of the original data (only one image in four is displayed). Edges are shown in Fig. 17.8, and the temporal tracking of the mitral valve and the auricle are presented in Figs. 17.9 and 17.10.

17.5 Tracking Points on the Deformable Structure

The new objective is now to perform locally the tracking of each individual point on the active deformable model. The approach we developed in (Cohen *et al.*, 1992a) was an improvement of the work of (Amini *et al.*, 1991; Duncan *et al.*, 1991) to analyze the deformations of curves in sequences of 2D images.

17.5.1 Privileging High Curvature Points

This approach is based on the paradigm that high curvature points usually possess an anatomical meaning, and are therefore good landmarks to guide the matching process. This is the case for instance when deforming patients skulls (Cutting, 1989; Guéziec & Ayache, 1991; Monga *et al.*, 1992), or when matching patient faces taken at different ages, when matching multipatients faces, or when analyzing images of a beating heart. In these cases, many lines of extremal curvatures (or ridges) are stable features which can be reliably tracked between the images (on a face they will correspond to the nose, chin and eyebrows ridges for instance, on a skull to the orbital, sphenoid, falx, and temporal ridges, on a heart ventricle to the papillary muscle etc. ...).

As Duncan's team (Amini *et al.*, 1991; Duncan *et al.*, 1991), we therefore propose a method based on the minimization of an energy which tends to preserve the matching of high curvature points, while ensuring a smooth field of displacement vectors everywhere. The energy minimization is obtained through the mathematical framework of Conforming Finite Elements analysis (Ciarlet, 1987), which provides a rigorous and efficient numerical solution. Moreover we show in (Cohen *et al.*, 1992a) that the approach can be generalized in 3D to analyze the deformations of surfaces.

This approach is particularly attractive in the absence of a reliable physical or deformable geometric model of the observed structures, which is often the case when studying medical images. When such a model is available, other approaches would involve a parametrization of the observed shapes (Metaxas & Terzopoulos, 1991a) (see chapter 5), a modal analysis of the displacement field (Horowitz & Pentland, 1991), or a parametrization of a subset of deformations (Bookstein, 1989; Mishra *et al.*, 1991). In fact we believe that our approach can always be used when some sparse geometric features provide reliable landmarks, either as a preprocessing to provide an

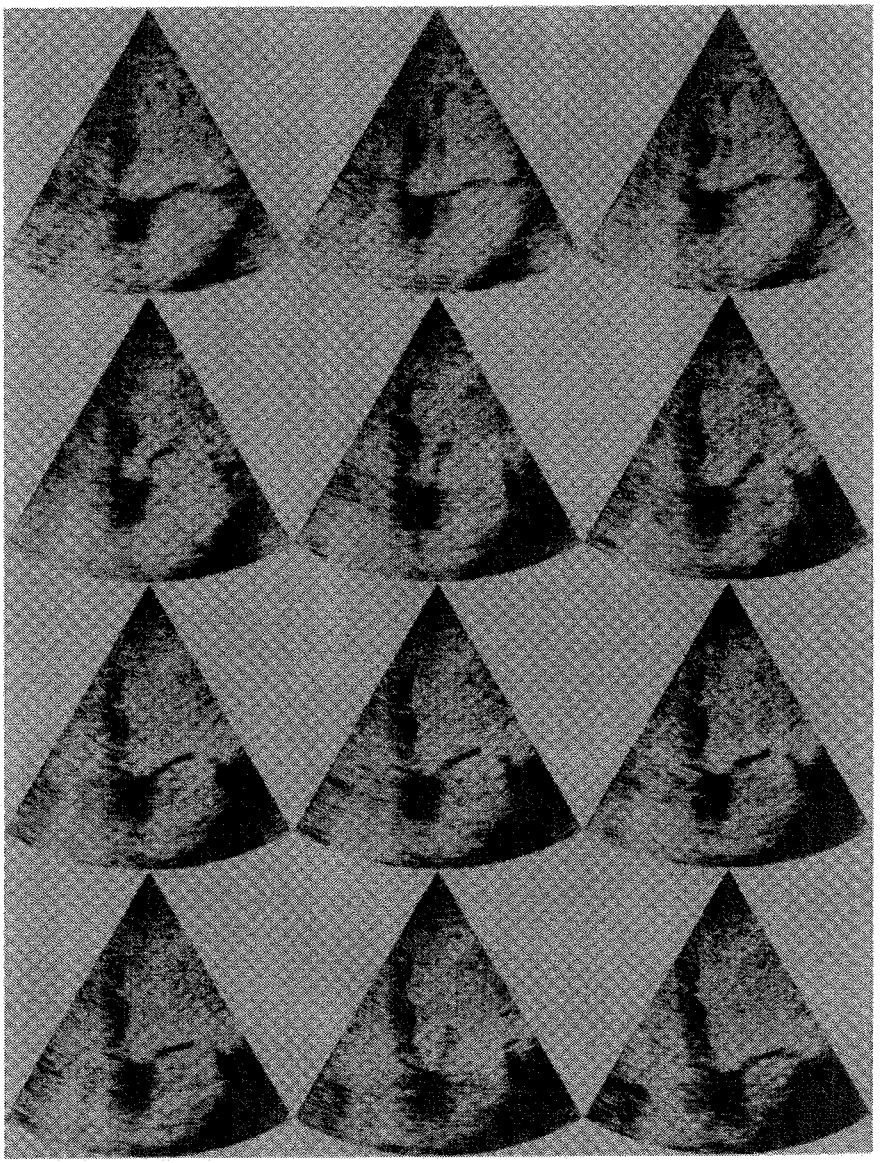

Figure 17.7
Data after scan-correction.

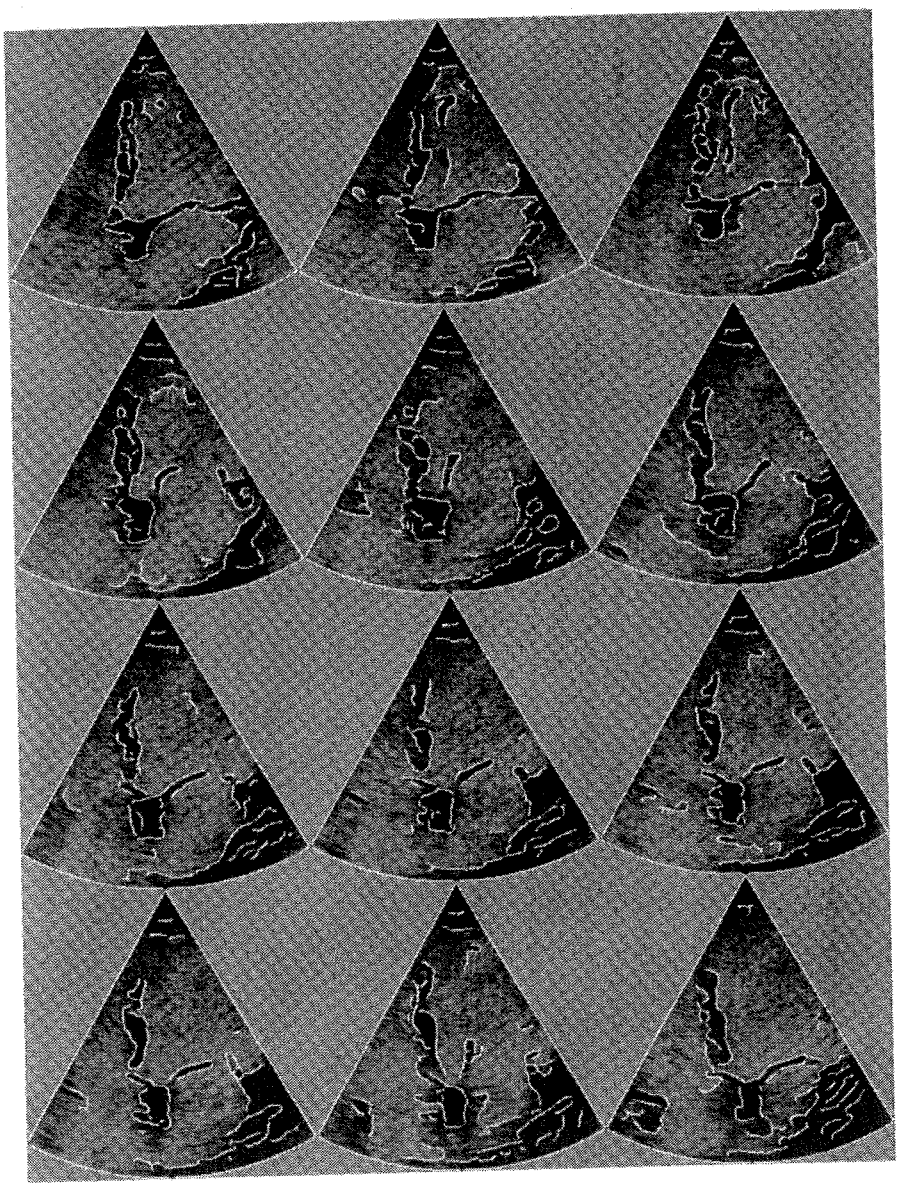

Figure 17.8
Edges.

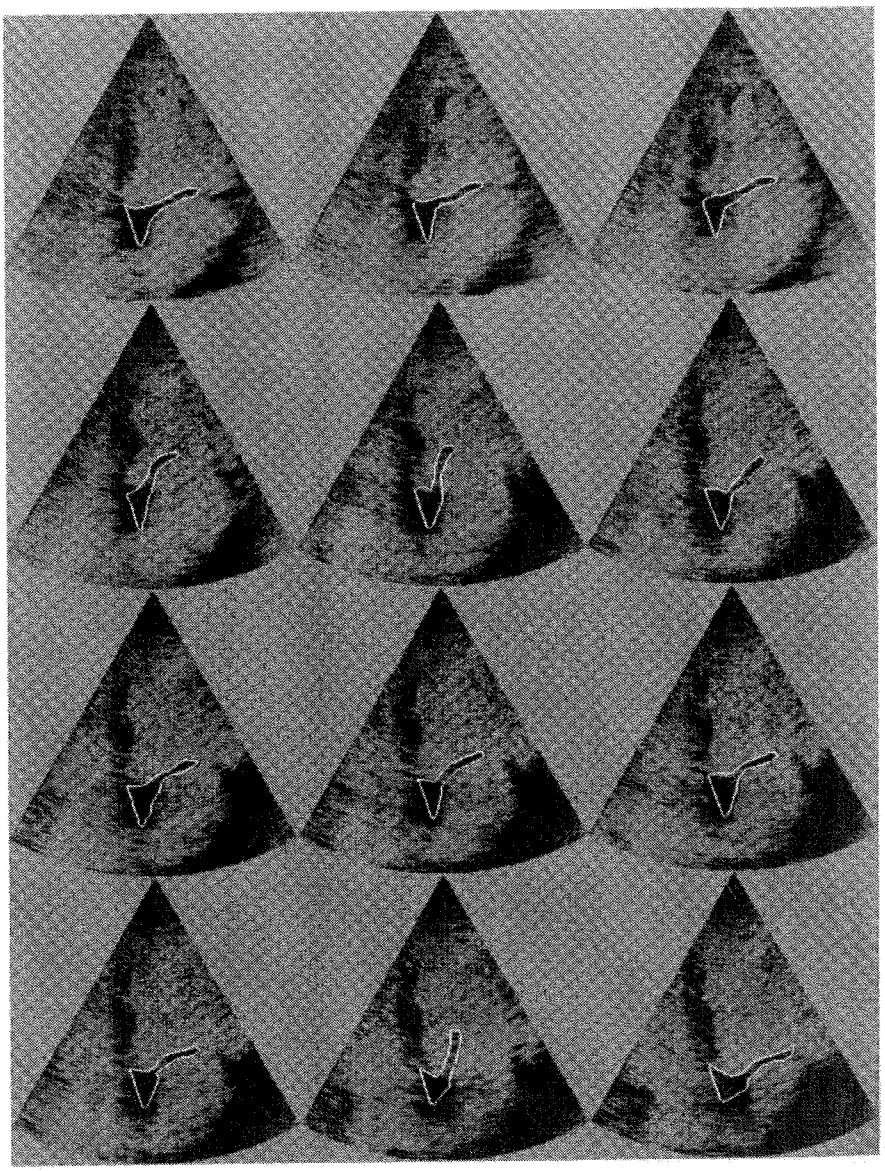

Figure 17.9
Temporal tracking of the mitral valve.

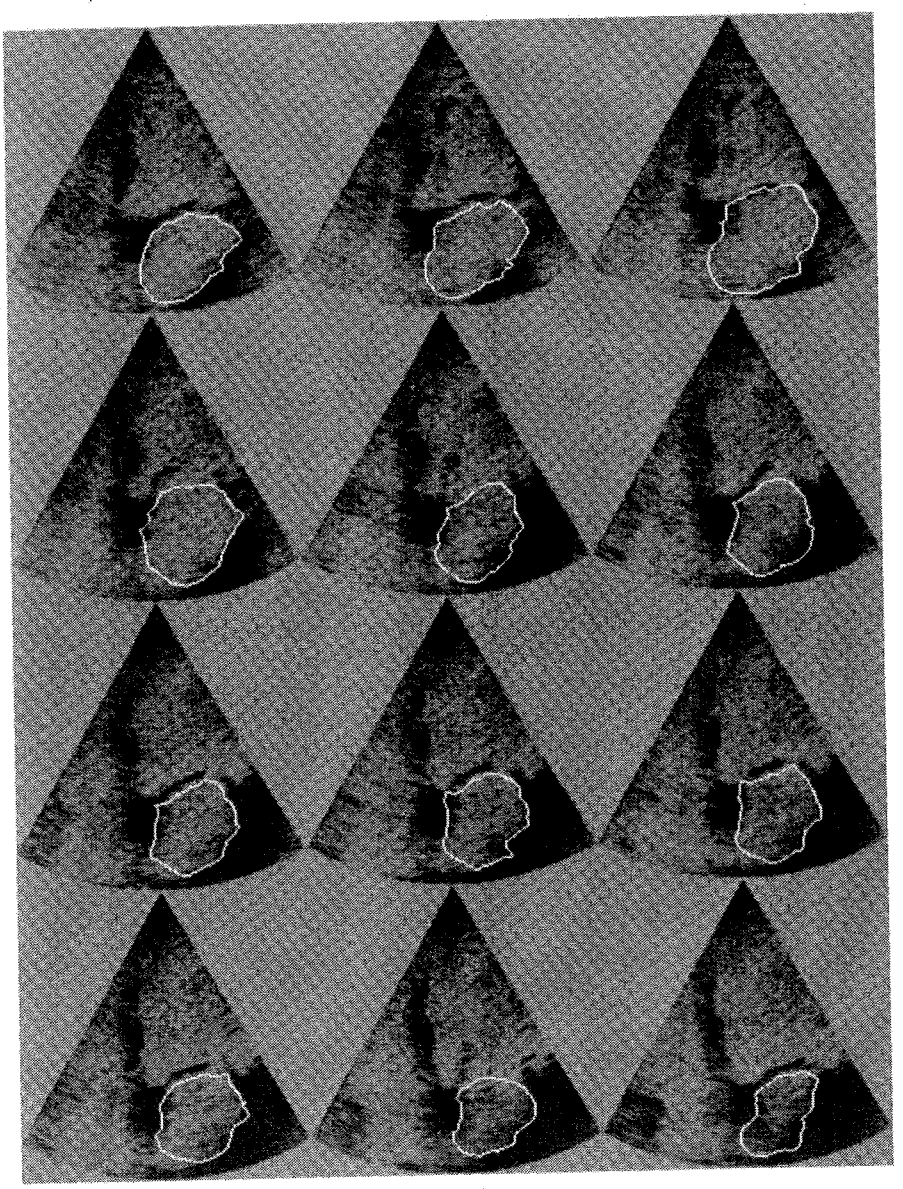

Figure 17.10
Temporal tracking of the left auricle.

initial solution to the other approaches, or as a post-processing to provide a final smoothing which preserves the matching of reliable landmarks.

17.5.2 Modelling the Problem

Let C_P and C_Q be two boundaries of the image sequence, the contour C_Q is obtained by a non rigid (or elastic) deformation of the contour C_P. The curves C_P and C_Q are parameterized by $P(s)$ and $Q(s')$ respectively.

The problem is to determine for each point P on C_P a corresponding point Q on C_Q. For doing this, we must define a similarity measure which will compare locally the neighborhoods of P and Q.

As explained in the previous section, we assume that points of high curvature correspond to stable salient regions, and are therefore good landmarks to guide the matching of the curves. Moreover, we can assume as a first order approximation, that the curvature itself remains invariant in these regions. Therefore, we can introduce an energy measure in these regions of the form:

$$E_{curve} = \frac{1}{2}\int_{\delta S} \left(K_Q(s') - K_P(s)\right)^2 ds \tag{17.5}$$

where K_P and K_Q denote the curvatures and s, s' parameterize the curves C_P and C_Q respectively. In fact, as shown by (Duncan *et al.*, 1991; Landau & Lifshitz, 1986), this is proportional to the energy of deformation of an isotropic elastic planar curve.

We also wish the displacement field to vary smoothly around the curve, in particular to insure a correspondence for points lying between two salient regions. Consequently we consider the following functional (similar to the one used by Hildreth to smooth a vector flow field along a contour (Hildreth, 1984)):

$$E = \int_{C_P} \left(K_Q(s') - K_P(s)\right)^2 ds + R\int_{C_P} \left\| \frac{\partial(Q(s') - P(s))}{\partial s} \right\|^2 ds \tag{17.6}$$

where

$$E_{regular} = \int_{C_P} \left\| \frac{\partial(Q(s') - P(s))}{\partial s} \right\|^2 ds$$

measures the variation of the displacement vector $\mathbf{PQ}$ along the curve C_P, and the $\|.\|$ denotes the norm associated to the euclidean scalar product $\langle .,. \rangle$ in the space $\mathcal{R}^2$.

The weighting parameter $R(s)$ depends on the shape of the curve C_P. Typically, $R(s)$ is inversely proportional to the curvature at P, to give a larger weight to E_{curve} in salient regions and conversely to $E_{regular}$ to points inbetween. This is done continuously without annihiling totally the weight of any of these two energies (see (Cohen *et al.*, 1992a)).

17.5.3 Solving the Matching Problem

Given two curves C_P and C_Q parameterized by $s \in [0,1]$ and $s' \in [0,\alpha]$ (where α is the length of the curve C_Q), we have to determine a function f such that:

$$\begin{array}{rcl} f : [0,1] & \rightarrow & [0,\alpha] \\ s & \rightarrow & s' \end{array} .$$

satisfying

$$f(0) = 0 \text{ and } f(1) = \alpha \tag{17.7}$$

and

$$f = \arg\min\big(E(f)\big) \tag{17.8}$$

where

$$E(f) = \int_{C_P} \big(K_Q(f(s)) - K_P(s)\big)^2 ds + R(s) \int_{C_P} \left\| \frac{\partial\big(Q(f(s)) - P(s)\big)}{\partial s} \right\|^2 ds. \tag{17.9}$$

The condition (17.7) means that the displacement vector is known for one point of the curve. In the trivial case, where the curves C_P and C_Q are identical, solving (17.8) leads to the identity function.

In the model defined above we assumed that:

- the boundaries have already been extracted,
- the curvatures K are known on the pair of contours.

These necessary data are obtained by preprocessing the image sequence (see (Guéziec & Ayache, 1991) for instance).

17.5.4 Numerical Solution

The characterization of a function f satisfying $f = \arg\min E(f)$ and the condition (17.7) is performed by a variational method. This method characterizes a local minimum f of the functional $E(f)$ as the solution of the Euler-Lagrange equation $\nabla E(f) = 0$.

Then solving the Euler-Lagrange equation $\nabla E(f) = 0$ leads to the solution of the partial differential equation:

$$\begin{cases} f'' \left\|\mathbf{Q}'(\mathbf{f})\right\|^2 + K_P \left\langle \mathbf{N}_P, \mathbf{Q}'(f) \right\rangle + \frac{1}{R} \left[K_P - K_Q(f)\right] K'_Q(f) = 0 \\ + \text{Boundary conditions} \end{cases} \tag{17.10}$$

where Q is a parametrization of the curve C_Q, $\mathbf{Q}'(\mathbf{f})$ the tangent vector of C_Q, K'_Q the derivative of the curvature of the curve C_Q and $\mathbf{N}_P$ is the normal vector to the curve C_P. The boundary conditions of equation (17.10) are $f(0) = 0$ and $f(1) = \alpha$ (condition (17.7)) that any solution must satisfy. In the following we consider null boundary conditions (this is done by a simple change of variables).

The equation (17.10) is solved by a finite element method. Details on the associated variational problem and the construction of the approximation space can be found in (Cohen *et al.*, 1992a) and the corresponding internal INRIA research report.

17.6 Local Tracking Results

The method was tested on a set of synthetic and real image sequences. Here, we only present the results obtained on the previously shown ultrasound images of a cardiac valve. This tracking can help the diagnosis of some heart diseases

Figure 17.11 show the tracking of each point of the valve for a selection of 8 successive pairs of contours (for other experiments, please refer to (Cohen *et al.*, 1992b). The results are given by specifying at each discretization point $P_i \quad i = 1 \ldots N$ of the curve C_P the displacement vector $\mathbf{u}_i = \mathbf{P}_i\mathbf{Q}_i$. At each point P_i the arrow represents the displacement vector $\mathbf{u}_i$.

These results were satisfying for the all sequence of 30 images, in the sense that they met quite well the objectives of preserving the matching of high curvature points while insuring a smooth displacement field.

The method was entirely automatic, except for 4 frames (out of 30) where the user had to provide an initial estimation. This was due to excessive deformations with respect to the current temporal resolution. Anyhow, one can visualize that the results meet the objectives of preserving the matching of high curvature points while insuring a smooth displacement field.

17.7 Conclusion

We presented in this chapter an application of active vision to the tracking of points in anatomical structures in time sequences of ultrasound medical images. We presented a chain of processes, built with the following generic tools:

- sonar space filtering, to extract local edgels,
- active deformable models, to track a global structure,
- local point trackers, to follow the motion of each individual point on the global structure.

We illustrated the applicability of the whole chain by demonstrating the tracking of each point of a cardiac valve in a time sequence of ultrasound images.

Further work will go in two directions: first we shall generalize this work in 3 dimensions, following the work of (Cohen *et al.*, 1992c) and (Amini *et al.*, 1991), and second, we shall incorporate some physical model in the local tracking.

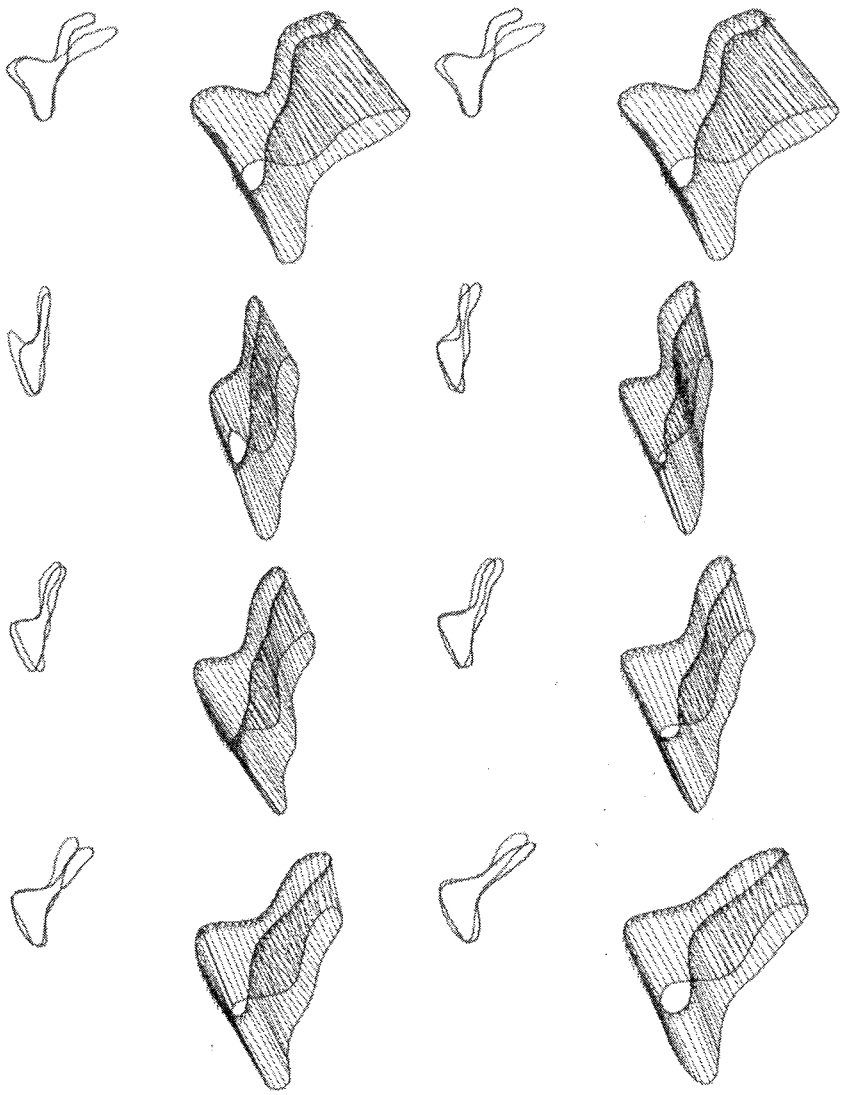

Figure 17.11
Experiments on a valve sequence (time 9–17).