Chapter 8. Experimental Validation with 3D Ultrasound

This chapter describes a series of *in vitro* and *in vivo* experiments validating the methods of this dissertation using RT3D ultrasound data. The first experiment identifies the cardiac left ventricle *in vivo* using a medial node model. Next, the effectiveness of core atoms for finding volumes is demonstrated on ultrasound images of fluid-filled balloons of known volume. The manual tracing technique for the LV is then described, which is used as a gold standard for the remaining two experiments. Core atoms are then applied to the task of determining LV volume, with rather poor results. Finally, the itereative Bayesian method developed in the previous chapter is applied to fuzzy segmentation of the LV and the errors analyzed.

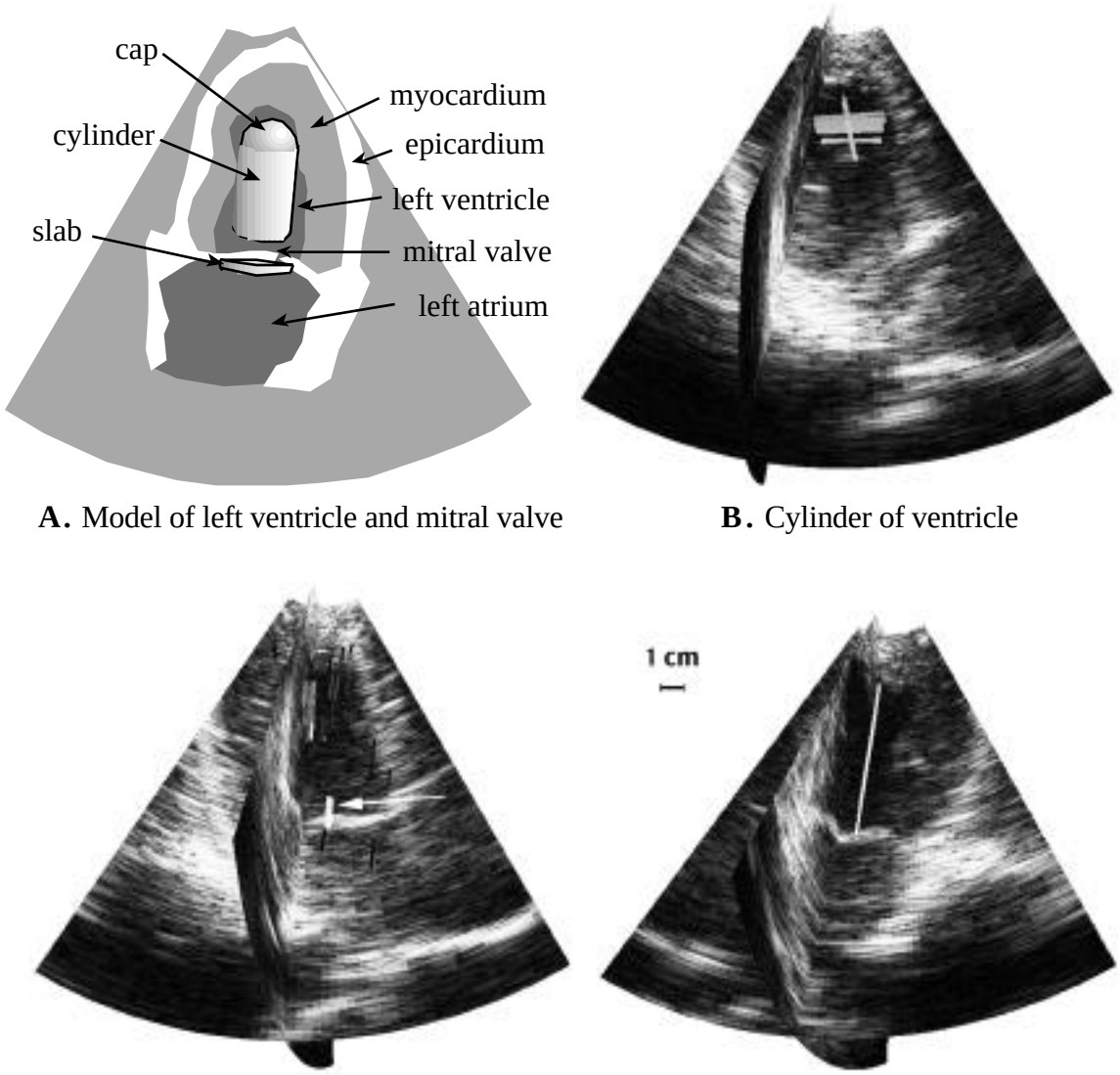
8A. Identifying the AMV Axis

This section describes experimental validation of the methods developed in Chapters 3-6 for object identification using a medial node model (Fig. 8.1A) and demonstrates the ability to automatically identify the Apex-to-Mitral-Valve (AMV) axis of the LV. Only apical ultrasound scans were used, with a further constraint that only time frames with the mitral valve closed were considered. Fortunately, this includes the entirety of systole and could be used, for example, to calculate stroke volume.

No pre-processing of the data was performed. Boundariness was found using a Difference of Gaussian measurement of intensity gradient, with Gaussian application accomplished by repeated convolution with a $2 \times 2 \times 2$ binomial kernel. Further constraints were applied as to the absolute intensity of the candidate boundary points.

To identify the cylinder in the image data, boundary points were determined with 4 applications of the binomial kernel. Core atoms with diameters 0.8 to 4.6 cm and face-to-faceness greater than 0.88 (see Eq. 3.2) were collected in bins on a regular lattice and ellipsoidal voting was applied. An example of the resulting clusters is displayed in Fig. 8.1B. Crosses are shown in the cylindrical chamber of the ventricle. Due to the pre-selection of core atoms by scale, no other significant clusters of core atoms were found. A single intensity constraint could not be found to reliably identify the endocardial boundary, because the intensity varied between images extending into the range of the outer boundary between the

myocardium and surrounding connective tissue (epicardium or septum). However, since this outer boundary also formed a cylinder roughly concentric to that of the endocardial boundary, core atoms forming from the outer boundary established approximately the same axis for the LV.



A. Model of left ventricle and mitral valve

B. Cylinder of ventricle

C. Slab of mitral valve

D. Automated apical-to-mitral-valve (AMV) axis

Fig. 8.1 Using a statistical model of medial primitives to automatically identify the axis of the cardiac left ventricle in Real Time 3D ultrasound data. A scale of 1 cm is shown in **D**.

Next, the mitral valve (MV) was sought. Boundary points were determined as above, but with only 2 applications of the binomial kernel to accommodate the finer structure of the MV. Core atoms with diameters 0.0 to 0.8 cm and face-to-faceness greater than 0.55 were collected. The lower threshold for face-to-faceness was necessary because of the smaller size of the core

atoms. As shown in Fig. 8.1C, the densest clusters formed at the center of the MV, although weaker false targets were also detected in the myocardium off to the side of the ventricle. To eliminate these false targets, the following criterion was established for the formation of appropriate pairs of clusters. Only slab-like clusters located further from the transducer than cylindrical clusters were permitted to form pairs. These pairs voted for their constituent clusters weighted by the product of the cosines between the orientations of each of the two clusters and the vector between them, in the spirit of *face-to-faceness* as (developed in Section 3A). The voting permitted a mean MV location and a mean LV cylinder location to be computed for each scan. The vector between these two mean locations established a cone for expected unpaired boundary points at the apex of the LV and the mean distance to such apical *core ions* was used to determine the location, along that vector, of the apical cap. Thus an apex-to-mitral valve (AMV) axis was determined, as shown in Fig. 8.1D.

The entire procedure that produced Figs. 8.1B, C, and D was automatic and required approximately 10 seconds per 3D scan on a 400 MHz Pentium computer (each scan holds approximately 2 million 8-bit voxels).

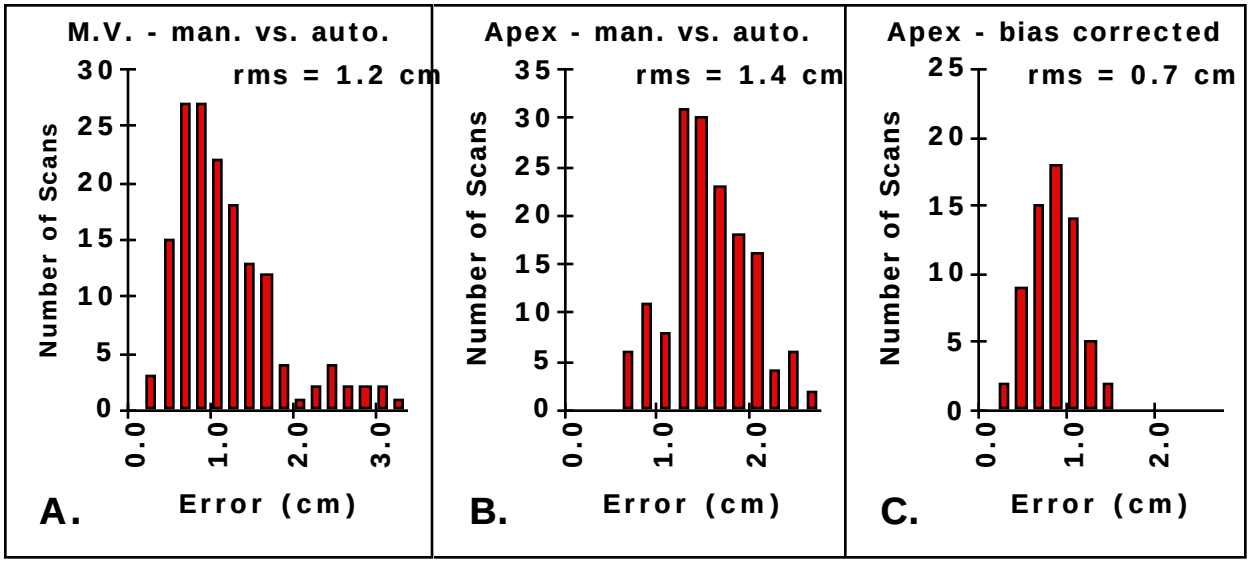


Fig. 8.2 **A.** Error (cm) between manual and automated placement of mitral valve (MV) end-point of apical-mitral-valve (AMV) axis for all 155 3D ultrasound scans. **B.** Error between manual and automated placement of apical end-point of AMV axis, for all 155 scans. **C.** Error for the apical end-point for a random subset of 65 scans corrected for bias measured in the remaining 90 scans, reducing the RMS error from 1.4 to 0.7 cm. (See Fig 15D for scale of 1 cm.)

The anatomical end-points of the AMV axis (the ventricular apex and the center of the MV) were also determined manually, as follows. A human operator was instructed to follow the general cylindrical shape of the left ventricle, marking the ventricular apex and the center of the MV on B-mode and C-mode slices.

The data included 18 RT3D scan sequences of *in vivo* human hearts, using a Volumetrics Model-1 scanner operating at 2.5 or 3.5 MHz. Of the 18 sequences, 12 were described as normal, 4 as dilated cardiomyopathy, 1 as akinetic, and 1 as pericardial effusion. In each sequence, only scans in which the MV was closed were used, for a total of 155 scans. All parameters for the method were established during its development using several normal sequences, none of which were included in these 18 sequences.

The locations of the manual end-points were compared to those determined automatically for all 155 scans, as shown in Figs. 8.2A and 8.2B. The error reported is simply the total physical distance (cm) in 3 dimensions between the manual and automated end-points. For reference, a scale of 1 cm is marked on the cardiac scan in Fig. 8.1D, with the pyramid of a typical scan having a height of approximately 15 cm. As can be seen in Fig. 8.2A, for most scans the center of the MV was correctly located within 2 cm (RMS error 1.2 cm). The greatest error was slightly more than 3 cm. For the apex of the left ventricle, the greatest error was approximately 2.5 cm (RMS error 1.4 cm).

These errors were improved by eliminating a consistent bias between manual and automated measurements. The scan sequences were divided blindly into two groups (training and experimental) without regard to image quality, distributing normal and abnormal hearts evenly, and placing the akinetic and pericardial effusion scans in the experimental group. Error for the LV apex is displayed in Fig. 8.2C, with only the experimental group (65 scans) shown, corrected for bias observed in the training group (90 scans). Correcting for bias yielded an RMS error of 0.7 cm ($x = 0.37$ cm, $y = 0.47$ cm, $z = 0.41$ cm) for placement of the LV apex, and 1.1 cm ($x = 0.48$ cm, $y = 0.62$ cm, $z = 0.72$ cm) for the MV. The automated placement of the LV apex tended to be further into the ventricular chamber (average of 1.2 cm further away from the transducer) than the manual measurement. This bias may have been due to greater local curvature at the LV apex than would have been expected from the simple model of an untapered cylinder with a hemispherical cap, since a typical LV actually narrows considerably as it approaches the apex. Another possible explanation for the bias is that the axis, as determined by the LV and MV clusters, is not centered perfectly as it extends into the apex, leading to contact along the wall rather than at the apex. Other sources of error include sampling in the boundary detection and ambiguities in the correct location of the landmarks by manual placement.

8B. Measuring Balloons with Core Atoms

This section describes the empirical validation of a method of determining volume using core atoms directly on fluid-filled balloons in a bath, in the same manner as mean core atom length was determined on parametric objects in Chapter 4 (see Table 4.2). The method worked quite well on the balloons, which represent uncluttered targets with sharp boundaries. The technique was less successful when applied to the left ventricle and was replaced with the Fuzzy Fill method for that application, as will be described in Sections 8C and 8D.

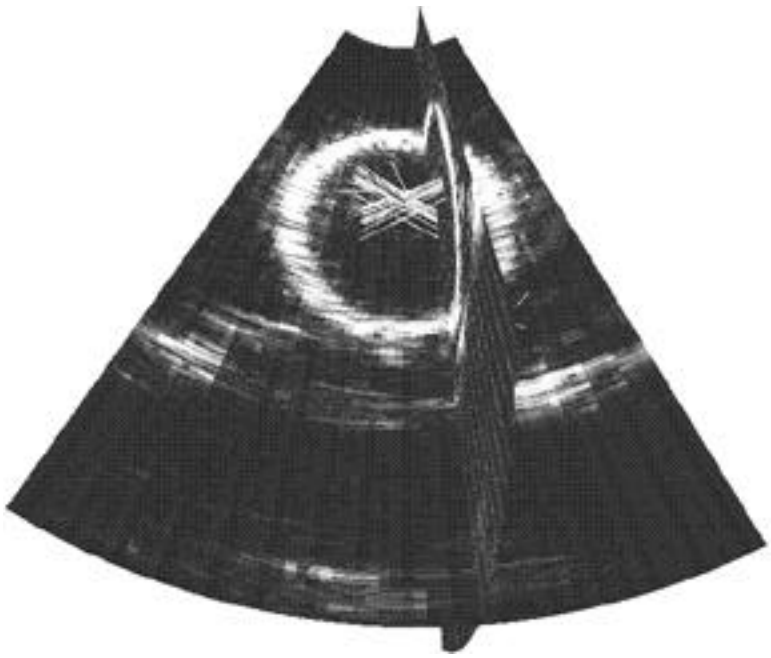


Fig. 8.3. Core atom clusters in a balloon (from the inside boundary of the intensity ridge) identified as cylindrical along the axis of the balloon, although a significant number of core atoms actually formed vertically through the balloon as well.

The ability to accurately measure volume of real objects in 3D ultrasound data using core atoms was demonstrated on a series of 7 balloons filled with an ethanol-water mixture of known density. The volume of the balloons was determined by weight, ranging from 58.0 ml to 83.1 ml. The balloons were scanned in a bath of the same mixture using the prototype RT3D ultrasound scanner developed at Duke University known as "T4". Boundariness was found using a Difference of Gaussian, with 6 applications of a binomial kernel (see Appendix B). Core atoms with lengths ranging from 2.3 to 7.8 cm and face-to-faceness of greater than 0.88 were collected from each scan. A center of mass was computed for all core atoms in each scan. The majority of core atoms formed a roughly spherical Koosh ball configuration as

shown for the sphere in Fig. 3.8. An example of clusters from one balloon is shown in Figs. 6.4 and 6.5 and visualized in 3D in Figs. 6.6 and 8.3. Volume measurements were performed simply by selecting core atoms within 1 cm of the center of mass of all core atoms in the image, since the image was uncluttered by other targets. The mean length of these core atoms (divided by 2) yielded an effective radius r from which a volume was calculated using $v = (4/3)\pi r^3$. This automatically determined volume was compared to balloon volume, determined by weight. This method of volume calculation is distinct from, and simpler than, the fuzzy segmentation technique applied to the LV in Section 8E.

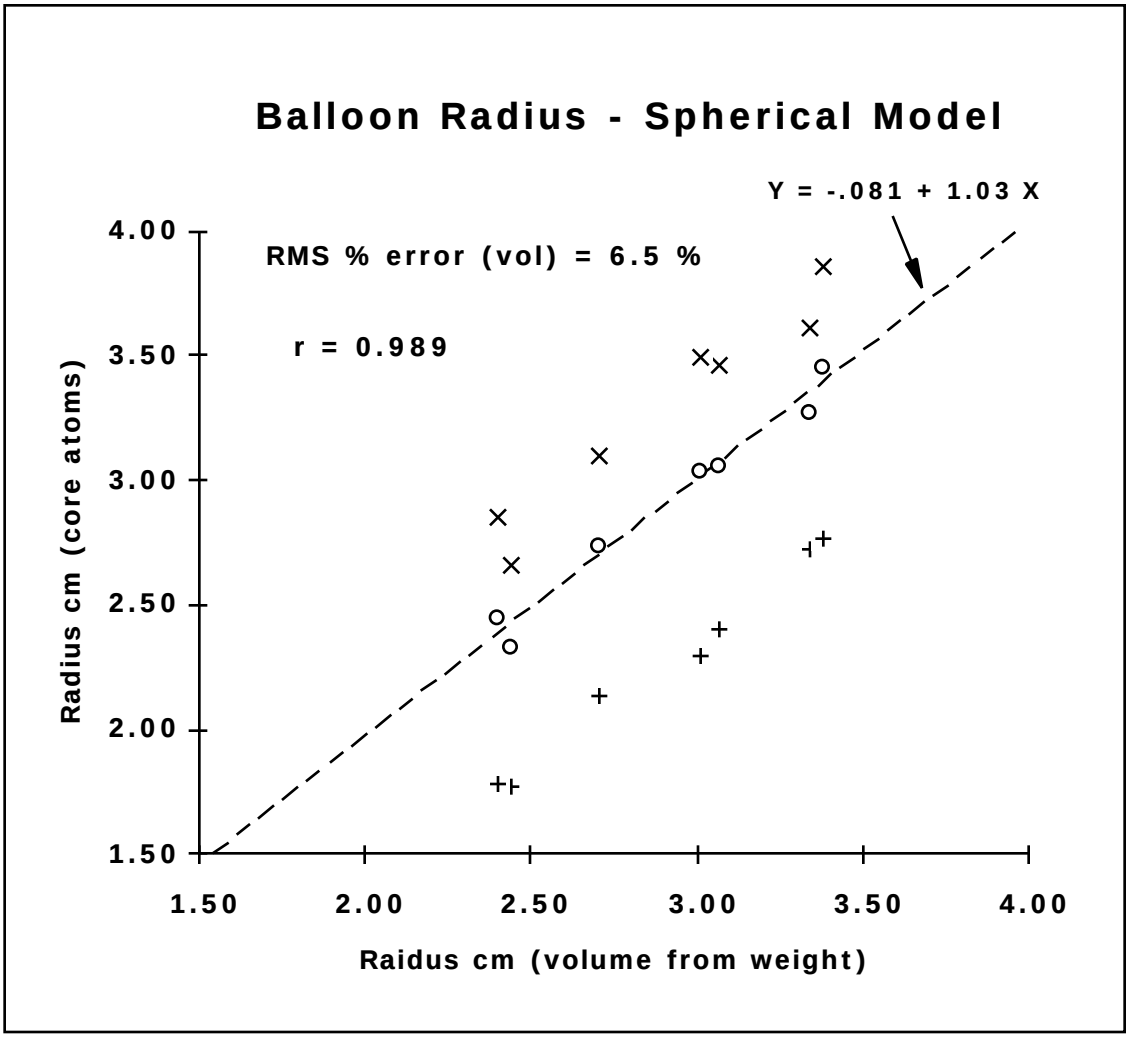


Fig. 8.4 Radius of fluid-filled balloons determined automatically using core atoms to analyze 3D ultrasound images, compared to radius of the same balloons determined by weight assuming a spherical shape. (**x**) outer facing boundary, (+) inner facing boundary, (**o**) weighted average of inner and outer boundaries, with weight determined to minimize RMS percent error by volume to 6.5%.

A complication arose because the skin of a balloon in ultrasound presents a dark-light-dark ridge in intensity rather than a single dark-light transition (again see Fig 8.3). This ridge causes boundary points to change orientation as one crosses the skin of the balloon. Using the inward facing boundary points to form core atoms yields a smaller radius than expected, while using the outward facing boundary points (by switching the polarity, defined in Section 3.1) yields a larger radius than expected. These results for the two cases are marked "**+**" and "**x**" respectively in Fig. 8.4.

An optimum weighted average radius $r = kr_I + (1-k)r_O$ was computed from the inner radius r_I and the outer radius r_O by minimizing the RMS percent error between the resulting calculated volume and that determined by weight, for the set of 7 balloons. The optimum value for k was 0.38, favoring the outward facing boundary points, yielding an RMS percent error by volume of 6.5%, which agrees favorably with previous measurements on the same balloons using a Hough transform (Stetten, Caines et al. 1995). The weighted average radii (marked "**o**" in Fig. 8.4) yielded a linear regression (dotted line) close to unity (slope = 1.03, intercept = 0.081 cm).

This measurement of balloon volume was completely automated and took advantage of the fact that many individual measurements were combined statistically. In a sense, each core atom serves as a yard-stick crossing the sphere near the center. Since generally the selected core atoms are chords of the sphere and not true diameters (as shown in Figs. 4.1C and 8.5), they can be expected to underestimate the actual diameter. This may explain the bias towards the outer surface in optimum value for k. Another reason for this bias may be that the balloons were generally ellipsoidal with a single major axis along the ultrasound beam, producing a propensity of shorter core atoms across the two minor axes.

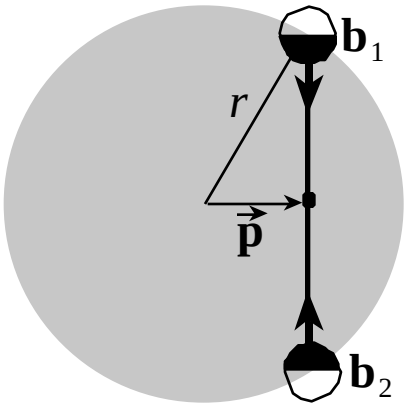


Fig. 8.5 Proposed correction for measurement with core atoms displaced from the center.

The displacement vector $\vec{\mathbf{p}}$ of a core atom relative to the center of a cluster (see Fig. 8.5) could theoretically be used to correct for the difference in length between a chord and a diameter. This would apply for the cylinder and the sphere, but not for the slab, since no such displacement occurs in a slab. This proposed correction has not been tested to date and remains in the list of future work.

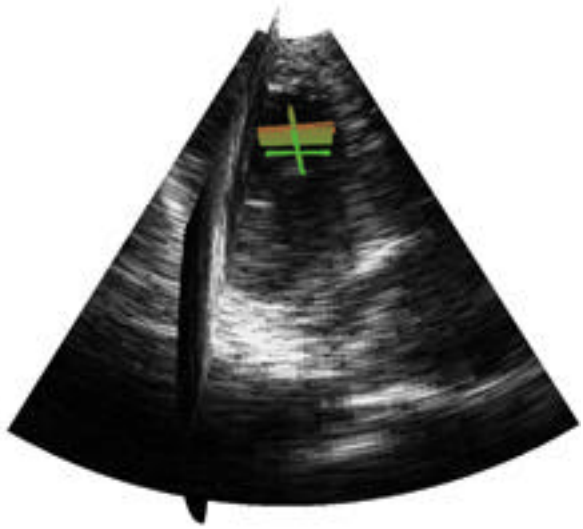
8C. Manual Tracing of LV Volume

Manual tracing was performed on short axis slices through the LV. The slices were produced orthogonal to manually identified endpoints of the AMV axis, using cell projection of I-mode slices created using the 3-Stripes algorithm developed in Section 6D. A stack of 16 parallel slices was collected for each scan. Volume was calculated for comparison with the automated method in Section 8D by multiplying the area within all of the traces by the slice thickness. The results were stored along with the physical coordinates of the traces for comparison with the automated method in 8E. Visualization of the traces superimposed in 3D on B-mode images was used to verify their accuracy (see Fig. 8.6).

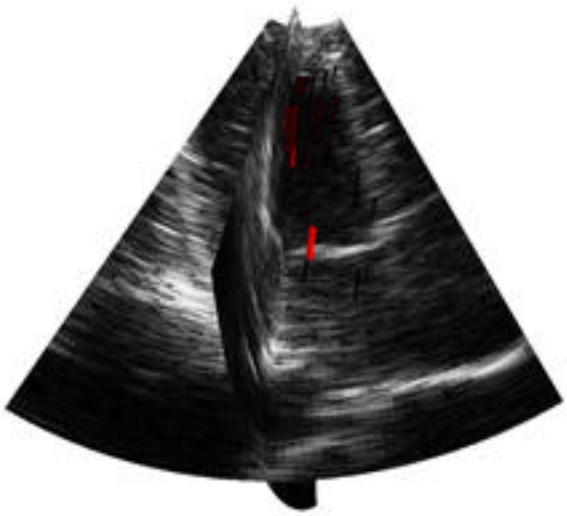
8D. Computing LV Volume with Core Atoms

The general approach used on balloons in Section 8B was applied to the left ventricle. Once the AMV axis was computed (as described in Section 8A), a set of wedges in a cylindrical coordinate system was established around the axis (as in Fig. 7.2B) and a mean radius to boundary points in each wedge computed, thus establishing a surface map for the LV. Missing portions of the surface map (containing no boundary points) were interpolated using 2D convolution on the surface map. Volume was computed for the ventricle using the cylindrical model. A sample is shown visualized in 3D in Fig. 8.6C, with the AMV axis in yellow and the surface map shown as a grid of dots with each dot representing a mean radius.

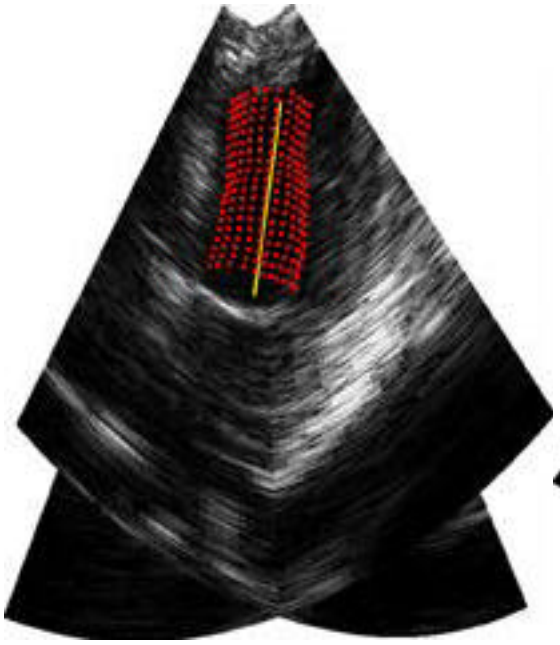
The results of this application of core atoms to measure volumes was far less encouraging than the corresponding application to determining balloon volume. The results are not reported here because the experiment was terminated. Although the core atoms had successfully found the AMV axis, they did so by forming from both endocardial and epicardial boundary points. These two surfaces are roughly concentric cylinders, and therefore they share the same axis. However, the core atom populations do not accurately represent the ventricular chamber because of the presence of the longer epicardial core atoms.



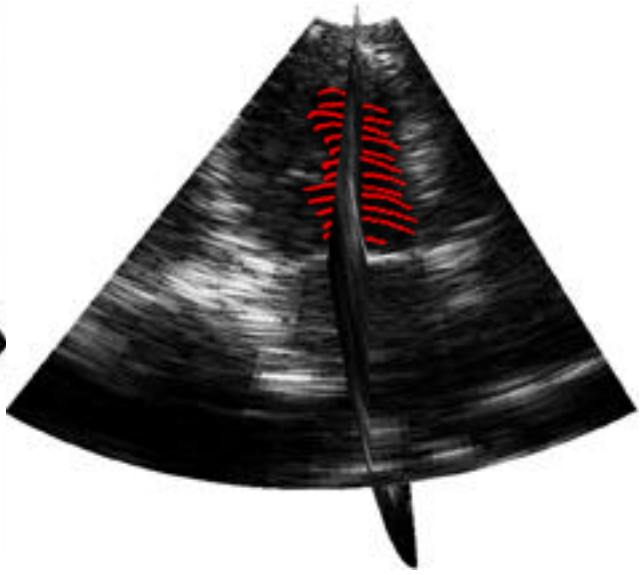
A. Cylinder of ventricle



B. Slab of mitral valve



C. Automated axis and surface map



D. Manual tracings of ventricle

Fig. 8.6 **A.** Color version of Fig. 8.1B with color mapped from lambda triangle as in Fig. 4.9. **B.** Color version of Fig. 8.1C. **C.** Automated surface map for the LV. **D.** Manual identification of left ventricle achieved by manual tracing on I-mode slices.

8E. Testing Fuzzy Segmentation of the LV

Although core atoms successfully identified the AMV in section 8A, they overestimated ventricular diameter in section 8D because they formed at both the endocardial and epicardial boundaries. This behavior led to the development of the fuzzy segmentation method based on voxel intensity and location described in section 7B.

The present section describes the application of the fuzzy segmentation method to the same set of *in vivo* human hearts for which the AMV axis was automatically established in section 8A. From the fuzzy segmentation, LV volume is computed and an effective endocardial surface determined. The results are analyzed against manual tracings as a gold standard.

The RT3D ultrasound sequences were divided into the same training and test sets as in Section 8A. The training set was used to optimize the method's parameters in terms of accuracy of volume measurement and to measure bias in the measurement using regression. Of the many parameters in the overall process, those which were adjusted included the shapes of the surface model in Fig. 7.10 and the probability curve in Fig. 7.12. Only a few iterations were performed. Fully optimizing these and other parameters in the method is beyond the scope of this dissertation. The method was then applied to the test set without further adjustment of parameters. The resulting measurements were compensated for the bias found in the training set.

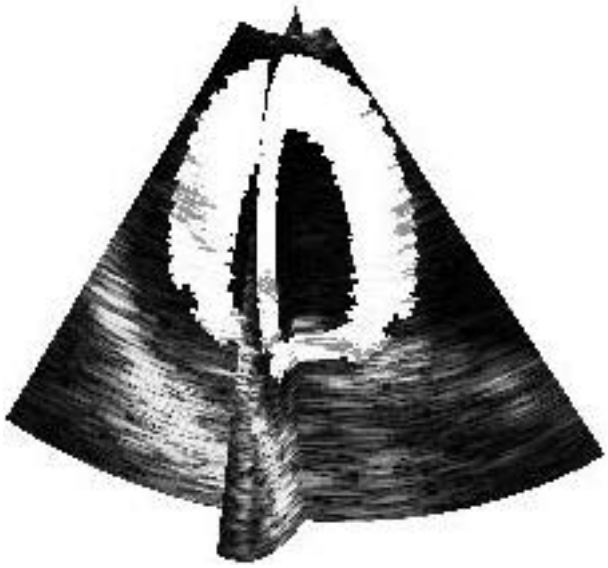


Fig 8.7 Region of interest (bright) and LV surface model (dark) superimposed on ultrasound data.

Fuzzy segmentation proceeded for each heart as follows: once the AMV axis was automatically established as in section 8A, a region of interest (ROI) and an expected boundary of the ventricle were established around the AMV axis using the surface model illustrated in Fig. 7.10. Scaling factors S_a' and S_m' were determined by the MNM model as described in section 7B. In Fig. 8.7 an example of an ROI (bright) encompassing an LV surface model (dark) is shown located, oriented, and scaled by the MNM. A probability $p_L(j)$ was computed for each voxel based on its r and ϕ relative to the surface model, using the probability function in Fig. 7.12. An example of $p_L(j)$ for one image is shown in Fig. 8.8.

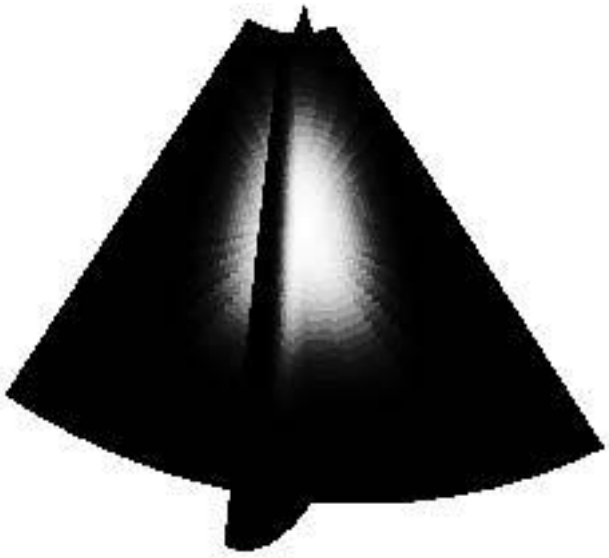


Fig. 8.8 The probability $p_L(j)$ for each voxel with values ranging from 1 (white) inside and 0 (black) outside the expected ventricle.

A weighted mean of voxel intensity $\bar{I}$ within the ROI was then computed using Eq. 7.19 with each voxel in the ROI being weighted by its $p_L(j)$. The weighted mean $\bar{I}$ was used to compute $p_I(j)$ for each voxel using the probability function in Fig. 7.13. An aggregate probability $p_A(j)$ was then computed for each voxel using Eq. 7.21 and a corresponding total volume v_A was calculated using Eq. 7.22. The aggregate probability is shown in Fig. 8.9 with values ranging from 1 (white) to 0 (black). As may be seen, $p_A(j)$ is affected by both $p_L(j)$, which dominates the overall distribution, and $p_I(j)$, which accounts for finer detail.

It is clear from Fig. 8.9 that the fuzzy segmentation method does not yield a reasonable probability for each individual voxel. The fuzzy representation lacks an explicit surface and therefore also lacks a distinct topology. The dark areas corresponding to non-ventricular structures, such as the mitral valve and epicardium, clearly do not make topological sense

floating disconnected from each other. Only in certain statistical operations, such as the computation of volume, is the fuzzy shape representation reasonable.

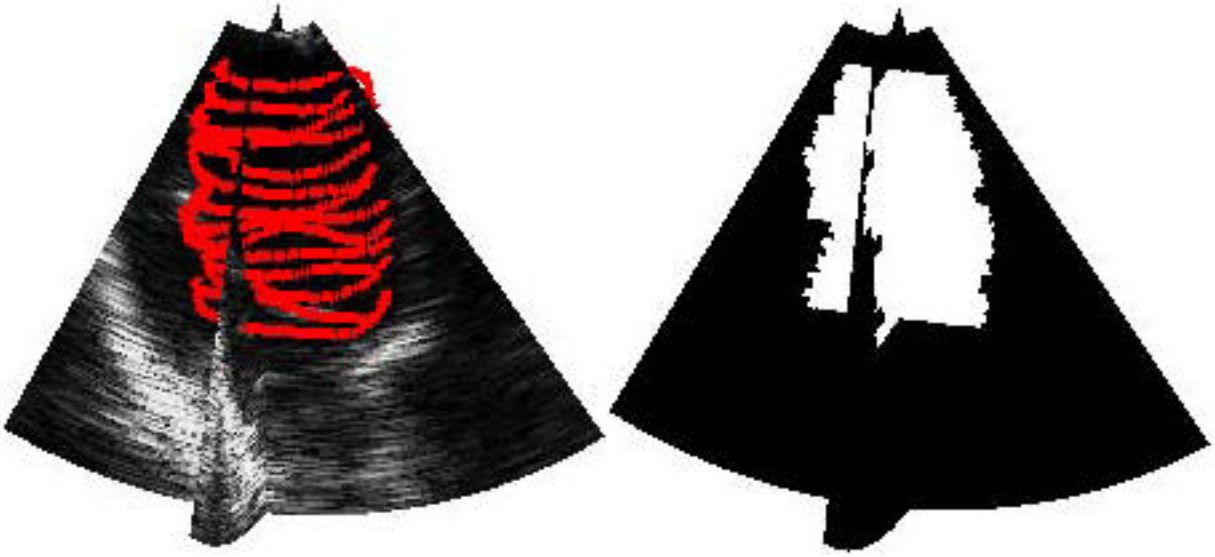
Fig. 8.9 The aggregate probability $p_A(j)$ for each voxel with values ranging from 1 (white) to 0 (black).

An analysis of the accuracy of the fuzzy segmentation method was based on manual tracings performed on the same data. Each image was manually analyzed as follows: the end points of the AMV axis were established manually and a stack of I-mode slices was created orthogonal to the axis, as described in Section 6D. Manual tracings were performed on each I-mode slice by a single operator, the author. An example is shown in Fig. 8.10A. Segmentation was then performed by examining each voxel in the data for its location relative to the traces. A probability $p_T(j)$ of the j^{th} voxel being within the manual tracings was computed by projecting the voxel location onto the nearest I-mode slice in the stack. If the projection placed it inside the corresponding manual tracing, then $p_T(j) = 1$. Otherwise, $p_T(j) = 0$. If the voxel was located such that it projected onto the AMV axis beyond either

endpoint , then $p_T(j) = 0$. Figure 8.10 B shows the result of this process. A volume v_T was then computed from the tracings by

$$v_T = \sum_{j \in ROI} p_T(j) v(j) \tag{8.1}$$

where $v(j)$ is the volume of voxel j. Recall that $v(j)$ is not constant in the azimuth-elevation coordinate system (see Section 6A). Manual tracings were performed three times by the observer, each time with individual manual placement of the AMV axis. The arithmetic mean $\bar{v}_T$ of the three resulting values of v_T was computed.



A **B**

Fig. 8.10 A. Manual tracings. **B.** Corresponding set of labeled voxels with white symbolizing $p_T(j) = 1$ and black, $p_T(j) = 0$.

Results for the training data are shown in Fig. 8.11 comparing v_A from the aggregate probability method against $\bar{v}_T$ from the 3 manual tracings. The correlation coefficient is $R = 0.88$ and the standard error of the estimate is $S_{x,y} = 10.4$ ml . The standard deviation between treatments for v_T computed from the 3 manual tracings is 20.7 ml. Using slopes and intercepts from the training data, compensated volumes were computed from automated volumes determined from the training data and compared to the mean volume from the 3

manual traces. Results are shown in Fig. 8.12 with an RMS error of 25.9 ml. This represents absolute accuracy using new data with compensation for bias from the training data.

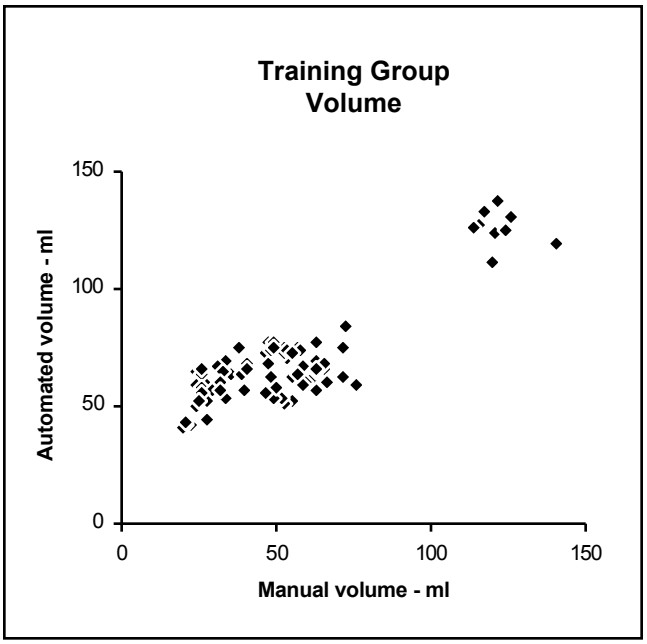


Fig. 8.11 Training data: automated volume v_A vs. manual volume $\bar{v}_T$ for all 88 scans.

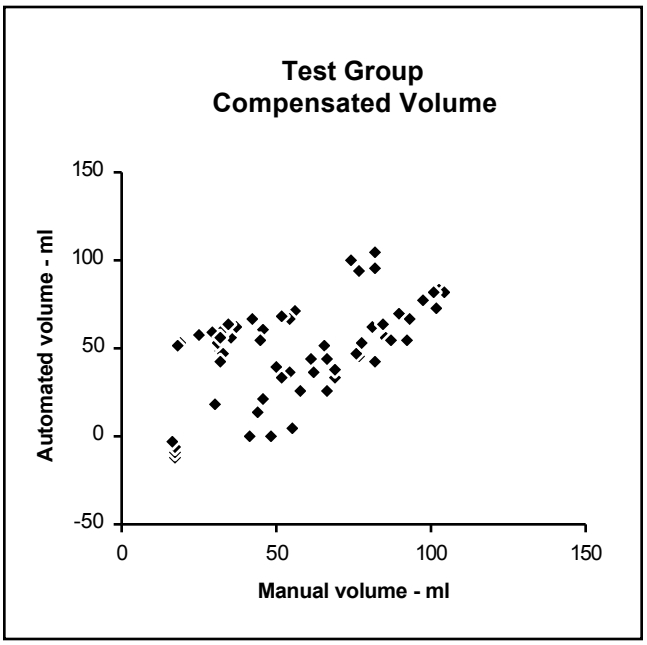


Fig. 8.12 Test data compensated with slope and intercept from regression on training data: automated volume v_A vs. manual volume $\bar{v}_T$ for all 65 scans.

Improved results are possible if, instead of total volume, change in volume for a given heart is measured. Closer inspection of Fig. 8.12 shows several distinct bands in the data, suggesting that the method is sensitive to variations between the scans on individual hearts. A further analysis shows this to be the case. Each heart in the data appears in a sequence containing between 3 and 14 images. The difference between volume from each image and the mean volume for its particular sequence is plotted in Fig. 8.13 for manual vs. automated compensated volumes performed on the test data. A reduced RMS error of 9.2 ml is seen for volume change within a given heart. The ability to accurately determine change in volume has clinical significance even if total volume is not as accurate, since volume change can yield stroke volume or provide a measure of heart function over the course of short-term therapy.

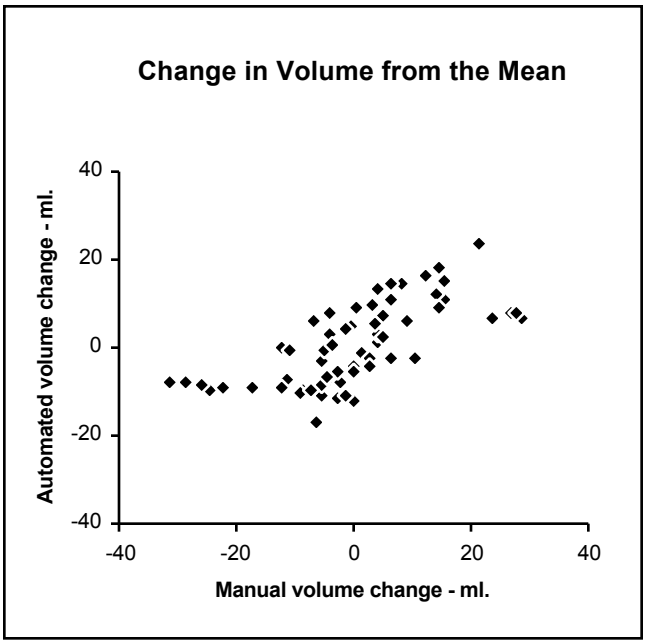


Fig. 8.13 Delta Volume for test data computed as difference from the mean for automated vs. manual volumes for all 65 scans.

It is desirable to conduct a more detailed comparison between the automated and manual segmentations, taking local boundary location into account. The accuracy of segmentation is not well judged solely by comparing total volume, since an object's volume says nothing about its location and since different shapes many have the same volume.

Using the medial framework already established, error between the manual and automated segmentations can be localized in spherical coordinates ϕ, ω, and r (see Fig. 7.10). Longitude ω is not a parameter in the ventricular surface model, but real ventricles are asymmetrical around the AMV axis and segmentation error are therefore expected to be likewise asymmetrical. Since all the scans in the data were performed using the same

transducer orientation including rotation around its axis, error can be correlated as a function of both ϕ and ω.

For each image, the mean location for the AMV axis from the 3 manual tracings served as a common frame of reference for comparing the three manual segmentations $p_T(j)$ against the automated segmentation $p_A(j)$. In each case, voxels were sorted by ϕ and ω to produce sets of voxels. Each set $S(\phi,\omega)$ contained all the voxels in a given wedge from which volumes $v_T(\phi,\omega)$ and $v_A(\phi,\omega)$ were calculated, respectively

$$v_T(\phi,\omega) = \sum_{j \; S(\phi,\omega)} v(j)p_T(j) \qquad \text{and} \qquad v_A(\phi,\omega) = \sum_{j \; S(\phi,\omega)} v(j)p_A(j) \tag{8.2}$$

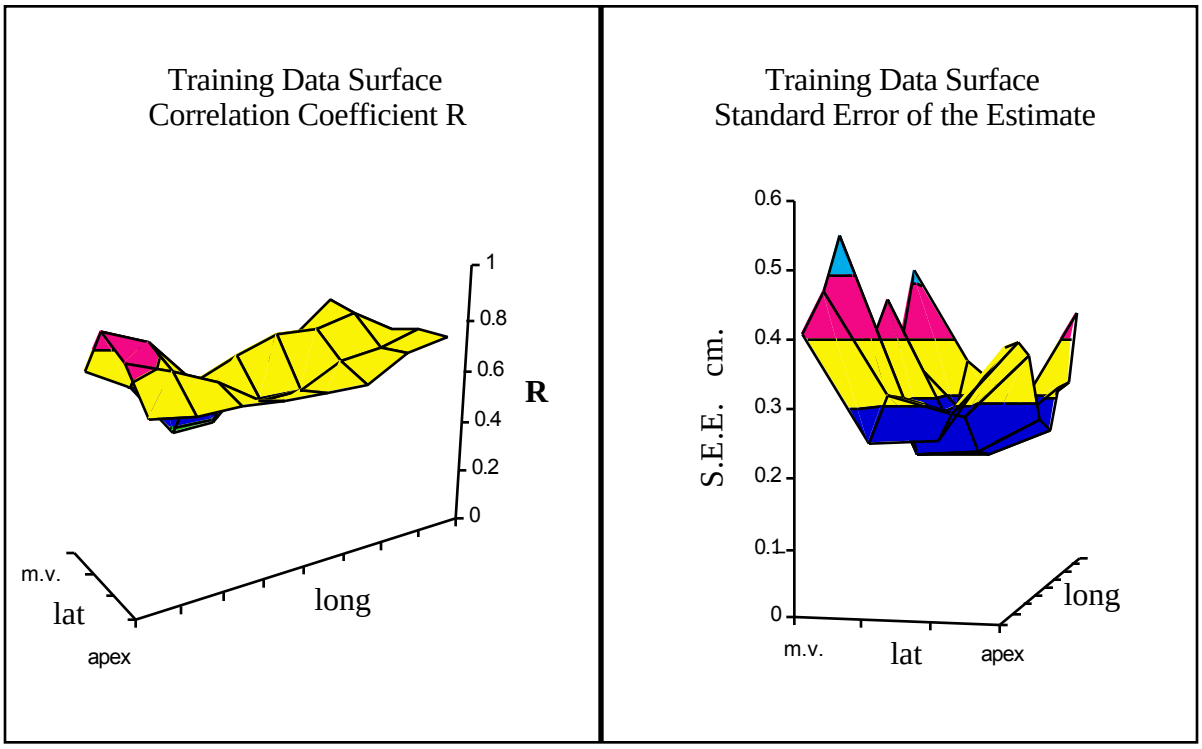
Although the segmentations were based on voxels, an effective radial distance within each wedge was computed using the standard formula for the volume v of a solid angle $\Delta\phi\Delta\omega$ in a sphere of radius r

$$v = \frac{r^3}{3}\sin\phi\Delta\omega\Delta\phi \quad . \tag{8.3}$$

The effective radius $r(\phi,\omega)$ for a particular wedge was determined by the volume $v(\phi,\omega)$ of that wedge,

$$r(\phi,\omega) = \left(\frac{3v(\phi,\omega)}{\sin\phi\Delta\omega\Delta\phi}\right)^{\frac{1}{3}} . \tag{8.4}$$

The effective radius $r(\phi,\omega)$ over the domain of ϕ and ω represents an endocardial surface map referenced to the AMV axis, and can provide a local measure of segmentation error. Using Eq. 8.4, an automated endocardial surface map $r_A(\phi,\omega)$ was computed from $v_A(\phi,\omega)$, and 3 manual surface maps $r_T(\phi,\omega)$ were computed from the respective determinations of $v_T(\phi,\omega)$ for the 3 sets of manual traces. An arithmetic mean surface map $\bar{r}_T(\phi,\omega)$ for the 3 individual $r_T(\phi,\omega)$ was also computed to represent a mean manual tracing.



A **B**

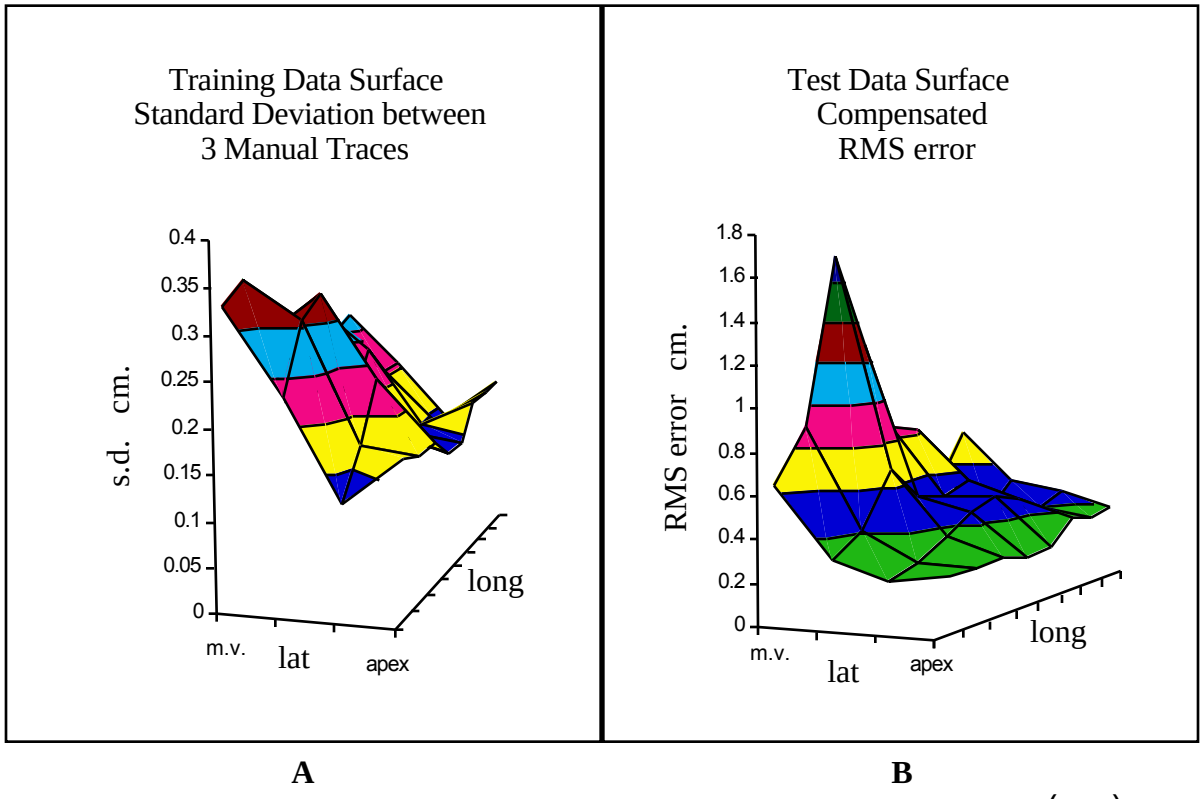
Fig. 8.14 Training data: Correlation between endocardial surface maps from automated $r_A(\phi,\omega)$ and manual $\bar{r}_T(\phi,\omega)$ analysis. Local correlation coefficient (**A**) and standard error of the estimate (**B**) shown as a function of ϕ and ω.

Results for the training data are shown in Fig. 8.14 comparing $r_A(\phi,\omega)$ with $\bar{r}_T(\phi,\omega)$ by plotting the correlation coefficient and standard error of the estimate as functions of ϕ and ω. The correlation coefficient within a given wedge is in the 0.6 to 0.8 range, representing fairly good correlation between boundary locations in different images. The standard error of the estimate falls in the range 0.2 to 0.6 cm, and is seen to be greatest at the apex and the mitral valve.

The test data were then analyzed and $r_A(\phi,\omega)$ calculated for those data. The linear regression computed from the training data was used to compensate $r_A(\phi,\omega)$ from the test data for expected bias. The RMS error with respect to the mean manual tracings $\bar{r}_T(\phi,\omega)$ for the test data is shown in Fig. 8.15B. In general, the error falls in the 0.2-0.8 cm range, with errors up to 1.8 cm occurring near the mitral valve. The large spike near the mitral valve may correspond to the aorta, which leaves the LV adjacent to the mitral valve and provides an ambiguity as to inclusion in the ventricle. Considering the general scale for the ultrasound scans of 15 cm, the general dimensions for the ventricle of 3-7 cm, and the poor visibility of the endocardium in the ultrasound images, these results seem a rather good achievement.

To gauge the precision of the manual tracings, the standard deviation between $r_T(\phi,\omega)$ from each of the 3 tracings was computed, shown in Fig. 8.15A. Intraobserver error is in the range 0.1 to 0.4 cm, less than the error from the automated results. For the manual tracings, the error is greatest at the mitral valve, with the apex located relatively accurately. *Inter*-observer error based on more than one human operator would likely be larger, perhaps comparable to that of the automated system. Determination of *inter*-observer error should be carried out in future work.

This section has shown empirically that fully automated segmentation of the cardiac LV in RT3D ultrasound data is possible using statistical analysis of medial primitives to establish the AMV axis and subsequent fuzzy segmentation based on voxel intensity and location relative to that axis. The next chapter will discuss the strengths and weaknesses of the various methods developed in this dissertation, and possible future directions for research.



A **B**

Fig. 8.15 A. Training data: Standard deviation in endocardial surface maps $r_T(\phi,\omega)$ from 3 sets of manual traces shown as a function of ϕ and ω. **B.** Test data: RMS error between endocardial surface maps $r_A(\phi,\omega)$ compensated for bias and $\bar{r}_T(\phi,\omega)$ as a function of ϕ and ω.