

# Semi-automated Measurement of Anatomical Structures Using Statistical and Morphological Priors

Edward A. Ashton and Tong Du  
VirtualScopics, LLC, 350 Linden Oaks, Rochester, NY, USA 14625

## ABSTRACT

Rapid, accurate and reproducible delineation and measurement of arbitrary anatomical structures in medical images is a widely held goal, with important applications in both clinical diagnostics and, perhaps more significantly, pharmaceutical trial evaluation. This process requires the ability first to localize a structure within the body, and then to find a best approximation of the structure's boundaries within a given scan. Structures that are tortuous and small in cross section, such as the hippocampus in the brain or the abdominal aorta, present a particular challenge. Their apparent shape and position can change significantly from slice to slice, and accurate prior shape models for such structures are often difficult to form. In this work, we have developed a system that makes use of both a user-defined shape model and a statistical maximum likelihood classifier to identify and measure structures of this sort in MRI and CT images. Experiments show that this system can reduce analysis time by 75% or more with respect to manual tracing with no loss of precision or accuracy.

## 1. INTRODUCTION

Accurate identification and measurement of various anatomical structures is a vital tool both for surgical planning and for evaluation of disease progression and patient response to therapy for numerous diseases. Measurement of hippocampal volume is an important endpoint for diagnosing and monitoring both intractable temporal lobe epilepsy and Alzheimer's disease. Identification of the aorta and associated vessels and measurement of various related parameters are vital tools for evaluation of abdominal aortic aneurism progression and response to treatment. Current standard methods for obtaining these data points are largely manual and subjective, and are therefore both error-prone and subject to inter- and intra-operator variability. In addition, manual tracing of structures such as the vascular system, which may appear on up to 800 images in a single study, requires both considerable expertise and a great deal of time.

Significant research effort has been devoted to the subject of identification of curvilinear and poorly defined structures in medical images, but there is at this time no generally accepted solution. de Bruijne *et al.* have demonstrated a method for identifying aortic stents after surgery through use of radio-opaque markers sewn into the stent prior to surgical implantation.<sup>1</sup> This technique, while effective, is somewhat invasive and clearly not useful for pre-surgical evaluation.

Ashton *et al.*<sup>2</sup> and Hsu *et al.*<sup>3</sup> have presented semi-automated methods for the identification and measurement of the hippocampus. Ashton *et al.*<sup>4</sup> have presented a method for identification and measurement of structures with simple (ovoid) shape, such as solid soft-tissue tumors. These methods are all reasonably well suited to the hippocampus measurement problem, but are not applicable to long, tortuous and brachiated structures such as the abdominal vascular system.

Taylor and Barrett have presented a method for segmentation of structures using competitive region growth without any *a priori* shape constraint.<sup>5</sup> This method is well suited to well-bounded regions such as the vascular lumen, but is not well suited to structures such as the vascular thrombus or the hippocampus, which frequently have ill-defined boundaries with adjacent structures.

Carlbon *et al.*<sup>6</sup> have presented a method for application of deformable templates to segmentation of neurological structures. Numerous researchers, including Cohen<sup>7</sup> and Chung<sup>8</sup> have demonstrated the use of 2-D active contours (snakes) and their derivatives in providing edge-based structural identification in medical images. These methods can be applied to the problems presented here, but will generally require significant user interaction on each image, making their use impractical for large data sets such as those seen in the abdominal cases.

Sato *et al.*<sup>9</sup> have described a segmentation method geared towards vascular and other curvilinear structures using a hierarchical filtering approach. Aylward and Bullitt<sup>10</sup> have proposed a method for identifying the center line of structures such as the vascular system. Krissian *et al.*<sup>11</sup> have demonstrated a method for identifying tubular structures such as the abdominal vasculature using a shape model approach. This approach and those described in the previous two references work well as long as the shape assumptions are valid. However, they have difficulty when these assumptions break down, as at bifurcations. In addition, these methods are not able to identify associated structures such as thrombus or calcifications, and have not been demonstrated to be effective in cases where significant artifacts are present, as in post-aortic endograft CT images.

Other methods that are able to segment aortic vessel boundaries but not thrombus, and which have significant difficulty with bifurcation points and tortuous vessels include those presented by Wink *et al.*,<sup>12</sup> Verdonck *et al.*,<sup>13</sup> Fiebich *et al.*,<sup>14</sup> and Bulpitt and Berry.<sup>15</sup>

## 2. MEASUREMENT TECHNIQUES

The anatomical feature delineation technique evaluated in this work makes use of three types of information: (1) A statistical description of the various tissue types present in the images. This information is obtained automatically through the use of a maximum likelihood classifier. (2) A statistical description of the tissue or tissues of interest. This information is obtained by making use of an anatomical atlas or user input – typically one or more seed regions or exemplars. The exemplars will ideally take the form of an identification of the structures of interest on one or more non-adjacent images. This identification can be carried out using manual tracing, or using one of several semi-automated feature identification techniques such as Geometrically Constrained Region Growth or LiveWire.<sup>16</sup> (3) A morphological description of the structure of interest. This information can be taken from an atlas-derived *a priori* shape model in the case of well-defined and localized structures such as the hippocampus. If such information is not available, a three-dimensional surface is fit to the provided exemplar regions. In the case of the abdominal vascular system, a surface is fit to the outer boundary of the lumen, and another is fit to the outer vessel wall. Other tissue types, such as thrombus and calcifications, are assumed to exist in the region between these two boundaries. These data serve as the inputs to a second-pass maximum likelihood classifier, which identifies the boundaries of the structure or structures of interest.

Maximum likelihood classification refers to the process of optimally separating an image into areas of similar statistical behavior. It is assumed that regions of similar statistical behavior will correspond to different tissue types. The goal of the MLC algorithm used in this invention is to globally maximize the following discriminant function:

$$g_i(x) = -\ln |R_i| - (x - m_i)' R_i^{-1} (x - m_i) \quad (1)$$

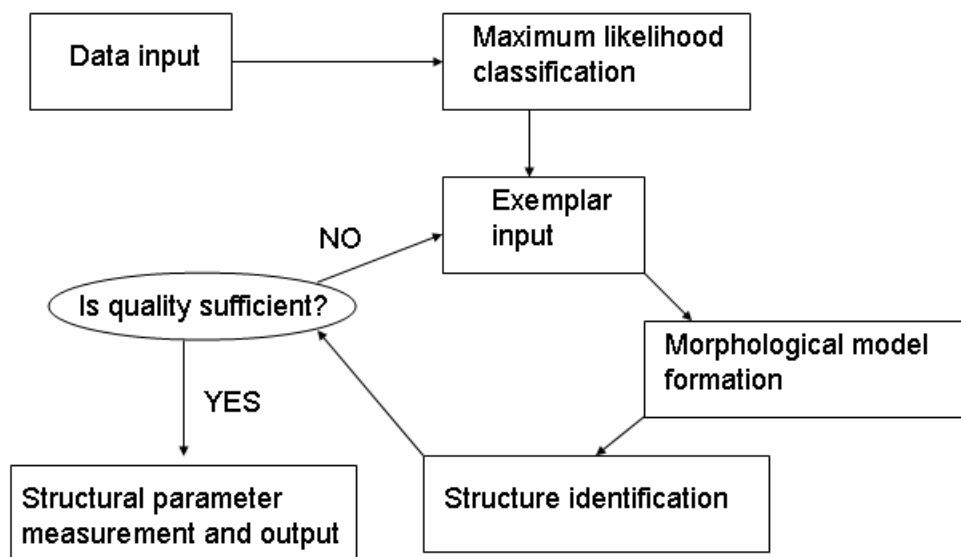
where  $R_i$  is the covariance matrix for class  $i$ ,  $m_i$  is the mean vector for class  $i$ , and  $x$  is the value vector describing the voxel under consideration. This discriminant function is applied to the first-pass case, where *a priori* spatial distribution probabilities are not available. In the second pass, where spatial priors are obtainable based on the result of the exemplar surface fit or anatomical atlas information, the discriminant function becomes

$$g_i(x) = \ln |p_i| - \frac{1}{2} \ln |R_i| - \frac{1}{2} (x - m_i)' R_i^{-1} (x - m_i) \quad (2)$$

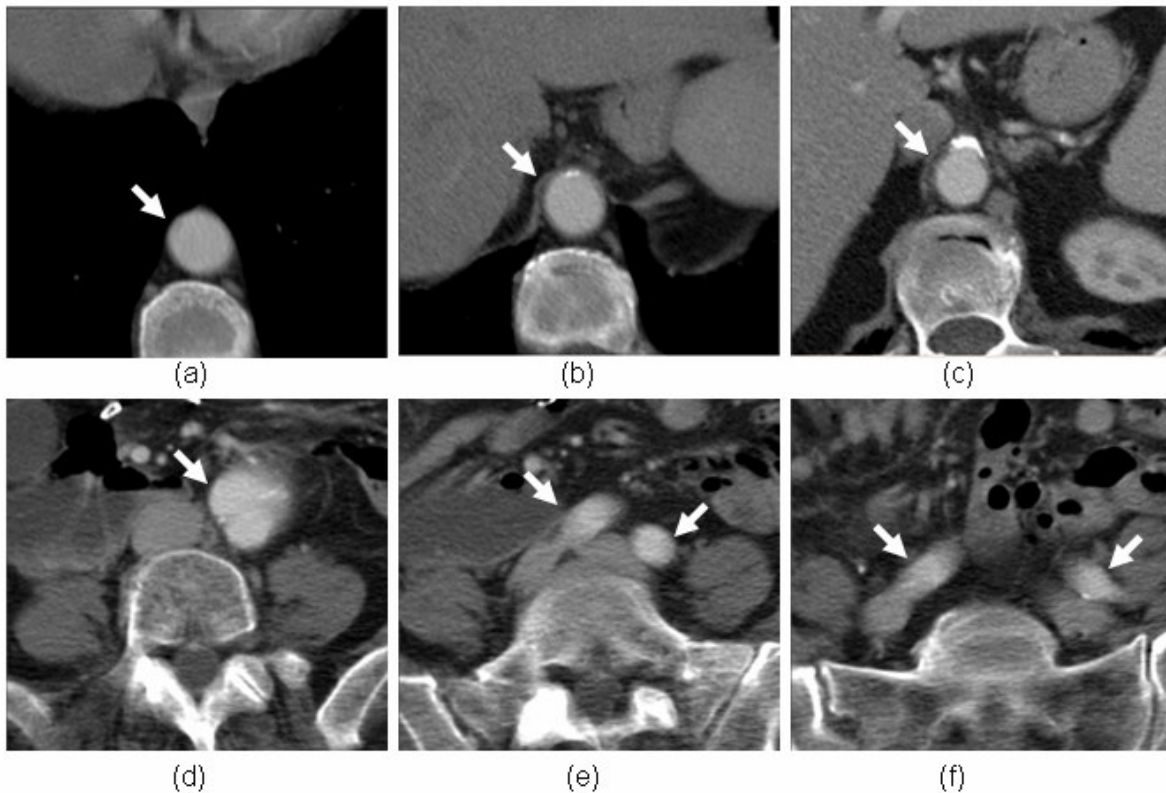
where  $p_i$  is the *a priori* probability of class  $i$  appearing at the voxel under consideration. These probabilities are generated by smearing the assumed boundaries for each tissue type taken from the anatomical atlas or surface fit using a Gaussian blur, then normalizing such that the *a priori* probabilities of all tissue classes sum to 1 at each voxel. Discriminant maximization can be accomplished using any of several known optimization techniques, such as iterated conditional modes (ICM),<sup>17</sup> or simulated annealing (SA).<sup>18</sup> A diagram describing information flow in this system is given in Figure 1.

The most important factor affecting performance of this system is the accuracy of the *a priori* structural shape model. Where anatomical atlas information is used to generate this model, accurate registration of the atlas to the experimental data is critical. Where shape information is generated using surface fitting, placement and accuracy of the exemplar regions is the most critical factor affecting system performance. In the experiments described below exemplar regions were placed at regular intervals. This was done in order to minimize the effects of user judgment on results, but clearly this is not an optimal approach. In the abdominal aorta, for example, the region above the aortic bifurcation is fairly regular in the axial plane and changes little from image to image. In this region, near-optimal results can be obtained with widely spaced exemplars. Below the aortic bifurcation, however, the branches of the aorta become smaller, more tortuous, and more skewed with respect to the axial plane. More closely spaced exemplars are therefore needed to provide a reasonably accurate shape model. This situation is illustrated in Figure 2.

Because of these factors, the number of exemplars required to provide a given level of performance in the experiments described below should be considered an upper limit. Intelligent placement of exemplars should in almost all cases provide superior results.



**Figure 1:** Information flow in the analysis system. Structure identification is carried out using maximum likelihood classification with statistical and morphological priors derived from the morphological model. The software allows for insertion of additional exemplars and formation of a new morphological model if the identification results are not sufficiently accurate.



**Figure 2:** Images from a CT scan of an abdominal aortic aneurism patient. Separation between (a) and (c) is approximately 4.8cm. The aorta changes very little in this region, allowing a close fit with only 2 exemplars. Separation between (d) and (f), which are just below the aortic bifurcation, is also 4.8cm. In this region the vasculature is smaller and more tortuous, requiring a greater number of exemplars for an accurate surface fit.

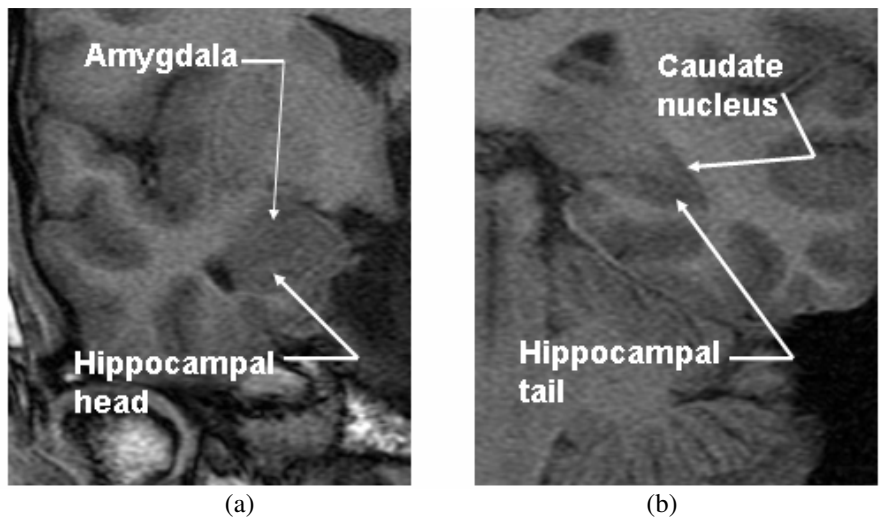
### 3. MATERIALS AND METHODS

Two experimental applications of this technique were tested in this work. In the first, the process was used to identify, delineate and measure the hippocampus in T1 weighted MRI images of normal volunteers. In the second, the process was used to identify, delineate and measure (separately) the lumen and surrounding thrombus in CT images of patients suffering from abdominal aortic aneurisms. Both experiments sought to quantify three variables describing system performance: processing time, accuracy with respect to manual tracing, and reproducibility in terms of coefficient of variability over repeated measurement of the same anatomy.

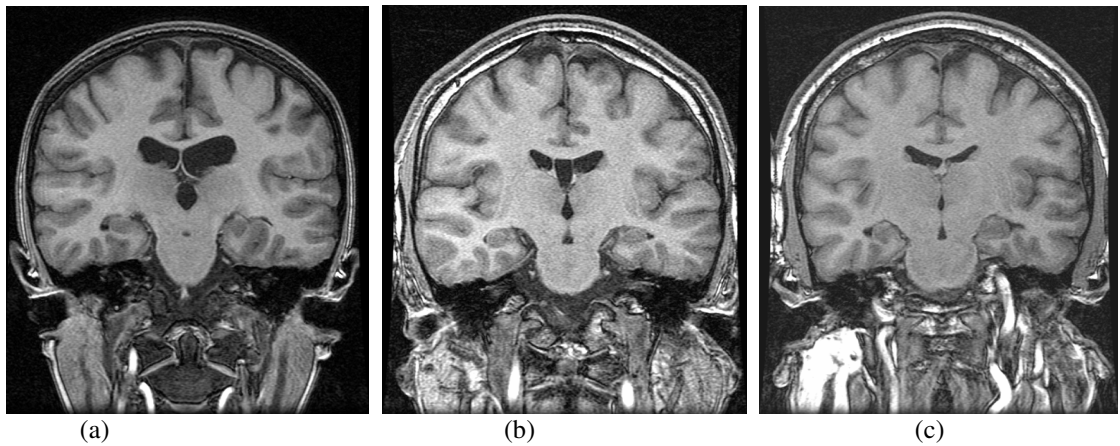
The first two variables are intertwined, since the goal of the system is to achieve optimal processing time while producing results statistically indistinguishable from those obtained by manual tracing. In order to determine the normal variability of manual measurement, the right hippocampus of one subject and the abdominal aorta of one subject were manually delineated four times by a single operator and once each by four operators. Automated results were considered to be indistinguishable from manual results if they fell within one standard deviation (as determined by these experiments) of the corresponding manual measurements.

The hippocampus is a gray matter structure of the human brain, located adjacent to the amygdala and the caudate nucleus and attached to the gray matter of the cerebral cortex. See Figure 3. Because the hippocampus is small, tortuous, and lacks clear boundaries with several adjacent structures, its identification and measurement are

particularly difficult. The object of the first experiment was to determine the accuracy, speed and precision of the system described in this work in identifying and measuring the hippocampus. A data set was obtained which consisted of 5 coronal T1 weighted MRI studies taken from normal volunteers. All volunteers provided informed consent prior to enrollment in this study. MR acquisition was 3D, with a slice thickness of 2.5mm. Sample images from this data set are given in Figure 4.



**Figure 3:** (a) Separation of the right hippocampal head from the basal nucleus of the amygdala. (b) Separation of the left hippocampal tail from the tail of the caudate nucleus.



**Figure 4:** Sample images from the coronal MRI data sets for: (a) Subject 1; (b) Subject 3; (c) Subject 5.

The second application considered in this work involved identifying and measuring the vessels and surrounding thrombus of the abdominal vascular system. Accurate mapping and measurement of abdominal aortic aneurisms and the surrounding vasculature is a vital tool for both surgical planning and patient follow-up. Current manual methods for vascular classification are very time consuming, since a typical abdominal CT scan may contain up to 800 individual images. In this case, the goal of the system is to provide a result that is statistically indistinguishable from a manual identification while enabling a substantial time savings.

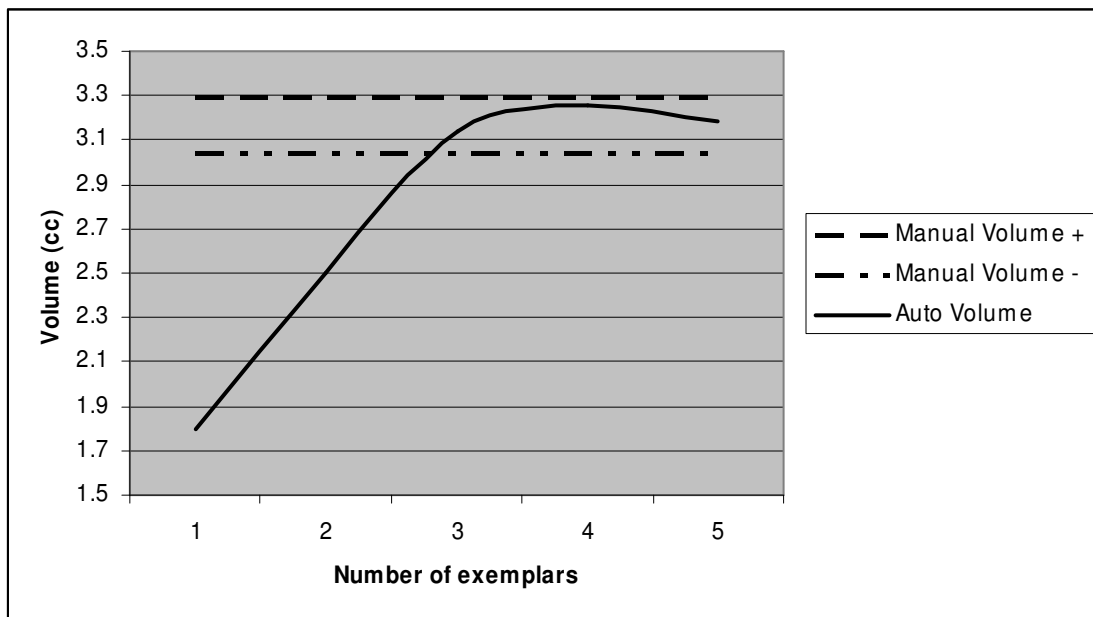
Experimental data for this experiment was obtained from an anonymized database of CT scans taken from pre- and post-surgical abdominal aortic aneurism patients. 5 subjects were analyzed. Images were standard spiral CT with in-plane resolution of .78mm and slice thickness of 2mm. Sample images from Subject 1 are shown in Figure 2.

## 4. RESULTS

### 4.1 Hippocampus Measurement Experiment

In order to establish a gold standard and an associated error margin, both hippocampi of each subject were identified by four expert analysts using a computer-aided manual tracing process. The experiment was intended to determine: (1) How many exemplars were required to produce an automated measurement that was statistically indistinguishable from a manual one? (2) What was the time savings associated with this process, as compared to manual tracing? (3) What was the reproducibility of the automated process?

In order to answer the first question, the right hippocampus for Subject 1 was measured, using a varying number of exemplars for morphological model formation. These results were compared to manual measurements of the same structure. The results of this experiment are given in Figure 5.



**Figure 5:** Plot of manual vs. automated volume for hippocampal measurement with varying numbers of exemplars. The manual volume + and manual volume - lines represent the mean manual measurement plus and minus one standard deviation. In this case, results with three or more exemplars are statistically indistinguishable from manual measurements.

The question of time savings can be answered by examining the number of exemplars required for adequate results. The hippocampus in this case extended over a total of 16 images. Because only three were needed as exemplars, time savings with respect to manual tracing of the entire structure should be at least 80%. In practice, because the exemplars were defined using single click geometrically constrained region growth<sup>4</sup> time savings were in excess of 90%.

In the second phase of this experiment, the right hippocampus of each of the five subjects was analyzed four separate times. The intent in this case was to establish the reproducibility of this technique. Results of this experiment are given in Table 1. Clearly, in the case of hippocampal measurement this invention provides clear advantages over current methods in terms of speed, accuracy, and precision.

	Repeat 1	Repeat 2	Repeat 3	Repeat 4	Mean	Std. Dev.	Coef. Var.
Subject 1	3.143	3.284	3.183	3.268	3.22	0.06	2.1%
Subject 2	3.34	3.351	3.379	3.154	3.306	0.10	3.1%
Subject 3	3.219	3.21	3.27	3.259	3.24	0.03	0.9%
Subject 4	2.647	2.836	2.79	2.748	2.755	0.08	2.9%
Subject 5	3.069	3.179	3.132	3.179	3.14	0.05	1.7%

**Table 1:** Results of hippocampus reproducibility experiment. Numbers are hippocampal volumes in cubic centimeters. Mean coefficient of variability is 2.14%. This compares to reported values of 5% - 7% for manual identification.

#### 4.2 Abdominal Aortic Aneurism Measurement Experiment

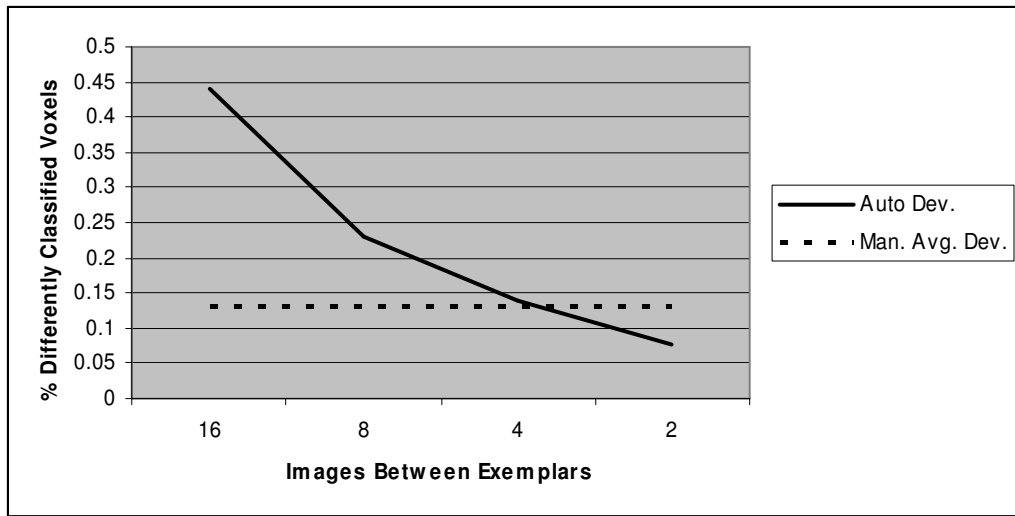
The intent of this experiment was to examine the speed, precision and accuracy of the analysis system in identifying and measuring the lumen and surrounding thrombus in CT scans of abdominal aortic aneurism patients with respect to that of manual tracing. In order to provide a precision baseline, a forty image section straddling the aortic bifurcation for one pre-surgical patient was analyzed manually five times over a period of one week by a single analyst. The first identification was considered to be baseline, while the next four were considered repeats. The parameter of interest in this case was the number of voxels classified differently on the baseline and each repeat, expressed as a percentage of total pixels of a given class in the baseline identification. Results of this experiment are given in Table 2. It should be noted that because this experiment was carried out using the less-complicated pre-surgical case, was carried out using a single operator, and had a wash-out period of only one to two days between identifications, the variability shown in Table 2 should be considered a minimum. Multiple analysts or less well conditioned data would most likely yield significantly higher variability numbers.

	Repeat 1	Repeat 2	Repeat 3	Repeat 4
% Lumen Difference	12.9	12.8	14.9	12.1
% Thrombus Difference	43.0	43.3	45.7	44.7
% Lumen Vol. Difference	5.4	10.1	11.9	1.3
% Thrombus Vol. Difference	5.6	5.6	9.2	15.6

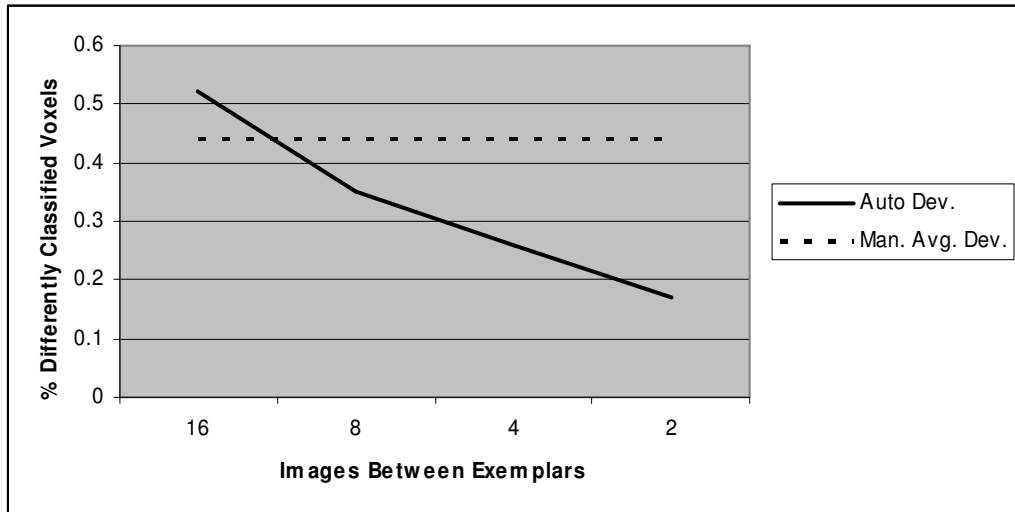
**Table 2:** Results of manual vasculature identification precision experiment. Note that volume differences are small relative to pixel classification differences, particularly for thrombus identification.

In order to determine both the accuracy and time savings possible using the method described here, five 228 image CT scans were fully analyzed manually, with all regions of both lumen and thrombus identified. This identification served as baseline. Varying numbers of exemplars were then used until the results fell within the bounds defined by the previous experiment. Results of this experiment are shown in Figure 6.

The results of this experiment are quite consistent with those of the hippocampus experiment. Results statistically indistinguishable from manual measurement are achieved with roughly one exemplar for every four images. This provides a potential time savings of 75% or more, with an accuracy equal to or better than that provided by manual measurement. It should be noted that all exemplars in this experiment were placed at regular intervals through the volume. As discussed previously, better results should be obtainable using intelligent selection of exemplar location.



(a)



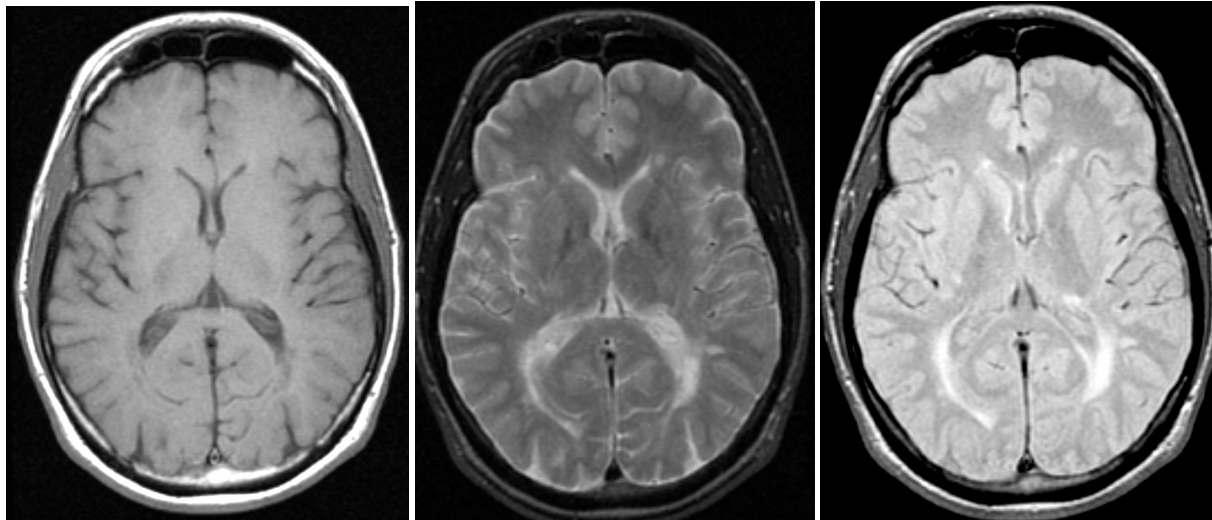
(b)

**Figure 6:** (a) Plot showing the decrease in differently classified lumen voxels ((false positives + false negatives)/ (total lumen voxels)) with increasing numbers of exemplars. Note that the result is statistically indistinguishable from a manual measurement at approximately 1 exemplar for every 4 slices. (b) Plot showing the decrease in differently classified thrombus voxels with increasing numbers of exemplars. Note that the result is statistically indistinguishable from a manual measurement at approximately 1 exemplar for every 12 slices.

## 5. DISCUSSION

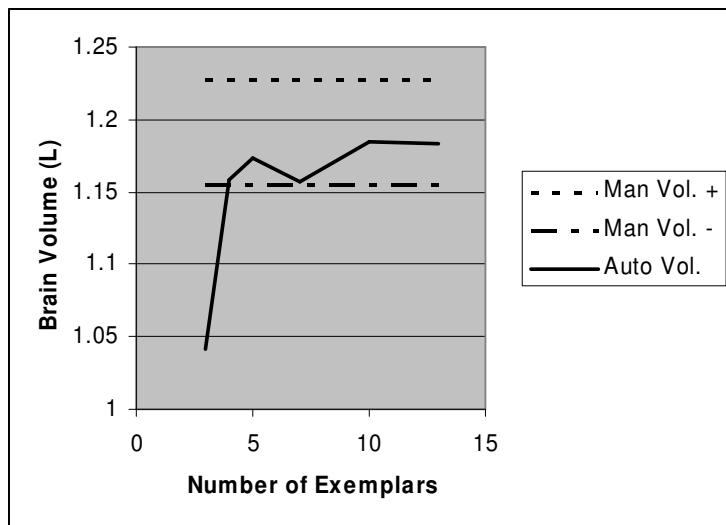
The results presented in the previous section raise a number of interesting questions. The first relates to the general applicability of the technique described here. Both applications presented previously involve single-modality data and structures that are generally much longer in the out-of-plane dimension than in the in-plane dimensions. In principle, however, this technique should work equally well on ovoid or irregular structures, and should be more effective using multi-spectral or multi-modality data sets due to the increased statistical separation between tissue classes. In order to test these hypotheses, our system was applied to the problem of total brain volume measurement in a three pulse sequence MRI head scan. This data set consisted of T1, T2 and proton density weighted MRI scans of a multiple sclerosis patient. This data set was originally collected in 1997 for use in an unrelated study and was provided to our

group without patient identification information. Slice thickness was 5mm with a slice center spacing of 5mm and a 256x192 matrix. In-plane resolution in these scans was 0.86mm. Sample images from this data set are shown in Figure 7.



**Figure 7:** T1, T2, and proton density weighted axial head scans of a multiple sclerosis patient. Total brain volume is an important biomarker used in evaluation of progression and response to treatment in multiple sclerosis.

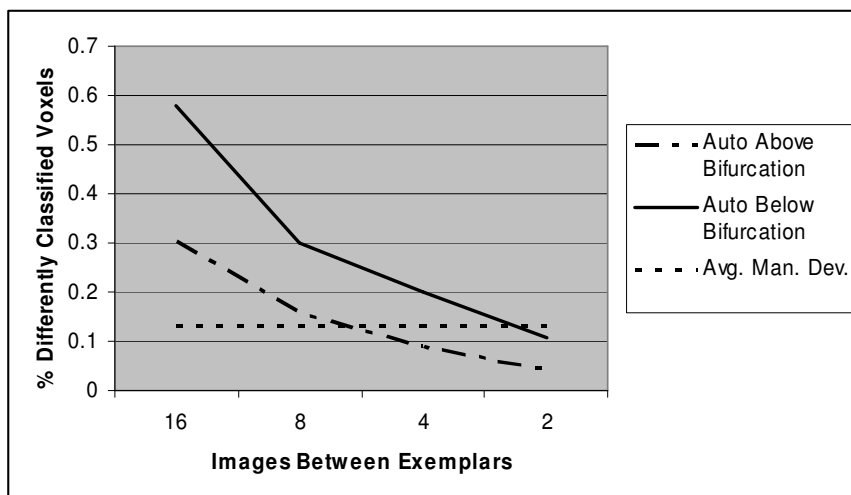
In this case, manual tracing yielded a total brain volume of 1.191L, with a coefficient of variability in volume measurement of 3%. This variability is somewhat lower than that seen in the previous experiments, as expected due to the more clearly defined structure boundaries. Results using the analysis system presented previously are given in Figure 8.



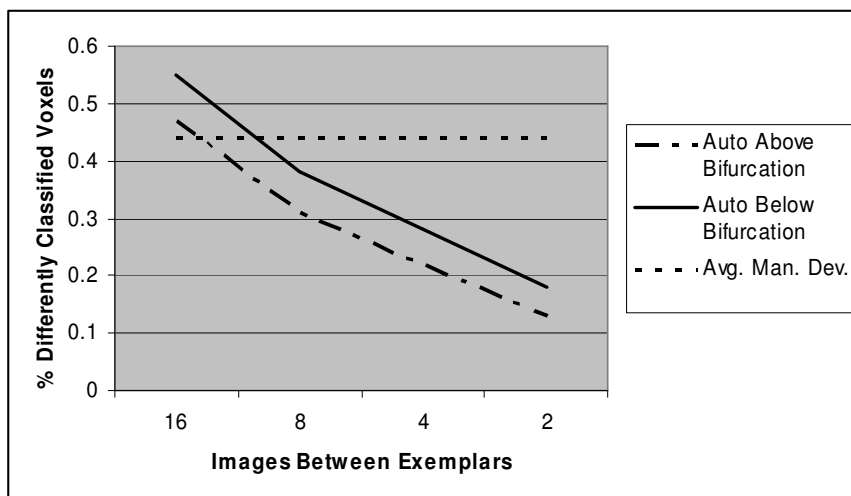
**Figure 8:** Application of the automated analysis system to calculation of total brain volume from a multi-spectral MRI head scan. In this case the brain spanned 25 images, so near-optimal results are obtained with one exemplar for every five images.

It is interesting to note that despite the differences in structure morphology and data characteristics, these results are quite consistent with those reported in the previous section. We consider this to be a good indicator for the broad applicability of this technique.

A second interesting question relates to the dependency of this analysis system on the regularity of the structures being measured. We have asserted previously that a higher density of exemplars is required in regions where the structure is complex or changes rapidly, while a lower density is required in regions where the structure is reasonably regular and does not change as rapidly. We tested this assertion by re-analyzing the abdominal aortic aneurysm data, keeping the images above and below the aortic bifurcation separate. Because the lumen is relatively regular above the aortic bifurcation, but more tortuous below, we expect to see significantly better performance above the bifurcation. The thrombus, however, is largely disjoint and has a complex shape both above and below the bifurcation. We therefore expect to see a smaller difference in thrombus performance. Results for this experiment are given in Figure 9.



(a)



(b)

**Figure 9:** (a) Classifier performance on lumen, above and below the aortic bifurcation. As expected, performance above the bifurcation is significantly better. (b) Classifier performance on thrombus, above and below the aortic bifurcation. Similar results are seen here, although the difference is smaller due to the fact that the thrombus is less regular overall.

As expected, these results show that performance is strongly influenced by the regularity of the structure being analyzed. This in turn indicates that system performance may be optimized through intelligent selection of exemplar placement, with exemplars placed frequently in rapidly changing regions and less so in stable regions.

Future work in this area will include system optimization and testing using other data types, including PET-CT and dynamic contrast enhanced CT and MRI.

## REFERENCES

1. M. de Bruijne, W. Niessen, J. Maintz, M. Viergever, "Localization and segmentation of aortic endografts using marker detection," *IEEE Trans. Medical Imaging* **22**(4), pp. 473 – 482, 2003
2. E. Ashton, K. Parker, M. Berg, C. Chen, "A novel volumetric feature extraction technique, with applications to MR images," *IEEE Trans. Medical Imaging* **16**(4), pp. 365 – 371, 1997.
3. Y. Hsu, N. Schuff *et al.*, "Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia," *Journal of MRI* **16**, pp. 305 – 310, 2002.
4. E. Ashton, C. Takahashi, M. Berg, A. Goodman, S. Totterman, S. Ekholm, "Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI," *Journal of MRI* **17**, pp. 300 – 308, 2003.
5. D. Taylor, W. Barrett, "Image segmentation using globally optimum growth in three dimensions with an adaptive feature set." *Visualization in Biomedical Computing 1994*, pp. 98 – 107, 1994.
6. I. Carlbon, D. Terzopoulos, K. Harris, "Computer assisted registration, segmentation and 3D reconstruction from images of neuronal tissue sections," *IEEE Trans. Med. Imaging*, pp. 351 – 362, 1994
7. L. Cohen, "On active contour models and balloons," *CVGIP: Graphical Models Image Processing*, pp. 211 – 218, 1991.
8. R. Chung, C. Ho, "3-D reconstruction from tomographic data using 2-D active contours," *Computers and Biomedical Research*, pp. 186 – 210, 2000.
9. Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, R. Kikinis, "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Medical Image Analysis*, pp. 143 – 168, 1998.
10. S. Aylward, E. Bullitt, "Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction," *IEEE Trans. Med. Imaging*, pp. 61 – 75, 2002.
11. K. Krissian, G. Malandain, N. Ayache, R. Vaillant, Y. Troussel, "Model-based detection of tubular structures in 3-D images," *Comput. Vis. Image Understanding*, pp. 130 – 171, 2000.
12. O. Wink, W. Niessen, M. Viergever, "Fast delineation and visualization of vessels in 3-D angiographic images," *IEEE Trans. Med. Imaging*, pp. 337 – 346, 2000.
13. B. Verdonck, I. Bloch, H. Maitre, D. Vandermeulen, P. Suentens, G. Marchal, "Accurate segmentation of blood vessels from 3D medical images," *IEEE Int. Conf. Image Processing*, pp. 311 – 314, 1996.
14. M. Fiebich, M. Tomiak, R. Engelmann, J. McGilland, K. Hoffman, "Computer assisted diagnosis in CT angiography of abdominal aortic aneurysms," *Proceedings of SPIE vol. 3034*, pp. 86 – 94, 1997.
15. A. Bulpitt, E. Berry, "Spiral CT of abdominal aneurysms: comparison of segmentation with an automatic 3D deformable model and interactive segmentation," in *Proceedings of SPIE vol. 3338*, pp. 938 – 946, 1998.
16. E. Ashton, L. Molinelli, S. Totterman, K. Parker, "Evaluation of reproducibility for manual and semi-automated feature extraction in CT and MR images," *Proceedings of IEEE International Conf. on Image Processing*, 2002.
17. J. Besag., "On the statistical analysis of dirty pictures," *J. Royal Stat. Soc.* **48**, pp. 259 – 302, 1986.
18. N. Metropolis, A. Rosenbluth, A. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.* **21**, 1087 – 1091, 1953.