User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability

Paul A. Yushkevich [a,*] Joseph Piven [c] Heather Cody Hazlett [c]

Rachel Gimpel Smith [c] Sean Ho [b] James C. Gee [a]

Guido Gerig [b,c]

[a] *Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania*

[b] *Departments of Computer Science and Psychiatry, University of North Carolina*

[c] *Neurodevelopmental Disorders Research Center, University of North Carolina*

Abstract

Active contour segmentation and its robust implementation using level set methods are well established theoretical approaches that have been studied thoroughly in the image analysis literature. Despite the existence of these powerful segmentation methods, the needs of clinical research continue to be fulfilled, to a large extent, using slice-by-slice manual tracing. To bridge the gap between methodological advances and clinical routine, we developed an open source application called ITK-SNAP, which is intended to make level set segmentation easily accessible to a wide range of users, including those with little or no mathematical expertise. This paper describes the methods and software engineering philosophy behind this new tool and provides the results of validation experiments performed in the context of an ongoing child autism neuroimaging study. The validation establishes SNAP intra/interrater reliability and overlap error statistics for the caudate nucleus and finds that SNAP is a highly reliable and efficient alternative to manual tracing. Analogous results for lateral ventricle segmentation are provided.

Key words: Computational Anatomy, Image Segmentation, Caudate Nucleus, 3D Active Contour Models, Open Source Software, Validation, Anatomical Objects

* Corresponding author. Address: 3600 Market St., Ste 320, Philadelphia, PA 19104, USA

   Email address: `pauly2@grasp.upenn.edu` (Paul A. Yushkevich).

1 Introduction

Segmentation of anatomical structures in medical images is a fundamental task in neuroimaging research. Segmentation is used to measure the size and shape of brain structures, to guide spatial normalization of anatomy between individuals and to plan medical intervention. Segmentation serves as an essential element in a great number of morphometry studies that test various hypotheses about the pathology and pathophysiology of neurological disorders. The spectrum of available segmentation approaches is broad, ranging from manual outlining of structures in 2D cross-sections to cutting-edge methods that use deformable registration to find optimal correspondences between 3D images and a labeled atlas (Haller et al., 1997; Goldszal et al., 1998). Amid this spectrum lie semiautomatic approaches that combine the efficiency and repeatability of automatic segmentation with the sound judgement that can only come from human expertise. One class of semiautomatic methods formulates the problem of segmentation in terms of active contour evolution (Zhu and Yuille, 1996; Caselles et al., 1997; Sethian, 1999), where the human expert must specify the initial contour, balance the various forces which act upon it, as well as monitor the evolution.

Despite the fact that a large number of fully automatic and semiautomatic segmentation methods has been described in the literature, many brain research laboratories continue to use manual delineation as the technique of choice for image segmentation. Reluctance to embrace the fully automatic approach may be due to the concerns about its insufficient reliability in cases where the target anatomy may differ from the norm, as well as due to high computational demands of the approach based on image registration. However, the

slow spread of semiautomatic segmentation may simply be due to the lack of readily available simple user interfaces. Semiautomatic methods require the user to specify various parameters, whose values tend to make sense only in the context of the method's mathematical formulation. We suspect that insufficient attention to developing tools that make parameter selection intuitive has prevented semiautomatic methods from replacing manual delineation as the tool of choice in the clinical research environment.

ITK-SNAP is a software application that brings active contour segmentation to the fingertips of clinical researchers. Our goals in developing this tool were (1) to focus specifically on the problem of segmenting anatomical structures, not allowing the kind of *feature creep* which would make the tool's learning curve prohibitively steep; (2) to construct a friendly and well documented user interface that would break up the task of initialization and parameter selection into a series of intuitive steps; (3) to provide an integrated toolbox for manual postprocessing of segmentation results; and (4) to make the tool freely accessible and readily available through the open source mechanism. SNAP is a product of over six years of development in academic and corporate environments and it is the largest end-user application bundled with the Insight Toolkit (ITK), a popular library of image analysis algorithms funded under the Visible Human Project by the U.S. National Library of Medicine (Ibanez et al., 2003). SNAP is available free of charge both as a stand-alone application that can be installed and executed quickly and as source code that can be used to derive new software.[1]

---

[1] SNAP binaries are available for download at `www.itksnap.org`; source code is managed at `www.itk.org`.

This paper provides a brief overview of the methods implemented in SNAP and describes the tool's core functionality. However, the paper's main focus is on the validation study, which we performed in order to demonstrate that SNAP is a viable alternative to manual segmentation. The validation was performed in the context of caudate nucleus segmentation in an ongoing child autism MRI study. Each caudate was segmented using both methods in multiple subjects by multiple highly trained raters and with multiple repetitions. The results of volume and overlap-based reliability analysis indicate that SNAP segmentation is very accurate, exceeding manual delineation in terms of efficiency and repeatability. We also demonstrate high reliability of SNAP in lateral ventricle segmentation.

The remainder of the paper is organized as follows. A short overview of automatic image segmentation, as well as some popular medical imaging tools that support it, is given in Sec. 2. A brief summary of active contour segmentation and level set methods appears in Sec. 3.1. Sec. 3.2 highlights the main features of SNAP's user interface and software architecture. Validation in the context of caudate and ventricle segmentation is presented in Sec. 4. Finally, Sec. 5 discusses the challenges of developing open-source image processing software, notes the limitations of SNAP segmentation and brings up the need for complimentary tools, which we plan to develop in the future.

2   Previous Work

In many clinical laboratories, biomedical image segmentation involves having a trained expert delineate the boundaries of anatomical structures in consecutive slices of 3D images. Although this approach puts the expert in full

5

control of the segmentation, it is time-consuming as well as error-prone. In the absence of feedback in 3D, contours traced in subsequent slices may become mismatched, resulting in unnatural jagged edges that pose a difficulty to applications such as shape analysis. Studies have demonstrated a frequent occurrence of significant discrepancies between the delineations produced by different experts as well as between repeated attempts by a single expert. For instance, a validation of caudate nucleus segmentation by Gurleyik and Haacke (2002) reports the interrater reliability of 0.84. Other studies report higher interrater reliability for the caudate, such as 0.86 in Naismith et al. (2002), 0.94 in Keshavan et al. (1998), 0.955 in Hokama et al. (1995) and 0.98 in Levitt et al. (2002). The reliability of lateral ventricle segmentation tends to be high, with Blatter et al. (1995), for example, reporting intraclass correlations of 0.99.

On the other side of the segmentation spectrum lie fully automated methods based on probabilistic models of image intensity, atlas deformation and statistical shape models. Intensity-based methods assign tissue classes to image voxels (Wells III et al., 1995; Alsabti et al., 1998; Van Leemput et al., 1999a,b) with high accuracy, but they can not identify the individual organs and anatomical regions within each tissue class. Methods based on elastic and fluid registration can identify anatomical structures in the brain by deforming a labeled probabilistic brain atlas onto the subject brain (Joshi and Miller, 2000; Avants et al., 2005; Davatzikos et al., 2001; Thirion, 1996). This type of registration assumes one-to-one correspondence between subject anatomies, which is not always the case, considering high variability in cortical folding and potential presence of pathology. In full-brain registration, small structures may be poorly aligned because they contribute a small portion to the overall

6

objective function optimized by the registration. When augmented by expert-defined landmarks, registration methods can achieve very high accuracy in structures like the hippocampus (Haller et al., 1997) but, without effective low cost software tools, they may lose their fully automatic appeal. Methods based on registration are also very computationally intensive, which may discourage their routine use in the clinical environment. Yet another class of *deformable template* segmentation methods uses a statistical model of shape and intensity to identify individual anatomical structures (Cootes et al., 1998; Joshi et al., 2002; Davies et al., 2002). The statistical prior model allows these methods to identify structure boundaries in absence of edges of intensity. However, shape priors must be learned from training sets that require a significant independent segmentation effort, which could benefit from a tool like SNAP.

In the field of biomedical image analysis software, SNAP stands out as a full-featured tool that is specifically devoted to segmentation. A number of other software packages provide semi-automatic segmentation capability, but these packages tend to be either very broad or very specific in the scope of functionality that they provide. For instance, large-scale packages such as Mayo Analyze (Robb and Hanson, 1995) and the open-source 3D Slicer (Gering et al., 2001) include 3D active contour segmentation modules, and the NIH MIPAV tool (McAuliffe et al., 2001) provides in-slice active contour segmentation. These tools carry a steep learning curve, due the large number of features that they provide. More specific tools include GIST (Lefohn et al., 2003), which has a very fast level set implementation but a limited user interface. In contrast, SNAP is both easy to learn, since it is streamlined towards one specific task, and powerful, including a full set of complimentary editing tools and a user interface that provides live feedback mechanisms intended to make parameter

7

selection easier for non-expert users.

3   Materials and Methods

3.1   Active Contour Evolution

SNAP implements two well-known 3D active contour segmentation methods: *Geodesic Active Contours* by Caselles et al. (1993, 1997) and *Region Competition* by Zhu and Yuille (1996). In both methods, the evolving estimate of the structure of interest is represented by one or more contours. An evolving contour is a closed surface $C(u, v; t)$ parameterized by variables u, v and by the time variable t. The contour evolves according to the following partial differential equation (PDE):

$$\frac{\partial}{\partial t} C(t, u, v) = F \vec{\mathbf{N}} , \tag{1}$$

where $\vec{\mathbf{N}}$ is the unit normal to the contour and F represents the sum of various forces that act on the contour in the normal direction. These forces are characterized as *internal* and *external*: internal forces are derived from the contour's geometry, and are used to impose regularity constraints on the shape of the contour, while external forces incorporate information from the image being segmented. Active contour methods differ by the way they define internal and external forces. Caselles et al. (1997) derive the external force from the gradient magnitude of image intensity, while Zhu and Yuille (1996) base it on voxel probability maps. Mean curvature of C is used to define the internal force in both methods.

8

In the Caselles et al. method, the force acting on the contour has the form

$$F = \alpha g_I + \beta \kappa g_I + \gamma (\nabla g_I \cdot \vec{\mathbf{N}}),$$ (2)

where g_I is the *speed function* derived from the gradient magnitude of the input image I, κ is the mean curvature of the contour, and α, β, γ are weights that modulate the relative contribution of the three components of F. The speed function must take values close to 0 at edges of intensity in the input image, while taking values close to 1 in regions where intensity is nearly constant. In SNAP, the speed function is defined as follows:

$$g_I(\mathbf{x}) = \frac{1}{1 + (\mathrm{NGM}_I(\mathbf{x})/\nu)^\lambda} \quad \mathrm{NGM}_I(\mathbf{x}) = \frac{\|\nabla (G_\sigma * I)\|}{\max_I \|\nabla (G_\sigma * I)\|}$$ (3)

where NGM_I is the normalized gradient magnitude of I; $G_\sigma * I$ denotes convolution of I with the isotropic Gaussian kernel with aperture σ; and ν and λ are user-supplied parameters that determine the shape of the monotonic mapping between the normalized gradient magnitude and the speed function, illustrated in Fig. 1. Note that since the speed function is non-negative, the first term in (2) acts in the outward direction, causing the contour to expand. This outward external force is counterbalanced by the so-called *advection force* $\gamma(\nabla g_I \cdot \vec{\mathbf{N}})$, which acts inwards when the contour approaches an edge of intensity to which it is locally parallel. An illustration of Caselles et al. (1997) evolution in 2D with and without the advection force, is given in Fig. 2.

Zhu and Yuille (1996) compute the external force by estimating the probability that a voxel belongs to the structure of interest and the probability that it belongs to the background at each voxel in the input image. In SNAP, these probabilities are estimated using fuzzy thresholds, as illustrated in Fig. 3. Alternatively, SNAP users can import tissue probability maps generated by

atlas-based and histogram-based tissue class segmentation programs. In the SNAP implementation, the external force is proportional to the difference of object and background probabilities, and the total force is given by

$$F = \alpha(P_{\mathrm{obj}} - P_{\mathrm{bg}}) + \beta\kappa \,. \tag{4}$$

This is a slight deviation from Zhu and Yuille (1996), who compute the external force by taking the difference between logarithms of the two probabilities. An example of contour evolution using region competition is shown in Fig. 4. Region competition is more appropriate when the structure of interest has a well-defined intensity range with respect to the image background. In contrast, the Caselles et al. (1997) method is well suited for structures bounded by strong image intensity edges.

Active contour methods typically solve the contour evolution equation using the *Level Set Method* (Osher and Sethian, 1988; Sethian, 1999). This approach ensures numerical stability and allows the contour to change topology. The contour is represented as the zeroth level set of some function ϕ, which is defined at every voxel in the input image. The relationship $\vec{\mathbf{N}} = \nabla\phi/\|\nabla\phi\|$ is used to rewrite the PDE (1) as a PDE in ϕ:

$$\frac{\partial}{\partial t}\phi(\mathbf{x};t) = F|\nabla\phi| \,. \tag{5}$$

Typically, such equations are not solved over the entire domain of ϕ but only at the set of voxels close to the contour $\phi = 0$. SNAP uses the highly efficient *Extreme Narrow Banding Method* by Whitaker (1998) to solve (5).

3.2 Software Architecture

This section describes SNAP functionality and highlights some of the more innovative elements of its user interface architecture. SNAP was designed to provide a tight but complete set of features that focus on active contour segmentation. It includes tools for viewing and navigating 3D images, manual labeling of regions of interest, combining multiple segmentation results, and post-processing them in 2D and 3D. Built on the ITK backbone, SNAP can read and write many image formats, and new features can be added easily.

3.2.1 Image Navigation and Manual Segmentation

SNAP's user interface emphasizes the 3D nature of medical images. As shown in Fig. 5a, the main window is divided into four panels, three displaying orthogonal cross-sections of the input image and the fourth displaying a 3D view of the segmented structures. Navigation is aided by a linked 3D cursor whose logical location is at the point where the three orthogonal planes intersect. The cursor can be repositioned in each slice view by mouse motion, causing different slices to be shown in the remaining two slice views. The cursor can also be moved out of the slice plane using the mouse wheel. This design ensures that the user is always presented with the maximum amount of detail about the voxel under the cursor and its neighborhood, while minimizing the amount of mouse motion needed to navigate through the image. Users can change the zoom in each slice view, and each of the slice views can be expanded to occupy the entire program window, as illustrated in Fig. 5b.

SNAP provides tools for manual tracing of regions of interest. Internally,

SNAP represents labeled regions using an integer-valued 3D mask image of the same dimensions as the input image. Each voxel in the input image can be assigned a single label, or no label at all. This approach has the disadvantage that partial volume segmentations can not be represented, but it allows users to view and edit regions simultaneously in the three slice views and in the 3D view. To define a region, the user makes a series of clicks, forming a closed contour, which can then be edited by moving, inserting and removing vertices. When integrating the contour with the 3D mask, the user assigns a label to the contour, and can choose to apply the update only to a subset of the labels that are already present in the mask, e.g., applying only to unlabeled voxels or to voxels with another label. The manual delineation mode can be used both to initialize automatic segmentation and to postprocess the results.

The 3D view is used to render the boundaries of segmented structures that are computed using the contour extraction, decimation and fairing algorithms in the Visualization Toolkit (VTK) (Lorensen and Cline, 1987; Schroeder et al., 1996). The user can reposition the 3D cursor by clicking on one of these surfaces. An arbitrary cut-plane can be defined in the 3D window, and labels on one side of the cut-plane can be replaced with another label. The cut-plane tool makes it easy to divide a segmentation into regions and to cut away extraneous tissue from a result that includes voxels outside of the structure of interest. A sequence of cut-plane operations can be used to perform complex editing operations in 3D, as illustrated in Fig. 6, where an automatic segmentation result is partitioned into the lateral ventricles and the third ventricle.

Other notable features of SNAP include the image input wizard, which allows users to read a number of recognized image file formats, includes specialized dialogs for DICOM series and raw data, and provides a graphical user interface

12

for specifying the mapping between image and anatomical coordinate systems. Once an image has been loaded using this wizard, the settings associated with the image are stored so that in the future it can be loaded without user interaction. SNAP offers linear and spline-based intensity windowing. SNAP is highly customizable, allowing users to reconfigure the arrangement and orientation of the slice views, choosing between radiological and neurological conventions; the users can change the appearance of various display elements in order to make them more or less prominent for presentation purposes.

3.2.2 Automatic Segmentation Workflow

The outcome of active contour segmentation depends on a number of parameters, including the choice of method, the way in which the input image is converted into a probability map or speed function, the initial contour, and the weights assigned to various internal and external forces that drive contour evolution. Even for users familiar with level sets, finding the right set of parameters can be difficult. To simplify this task, SNAP organizes parameter specification into a wizard-like workflow and relies extensively on live feedback mechanisms. The workflow is divided into three logical stages.

In the first stage, the user chooses between Zhu and Yuille (1996) and Caselles et al. (1997) methods and, depending on the method chosen, computes the probability map $P_{\mathrm{obj}} - P_{\mathrm{bg}}$ or the speed function g_I. Probability maps are computed, as shown in Fig. 7a, by applying a smooth threshold, which can be one-sided or two-sided, depending on whether the intensity range of the structure of interest lies at one of the ends or in the middle of the histogram. Taking advantage of the flexible ITK architecture that can apply image processing fil-

13

ters to arbitrary image subregions, the orthogonal slice views in SNAP provide immediate feedback in response to changes in parameter values. Alternatively, the user may import an external image, such as a tissue class segmentation, as the probability map or speed image.

In the second stage, the user initializes the segmentation by placing one or more spherical 'seeds' in the image (Fig. 7b). The user can also initialize the active contour with a result of an earlier manual or automatic segmentation. Level set methods allow contours to change topology, and it is common to place several seeds within one structure, letting them merge into a single contour over the course of evolution.

The last stage of the segmentation workflow is devoted to specifying the weights of the various terms in the active contour evolution PDE and running the evolution interactively. In order to accommodate a wider range of users, SNAP provides two separate modes for choosing weights. In the casual user mode, weights α, β and γ from the active contour equations are described verbally in terms of their impact on the behavior of the evolving contour, accompanied by an interactive, dynamically updated 2D curve evolution illustration that shows the effect of each of the parameters on the total force acting on the interface (Fig. 8). The other mode is for users familiar with the mathematics of active contours, and allows them to specify the weights in a generic evolution formulation that incorporates both (2) and (4):

$$F = \alpha h^a - \beta h^b \kappa - \gamma h^c \left(\nabla h \cdot \vec{\mathbf{N}} \right) \ . \tag{6}$$

By setting $h = g_I, a = 1, b = 1, c = 0$, the user arrives at the Caselles et al. (1997) formulation, and with $h = P_{\text{obj}} - P_{\text{bg}}, a = 1, b = 0, c = 0$, the Zhu and Yuille (1996) formulation is obtained. Formulation (6) corresponds to SNAP's

internal representation of the evolution equation.

The actual contour evolution is controlled by a VCR-like interface. When the user clicks the 'play' button, the contour begins to evolve, and the slice views and, optionally, the 3D view are updated after each iteration. The user can use 'stop', 'rewind' and 'single step' buttons to control and terminate contour evolution. Fig. 9 shows SNAP before and and after contour evolution.

Before entering the automatic segmentation mode, the user may choose to restrict segmentation to a 3D region of interest in order to reduce computational cost and memory use. An option to resample the region of interest using nearest neighbor, linear, cubic, or windowed sinc interpolation is provided; this is recommended for images with anisotropic voxels. When returning from automatic mode to the manual mode, SNAP converts the level set segmentation result to a binary mask and merges it with the 3D mask image. In the process, the sub-voxel accuracy of the segmentation is compromised. Before merging, the user has the option to export the segmentation as a real-valued image.

4   Results

The new SNAP tool, with its combination of user-guided 3D active contour segmentation and post-processing via manual tracing in orthogonal slices or using the 3D cut-plane tool, is increasingly replacing conventional 2D slice editing for a variety of image segmentation tasks. SNAP is used in several large neuroimaging studies at UNC Chapel Hill, Duke University and the University of Pennsylvania. Segmentation either uses the soft threshold option for the definition of foreground and background, e.g. for the segmentation of the caudate

15

nucleus in head MRI, or employs existing tissue probability maps that define object to background probabilities. This option is used for the segmentation of ventricles based on cerebrospinal fluid probabilistic segmentations.

An ongoing child neuroimaging autism study serves as a testbed for validation of the new tool as a prerequisite to its use in a large clinical study. In particular, we chose the segmentation of the caudate nucleus to establish intrarater and interrater reliability of applying SNAP and also to test validity in comparison to manual rater segmentation. The following sections describe validation of SNAP versus manual rater contour drawing in more details. In addition, we provide reliability results for the lateral ventricle segmentation in SNAP, but without a comparison to manual segmentation.

4.1 Validation of SNAP: Caudate Segmentation

From our partnership with the UNC Psychiatry department, we have access to a morphologic MRI study with a large set of autistic children (N=56), developmentally delayed subjects (N=11) and control subjects (N=17), scanned at age two. SNAP was chosen as an efficient and reliable tool to segment the caudate nucleus from high-resolution MRI. Before replacing conventional manual outlining by this new tool, we designed a validation study to test the difference between methods, the difference between operators, and the variability for each user.

4.1.1 Gray-level MRI data

Caudate segmentation uses high-resolution T1-weighted MRI with voxel size $0.78 \times 0.78 \times 1.5 \, \text{mm}^3$. The protocol established by the UNC autism image analysis group rigidly aligns these images to the Talairach coordinate space by specifying anterior and posterior commissure (AC-PC) and the interhemispheric plane. The transformation also interpolates the images to the isotropic voxel size of 1mm^3. Automatic atlas-based tissue segmentation using EMS (Van Leemput et al., 1999a,b) results in a hard segmentation and separate probability maps for white matter, gray matter and cerebrospinal fluid. These three-tissue maps are used for SNAP ventricle segmentation, but not for the caudate nucleus, because in some subjects, the intensity distribution of the subcortical gray matter is different from the cortex.

4.1.2 Reliability series and validation

Five MRI datasets were arbitrarily chosen from the whole set of 100+ images. Data were replicated three times and blinded to form a validation database of 15 images. Three highly trained raters participated in the validation study; rater A segmented each image manually and in SNAP, while rater B only used SNAP and rater C only performed manual segmentation. Segmentation results between pairs of raters or methods were analyzed using common intraclass correlation statistics (ICC) as well as using overlap statistics.

4.1.3 Caudate nucleus segmentation

At first sight, the caudate seems easy to segment since the largest fraction of its boundary is adjacent to the lateral ventricles and white matter. Portions of

the caudate boundary can be localized with standard edge detection. However, the caudate is also adjacent to the nucleus accumbens and the putamen where there are no visible boundaries in MRI (see Fig. 10). The caudate, nucleus accumbens and putamen are distinguishable on histological slides, but not on T1-weighted MRI of this resolution. Another "trouble-spot" for the caudate is where it borders the putamen; there are "fingers" of cell bridges adjacent to blood vessels which span the gap between the two.

4.1.4 Manual boundary drawing

Using the drawing tools in SNAP, we have developed a highly reliable protocol for manual caudate segmentation using slice-by-slice boundary drawing in all three orthogonal views. In addition to boundary overlays, the segmentation is supported by a 3D display of the segmented structure. The coupling of cursors between 2D slices and the 3D display help significantly reduce slice-by-slice jitter that is often seen in this type of segmentations. Segmentation time for left and right caudate is approximately 1.5 hours for experienced experts.

4.1.5 3D Active Contour Segmentation

We developed a new segmentation protocol for caudate segmentation based on the T1 gray level images with emphasis on efficiency and optimal reliability. The caudate nucleus is a sub-cortical gray matter structure. T1 intensity values of the caudate regions in the infant MRI showed significant differences among subjects and required individual adjustment. We place sample regions in four axial slices of the caudate, measure the intensity and standard deviation of the sample regions, then use those values to guide the setting of the upper

18

and lower thresholds in the pre-processing step of SNAP. This results in foreground/background maps which guide the level set evolution. Parameters for smoothness and speed were trained in a pilot study and were kept constant for the whole study. The same regions were used as initialization regions. Evolution was stopped after the caudate started to bleed into the adjacent putamen. Optimal intensity window selection and active contour evolution takes only about 5 minutes for the left and right caudate. In some caudate segmentation protocols, the inferior boundary is cut off by the selection of an axial cut plane, which only takes a few additional seconds using the cut-plane feature in SNAP. In our autism project, we decided to add a precise separation from the putamen and a masking of the left and right nucleus accumbens. This is a purely manual operation since there are no visible boundaries between caudate and nucleus accumbens in the MR image. This step added another 30 minutes to the whole process. The total segmentation time was reduced from originally 1.5 hours for slice-by-slice contour drawing to 35 minutes, with the option to be reduced to only 5 minutes if simple cut-planes for inferior boundaries would be sufficient for the given task, which is similar to the protocol applied by Levitt et al. (2002). The raters also reported that they felt much more comfortable with the SNAP tool since they could focus their effort on a small part of the boundary that is most difficult to trace.

4.1.6 Volumetric Analysis

Table 1 lists the left and right caudate volumes for manual segmentation (slice by slice contouring) and user-assisted 3D active contour segmentation (SNAP). Results of the reliability analysis using one-way random effects intraclass correlation statistics (Shrout and Fleiss, 1979) are shown in Table 2.

19

The table shows not only the excellent reliability of SNAP segmentation but also reflects the excellent reliability of the manual experts trained over several months. Therefore, reliability between methods is not significantly different. On the basis of volume comparisons, the SNAP segmentation, which requires much less training and is significantly more efficient, is shown equivalent to the manual expert, both with respect to intra-method reliability and inter-method validity. However, this is to be compared with the significantly reduced segmentation time and short rater training time of SNAP.

4.1.7 Overlap Analysis

In addition to volume-based reliability analysis, we compare SNAP and manual methods in terms of overlap between different segmentations of each instance of the caudate. Overlap is a more accurate measure of agreement between two segmentations than volume difference because the latter may be zero for two completely different segmentations. For every subject-caudate combination, we measure the overlap between all ordered pairs of available segmentations. There are 10 structures (5 subjects, left and right) and for each structure there are 12 different segmentations (2 raters, 2 methods, 3 repetitions) and $\binom{12}{2} = 66$ ordered pairs. For each pair, we count the number of voxels in the input image that belong to both segmentations. Following the statistical approach described in (Zou et al., 2004), we define the overlap between segmentations S_1 and S_2 as the Dice Similarity Coefficient (DSC):

$$\mathrm{DSC}(S_1, S_1) = \frac{2\,\mathrm{Vol}(S_1 \cap S_2)}{\mathrm{Vol}(S_1) + \mathrm{Vol}(S_2)} . \tag{7}$$

This symmetric measure of segmentation agreement lies in the range $[0, 1]$, with larger values indicating greater overlap.

20

Table 3 lists means and standard deviations of the overlaps for the left and right caudate within 13 different categories of comparisons. These categories fall into three larger classes: SNAP-to-SNAP comparisons, (i.e., pairs where both segmentations were generated with SNAP), manual-to-manual comparisons, and mixed SNAP-to-manual comparisons. Within each class, categories include (1) all pairs where both segmentations were performed by a given rater, (2) all pairs where both segmentations were performed by the same rater, and (3) all pairs where segmentations were performed by different raters. Fig. 11 displays box and whisker plots of the same 13 categories. It is immediately noticeable that the comparisons in the SNAP-to-SNAP group yield significantly better overlaps than other types of comparisons; in fact, the worst overlap between any pair of SNAP segmentations is still better than the best overlap between any pair of manual segmentations or any pair where both methods are used. This indicates that SNAP caudate segmentation exhibits significantly better repeatability than manual segmentation.

To confirm this finding quantitatively, we perform an ANOVA experiment adopted from Zou et al. (2004), who studied repeatability in the context of pre- and post-operative prostate segmentation. Variance components in the ANOVA model include the *method* ($M \in \{\text{SNAP}, \text{Manual}\}$), *case* (i.e., subject; $C \in \{1, 2, 3, 4, 5\}$), *anatomy* $A \in \{\text{l. caudate}, \text{r. caudate}\}$ and *repetition pair* $P \in \{(a,b), (b,c), (a,c)\}$, i.e., one of the three possible ordered pairs of segmentations performed by a given rater in a given case on a given structure (symbols a, b, c correspond to the order in which the rater performed the segmentations). Zou et al. (2004) pools the model over all raters, but in our case, since only rater A performed segmentation using both methods, we just include segmentations by rater A in the model. Following Zou et al. (2004),

the model includes two interaction terms: $M \times S$ and $C \times S$, and the outcome variable is derived by standardizing the DSC using the *logit* transform:

$$\text{LDSC}(S_1, S_2) = \ln \left(\frac{\text{DSC}(S_1, S_2)}{1 - \text{DSC}(S_1, S_2)} \right) . \tag{8}$$

Our ANOVA results are listed in Table 4. We conclude that SNAP segmentations are significantly more reproducible than manual segmentation ($F = 1066, p \ll 0.001$). We observe a significant effect of case on repeatability ($F = 5.176, p = 0.001$), implying that reproducibility varies by subject. There is no evidence to support the hypothesis that repeatability improves with training, as the choice of repetition pair has no significant effect on the overlap ($p = 0.287$). We also do not find a significant difference in repeatability between left and right caudates. A limitation of the above analysis is that segmentations from only one rater are included in the ANOVA. The other two raters could not be included because rater B used SNAP for all caudate segmentations, and all segmentations by rater C were manual. A stronger case could have been made if each of these raters had used both methods, allowing us to treat rater as a random effect. However, visual analysis of box plots in Fig. 11 suggests that while repeatability varies by rater within each method, this difference is smaller than the overall difference in repeatability between the methods.

4.2 Lateral Ventricle Segmentation

Unlike the caudate, which has a simple shape but lacks clearly defined MRI intensity boundaries, the lateral ventricles are complex geometrically yet have an easily identifiable boundary. To demonstrate the breadth of SNAP segmen-

tation, we present the results of a repeatability experiment for lateral ventricle segmentation. As for the caudate, we randomly selected five MRI images from the child autism database and applied our standard image processing pipeline, including tissue class segmentation using EMS. Our ventricle segmentation protocol involves placing three initialization seeds in each ventricle and running the active contour segmentation until there is no more expansion into the horns. Afterwards, the 3D cut-plane tool is used to separate the left ventricle from the right. In some cases, active contour segmentation will bleed into the third ventricle, which can be corrected using the cut-plane tool or by slice-by-slice editing. If the ventricles are very narrow, the evolving interface may not reach the inferior horns, and there also are rare cases where parts of the ventricles are so narrow that they are misclassified by EMS. These problems are corrected by post-processing, which involves reapplying active contour segmentation to the trouble regions or correcting the segmentation manually. Nevertheless, the approximate average time to segment a pair of lateral ventricles is 15 minutes, including initialization, running the automatic pipeline, reviewing and editing.

Using SNAP, each of the five selected images was segmented three times by two blinded raters; in contrast to caudate validation, manual segmentation was not performed. The volumes of the ventricle segmentations are listed in Table 5 and the volume-based reliability statistics are given in Table 6. SNAP interrater reliability coefficients exceed 0.99 for both ventricles, matching the reliability of manual segmentation, as reported by Blatter et al. (1995). Overlap statistics are summarized in Table 7 and Fig. 12. Despite the ventricles' complex shape, the average DSC values (0.989 for the left ventricle and 0.983 for the right) are higher than for caudate segmentation.

5 Discussion

The caudate segmentation validation, which compares the SNAP tool to manual segmentation by highly trained raters, demonstrates the excellent reliability of the tool for efficient three-dimensional segmentation. While the volume-based reliability analysis shows a similar range of intra-method reliability for both segmentation approaches, overlap analysis reveals that SNAP segmentation exhibits significantly improved repeatability. SNAP cut the segmentation time by a factor of three and also significantly reduced the training time to establish expert reliability. Besides repeatability and efficiency, our experts preferred using SNAP over slice contouring due to the tool's capability to display 3D segmentations in real time and due to the simple option to postprocess the automated segmentation using 3D tools.

In addition to brain structure extraction in MRI, SNAP has found a variety of uses in other imaging modalities and anatomical regions. For example, in radiation oncology applications, SNAP has proven useful for segmenting the liver, kidneys, bony structures and tumors in thin-slice computer tomography (CT) images. In emphysema research involving humans and mice, SNAP has been used to extract lung cavities in CT, as well as pulmonary vasculature in MRI. SNAP has also proven invaluable as a supporting tool for developing and validating medical image analysis methodology. It has found countless uses in our own labs, such as to postprocess the results of automatic brain extraction, to identify landmarks that guide image registration and to build anatomical atlases for template deformation morphology (Yushkevich et al., 2005).

Despite SNAP's versatility, its automatic segmentation pipeline is limited to

a specific subset of segmentation problems where the structure of interest has a different intensity distribution from most of the surrounding tissues. Future development of SNAP will focus on simplifying the segmentation of structures whose intensity distribution is indistinguishable from some of its neighbors. This will be accomplished by (1) preventing the evolving interface from entering certain regions via special seeds placed by the user, which push back on the interface, similar to the "volcanoes" in the seminal paper by Kass et al. (1988); and (2) providing additional 3D postprocessing tools that will make it easier to cut away parts of the interface that has leaked. These postprocessing tools will be based on graph-theoretic algorithms. One such tool will allow the user to trace a closed path on the surface of the segmentation result, after which the minimal surface bounded by that path will be computed and used to partition the segmentation in two. Another future feature of SNAP will be an expanded user interface for defining object and background probabilities in the Zhu and Yuille (1996) method, where the user will be able to place a number of sensors inside and outside of the structure in order to estimate the intensity distribution within the structure. In order to account for intensity inhomogeneity in MRI, SNAP will include an option to let the estimated distribution vary spatially. Finally, we plan to tightly integrate SNAP will external tools for brain extraction, tissue class segmentation and inhomogeneity field correction.

6 Conclusion

ITK-SNAP is an open source medical image processing application that fulfills a specific and pressing need of biomedical imaging research by providing

a combination of manual and semiautomatic tools for extracting structures in 3D image data of different modalities and from different anatomical regions. Designed to maximize user efficiency and to provide a smooth learning curve, the user interface is focused entirely on segmentation, parameter selection is simplified using live feedback, and the number of features unrelated to segmentation kept to a minimum. Validation in the context of caudate nucleus and lateral ventricle segmentation in child MRI demonstrates excellent reliability and high efficiency of 3D SNAP segmentation, and provides strong motivation for adopting SNAP as the segmentation solution for clinical research in neuroimaging and beyond.

Acknowledgements

26

Eric Larsen, Sanjay Sthapit, Ashraf Farrag, Amy Henderson, Robin Munesato, Ming Yu, Nathan Moon, Thorsten Scheuermann, Konstantin Bobkov, Nathan Talbert, Yongjik Kim, Pierre Fillard (UNC Chapel Hill student projects, 1999-2003, supervised by Guido Gerig), Daniel S. Fritsch and Stephen R. Aylward (ITK Integration, 2003-2004). Special thanks are extended to Terry S. Yoo, Joshua Cates, Luis Ibáñez, Julian Jomier, and Hui Zhang.

References

Alsabti, Ranka, and Singh. An efficient parallel algorithm for high dimensional similarity join. In *IPPS: 11th International Parallel Processing Symposium*. IEEE Computer Society Press, 1998.

Avants, B., Schoenemann, P., and Gee, J. Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex. *Medical Image Analysis*, 2005. In Press.

Blatter, D. D., Bigler, E. D., Gale, S. D., Johnson, S. C., Anderson, C. V., Burnett, B. M., Parker, N., Kurth, S., and Horn, S. D. Quantitative volumetric analysis of brain MR: normative database spanning 5 decades of life. *AJNR Am J Neuroradiol*, 16(2):241–51, 1995.

Caselles, V., Catte, F., Coll, T., and Dibos, F. A geometric model for active contours. *Numerische Mathematik*, 66:1–31, 1993.

Caselles, V., Kimmel, R., and Sapiro, G. Geodesic active contours. *International Journal of Computer Vision*, 22:61–79, 1997.

Cootes, T., Edwards, G., and Taylor, C. Active appearance models. In *European Conference on Computer Vision*, volume 2, pages 484–498. Freiburg, Germany, 1998.

Davatzikos, C., Genc, A., Xu, D., and Resnick, S. M. Voxel-based morphom-

etry using the ravens maps: methods and validation using simulated longitudinal atrophy. *Neuroimage*, 14(6):1361–9, 2001.

Davies, R. H., Twining, C. J., Cootes, T. F., Waterton, J. C., and Taylor, C. J. A minimum description length approach to statistical shape modeling. *IEEE Trans Med Imaging*, 21(5):525–37, 2002.

Gering, D., Nabavi, A., Kikinis, R., Hata, N., Odonnell, L., Grimson, W. E. L., Jolesz, F., Black, P., and III, W. W. An integrated visualization system for surgical planning and guidance using image fusion and an open mr. *Journal of Magnetic Resonance Imaging*, 13:967–975, 2001.

Goldszal, A., Davatzikos, C., Pham, D. L., Yan, M. X. H., Bryan, R. N., and Resnick, S. M. An image processing system for qualitative and quantitative volumetric analysis of brain images. *J. Comp Assist. Tomogr.*, 22(5):827–837, 1998.

Gurleyik, K. and Haacke, E. M. Quantification of errors in volume measurements of the caudate nucleus using magnetic resonance imaging. *Jounral of Magnetic Resonance Imaging*, 15(4):353–363, 2002.

Haller, J., Banerjee, A., Christensen, G., Gado, M., Joshi, S., Miller, M., Sheline, Y., Vannier, M., and Csernansky, J. Three-dimensional hippocampal MR morphometry by high-dimensional transformation of a neuroanatomic atlas. *Radiology*, 202:504–510, 1997.

Hokama, H., Shenton, M. E., Nestor, P. G., Kikinis, R., Levitt, J. J., Metcalf, D., Wible, C. G., O'Donnell, B. F., Jolesz, F. A., and McCarley, R. W. Caudate, putamen, and globus pallidus volume in schizophrenia: a quantitative MRI study. *Psychiatry Res*, 61(4):209–29, 1995.

Ibanez, L., Schroeder, W., Ng, L., and Cates, J. *The ITK Software Guide*. Kitware, Inc., 2003.

Joshi, S. and Miller, M. I. Landmark matching via large deformation dif-

feomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000.

Joshi, S., Pizer, S., Fletcher, P., Yushkevich, P., Thall, A., and Marron, J. Multi-scale deformable model segmentation and statistical shape analysis using medial descriptions. *IEEE Transactions on Medical Imaging*, 21(5):538–550, 2002.

Kass, M., Witkin, A., and Terzopoulos, D. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

Keshavan, M. S., Rosenberg, D., Sweeney, J. A., and Pettegrew, J. W. Decreased caudate volume in neuroleptic-naive psychotic patients. *Am J Psychiatry*, 155(6):774–8, 1998.

Lefohn, A. E., Cates, J. E., and Whitaker, R. T. Interactive, gpu-based level sets for 3d segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 564–572. 2003.

Levitt, J. J., McCarley, R. W., Dickey, C. C., Voglmaier, M. M., Niznikiewicz, M. A., Seidman, L. J., Hirayasu, Y., Ciszewski, A. A., Kikinis, R., Jolesz, F. A., and Shenton, M. E. MRI study of caudate nucleus volume and its cognitive correlates in neuroleptic-naive patients with schizotypal personality disorder. *Am J Psychiatry*, 159(7):1190–7, 2002.

Lorensen, W. E. and Cline, H. E. Marching cubes: a high resolution 3D surface construction algorithm. *Computer Graphics*, 21(4):163–169, 1987.

McAuliffe, M. J., Lalonde, F. M., McGarry, D., Gandler, W., Csaky, K., and Trus, B. L. Medical image processing, analysis & visualization in clinical research. In *CBMS '01: Proceedings of the Fourteenth IEEE Symposium on Computer-Based Medical Systems*, page 381. IEEE Computer Society, Washington, DC, USA, 2001.

Naismith, S., Hickie, I., Ward, P. B., Turner, K., Scott, E., Little, C., Mitchell,

P., Wilhelm, K., and Parker, G. Caudate nucleus volumes and genetic determinants of homocysteine metabolism in the prediction of psychomotor speed in older persons with depression. *Am J Psychiatry*, 159(12):2096–8, 2002.

Osher, S. and Sethian, J. Fronts propagating with curvature speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

Robb, R. A. and Hanson, D. P. The ANALYZE(tm) software system for visualization and analysis in surgery simulation. In S. Lavallé, R. Taylor, G. Burdea, and R. Mösges, editors, *Computer Integrated Surgery*, pages 175–190. MIT Press, 1995.

Schroeder, W. J., Martin, K. M., and Lorensen, W. E. The design and implementation of an object-oriented toolkit for 3D graphics and visualization. In R. Yagel and G. M. Nielson, editors, *IEEE Visualization '96*, pages 93–100. 1996.

Sethian, J. A. *Level set methods and fast marching methods*. Cambridge University Press, 1999.

Shrout, P. and Fleiss, J. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, 86:420–428, 1979.

Thirion, J.-P. Non-rigid matching using demons. In *CVPR '96: Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, page 245. IEEE Computer Society, 1996. ISBN 0-8186-7258-7.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18:885–896, 1999a.

Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. Automated model-based tissue classification of MR images of the brain. *IEEE Trans-*

actions on Medical Imaging, 18:897–908, 1999b.

Wells III, W. M., Grimson, W. E. L., Kikinis, R., and Jolesz, F. A. Adaptive segmentation of MRI data. In N. Ayache, editor, *Computer Vision, Virtual Reality and Robotics in Medicine*. Springer-Verlag, 1995.

Whitaker, R. T. A level-set approach to 3d reconstruction from range data. *Int. J. Comput. Vision*, 29(3):203–231, 1998. ISSN 0920-5691.

Yushkevich, P. A., Dubb, A., Xie, Z., Gur, R., Gur, R., and Gee, J. C. Regional structural characterization of the brain of schizophrenia patients. *Academic Radiology*, 12(10):1250–1261, 2005.

Zhu, S. C. and Yuille, A. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(9):884–900, 1996. ISSN 0162-8828.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., III, W. M. W., Jolesz, F. A., and Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index. *Academic Radiology*, 11(2):178–189, 2004.

Table 1: Caudate volumes (in mm^3) from the validation study comparing SNAP reliability to manual segmentation. Five test cases, replicated three times (column one) have been segmented by three raters who were blind to the cases.

| case \ rater | Right Caudate Volumes | | | | Left Caudate Volumes | | | |
| | SNAP | | Manual | | SNAP | | Manual | |
	A	B	A	C	A	B	A	C
gr1a	3720	3738	3690	3724	3734	3781	3621	3529
gr1b	3713	3780	3631	3551	3715	3826	3552	3482
gr1c	3790	3786	3735	3749	3758	3790	3521	3510
mean(gr1)	3741	3768	3685	3675	3735	3799	3565	3507
gr2a	4290	4353	4267	4120	4087	4141	4137	4262
gr2b	4229	4247	4254	4164	4083	4103	4189	4179
gr2c	4233	4232	4303	4253	4025	4143	4168	4149
mean(gr2)	4251	4277	4275	4179	4065	4129	4165	4196

Continued from the last page ...

		Right Caudate Volumes				Left Caudate Volumes			
		SNAP		Manual		SNAP		Manual	
case	rater	A	B	A	C	A	B	A	C
gr3a		4211	4250	4263	4397	4590	4495	4416	4482
gr3b		4289	4155	4221	4149	4562	4506	4444	4417
gr3c		4264	4257	4354	4174	4416	4583	4323	4332
mean(gr3)		4255	4221	4279	4240	4523	4528	4394	4410
gr4a		4091	4105	4063	4122	3967	4129	4006	4066
gr4b		4151	4150	4144	4116	4058	4135	3934	4029
gr4c		4081	4149	4103	4037	4081	4141	4001	3995
mean(gr4)		4108	4134	4103	4092	4036	4135	3980	4030
gr5a		4112	4197	4167	4125	4355	4295	4278	4321
gr5b		4165	4226	4143	4039	4326	4273	4253	4317
gr5c		4191	4237	4089	4226	4311	4288	4127	4087
mean(gr5)		4156	4220	4133	4130	4330	4285	4219	4242

Table 2: Intrarater and interrater reliability of caudate segmentation. Reliability was measured based on 3 replications of 5 test datasets by two raters. Reliability values for (1) manual segmentation by two experts; (2) manual vs. SNAP segmentation by the same expert; and (3) SNAP segmentation by two experts show the excellent reliability of both methods and the excellent agreement between manual expert's segmentation and SNAP. SNAP reduced segmentation time from 1.5 hours to 30 minutes, while the training period to establish reliability was several months for the manual method and significantly shorter for SNAP.

Validation type	Side	Intra A	Intra B	Intra AB	Inter AB
Intra/interrater Manual	Right	0.963	0.845	0.902	0.916
(A Manual, B Manual)	Left	0.970	0.954	0.961	0.967
Manual vs. SNAP	Right	0.963	0.967	0.964	0.967
(A Manual, B SNAP)	Left	0.970	0.969	0.969	0.907
Intra/interrater SNAP	Right	0.967	0.958	0.962	0.958
(A SNAP, B SNAP)	Left	0.969	0.990	0.978	0.961

Table 3: Overlap statistics between pairs of caudate segmentations, categorized by different methods and raters. Letters A, B and C refer to individual raters. Overlap values are given in percent, i.e., overlap$(S_1, S_2) = 100\% * \text{DSC}(S_1, S_2)$. The number of pairs in each category is given in the last column (n).

	Category	μ_{Left}	σ_{Left}	μ_{Right}	σ_{Right}	n
1.	SNAP A vs. SNAP A	97.5	0.836	97.7	0.598	15
2.	SNAP B vs. SNAP B	98.8	0.291	98.7	0.679	15
3.	SNAP intrarater average	98.1	0.931	98.2	0.823	30
4.	SNAP A vs. SNAP B	97.3	0.810	97.5	0.686	45
5.	Manual A vs. Manual A	94.0	0.802	94.2	0.617	15
6.	Manual C vs. Manual C	93.4	0.568	93.1	0.504	15
7.	Manual intrarater average	93.7	0.752	93.6	0.783	30
8.	Manual A vs. Manual C	92.5	0.913	91.9	0.697	45

Continued from the last page ...

Category	μ_{Left}	σ_{Left}	μ_{Right}	σ_{Right}	n
9. SNAP A vs. Manual A	92.2	0.991	92.6	0.515	45
10. SNAP A vs. Manual C	91.3	0.602	91.0	0.893	45
11. SNAP B vs. Manual A	91.9	1.11	92.5	0.421	45
12. SNAP B vs. Manual C	91.2	0.635	91.1	0.733	45
13. SNAP vs. Manual interrater average	91.5	0.860	91.5	0.999	135

Table 4: Results of ANOVA experiment to determine whether segmentation repeatability measured in terms of overlap varies according to method (M), rater (R), case (C), anatomical structure (A) or the pair of segmentations involved in the overlap computation (P).

Variance Comp.	D.o.F.	Sum Sq.	Mean Sq.	F-Stat.	P-Value
Method (M)	1	13.886	13.886	220.639	0.
Segmentation pair (P)	2	0.027	0.013	0.214	0.808
Case (C)	4	0.759	0.190	3.015	0.029
Anat. structure (A)	1	0.033	0.033	0.520	0.475
$M \times P$	2	0.117	0.059	0.931	0.402
$C \times P$	8	0.167	0.021	0.337	0.947
Residuals	41	2.580	0.063	-	-

Table 5: Ventricle volumes (in mm^3) from the SNAP reliability experiment. Five test cases, replicated three times (column one) have been segmented by two raters (A, B), who were blinded to the cases.

	Left		Right			Left		Right	
Case	A	B	A	B	Case	A	B	A	B
gr1a	2935	2942	3375	3389	gr4a	1605	1581	1719	1752
gr1b	2954	2953	3363	3404	gr4b	1578	1605	1719	1725
gr1c	2955	2936	3380	3386	gr4c	1606	1607	1719	1725
mean(gr1)	2948	2944	3373	3393	mean(gr4)	1596	1598	1719	1734
gr2a	5565	5572	6833	6824	gr5a	3687	3562	7776	7436
gr2b	5561	5579	6825	6830	gr5b	4178	3511	7725	7541
gr2c	5564	5577	6830	6824	gr5c	3871	3561	7758	7389
mean(gr2)	5563	5576	6829	6826	mean(gr5)	3912	3544	7753	7455
gr3a	2775	2781	6778	6963					
gr3b	2682	2780	7307	6971					
gr3c	2734	2772	7217	6982					
mean(gr3)	2730	2778	7101	6972					

Table 6: Intrarater and interrater reliability of lateral ventricle segmentation in SNAP. Reliability was measured based on 3 replications of 5 test datasets by two raters.

Validation type	Side	Intra A	Intra B	Intra AB	Inter AB
Intra/interrater SNAP	Right	0.9942	0.9999	0.9970	0.9917
(A SNAP, B SNAP)	Left	0.9977	0.9998	0.9987	0.9976

Table 7: Overlap statistics for left and right lateral ventricle segmentations. Overlap values are given as percentages, i.e., $\text{overlap}(S_1, S_2) = 100\% * \text{DSC}(S_1, S_2)$.

Category	μ_{Left}	σ_{Left}	μ_{Right}	σ_{Right}	n
1. SNAP A vs. SNAP A	99.5	0.481	99.5	0.482	15
2. SNAP B vs. SNAP B	99.0	1.16	98.6	1.69	15
3. SNAP intrarater average	99.3	0.898	99.1	1.31	30
4. SNAP A vs. SNAP B	98.9	0.914	98.3	2.20	45

Figure Captions

Fig. 1. An illustration of the parameters κ and λ that determine the shape of the function g that inversely maps the values of image gradient magnitude to the range $(0, 1]$. The top row shows the shapes of the mapping function under different values of the parameters, and the bottom row shows the resulting feature images.

Fig. 2. Examples of edge-based contour evolution before (top) and after (bottom) adding the advection term. Without advection, the contour leaks past the boundaries of the caudate nucleus because the external force is non-negative.

Fig. 3. The plot on the left gives an example of three smooth threshold functions with different values of the smoothness parameter κ. To the right of the plot are the input grayscale image and the feature images corresponding to the three thresholds.

Fig. 4. Active contour evolution using the feature image based on region competition. The propagation force acts outwards over the 'foreground' region (red) and inwards over the 'background' region (blue), causing the active contour to reach equilibrium at the boundary of the regions.

Fig. 5. **a.** SNAP user interface shows three orthogonal views of a volumetric image, linked by a common cursor (light blue crosshairs). A fourth panel is used to view the segmented structures in three dimensions. **b.** Alternatively, SNAP can be focused on a single slice, and zoom facilities are provided for

segmenting high resolution images.

Fig. 6. An example of using cut-plane operations to relabel a segmentation into three components (left lateral ventricle, right lateral ventricle and third ventricle.)

Fig. 7. **a.** User interface for feature image specification: the user is setting the values of the smooth threshold parameters, and the feature image is displayed in the orthogonal slice views using a color map.**b.** User interface for active contour initialization: the user has placed two spherical bubbles in the caudate nucleus.

Fig. 8. User interface for specifying contour evolution parameters, including the relative weights of the forces acting on the contour. The intuitive interface is shown on the left and the mathematical interface on the right. The parameter specification window also shows how the forces interrelate in a two-dimensional example.

Fig. 9. The user interface for contour evolution. The image on the left shows SNAP before evolution is run, and the image on the right is taken after the contour runs for a few seconds. In this example, an edge based feature image is used.

Fig. 10. Two and three-dimensional views of the caudate nucleus. Coronal slice of the caudate: Original T1-weighted MRI (left), and overlay of segmented structures (middle). Right and left caudate are shown shaded in green and red; left and right putamen are sketched in yellow, laterally exterior to the

caudates. The nucleus accumbens is sketched in red outline. Note the lack of contrast at the boundary between the caudate and the nucleus accumbens, and the fine-scale cell bridges between the caudate and the putamen. At right is a 3D view of the caudate and putamen relative to the lateral ventricles.

Fig. 11. Box and whisker plots showing order statistics (minimum, 25% quantile, median, 75% quantile, maximum) of overlaps between pairs of segmentations of the caudate nucleus. The horizontal axis represents 13 different categories of pairwise comparisons that are listed in Table 3, and the vertical axis plots the Dice Similarity Coefficient (DCS). Columns 1-4 are SNAP-to-SNAP comparisons, columns 5-8 are manual-to-manual comparisons, and columns 9-13 are mixed-method comparisons. Orange boxes represent intrarater comparisons for specific raters; red boxes stand for intrarater comparisons pooled over available raters; light blue boxes are interrater comparisons for specific pairs of raters; and dark blue boxes are interrater comparisons pooled over available rater pairs.

Fig. 12. Box and whisker plots of overlaps between pairs of segmentations of the lateral ventricles nucleus. All columns represent SNAP-to-SNAP comparisons; columns 1 and 2 represent intrarater comparisons for specific raters; column 3 plots intrarater comparisons pooled over the two raters; and column 4 shows interrater comparisons.

Figures

Fig. 1. An illustration of the parameters κ and λ that determine the shape of the function g that inversely maps the values of image gradient magnitude to the range $(0, 1]$. The top row shows the shapes of the mapping function under different values of the parameters, and the bottom row shows the resulting feature images.
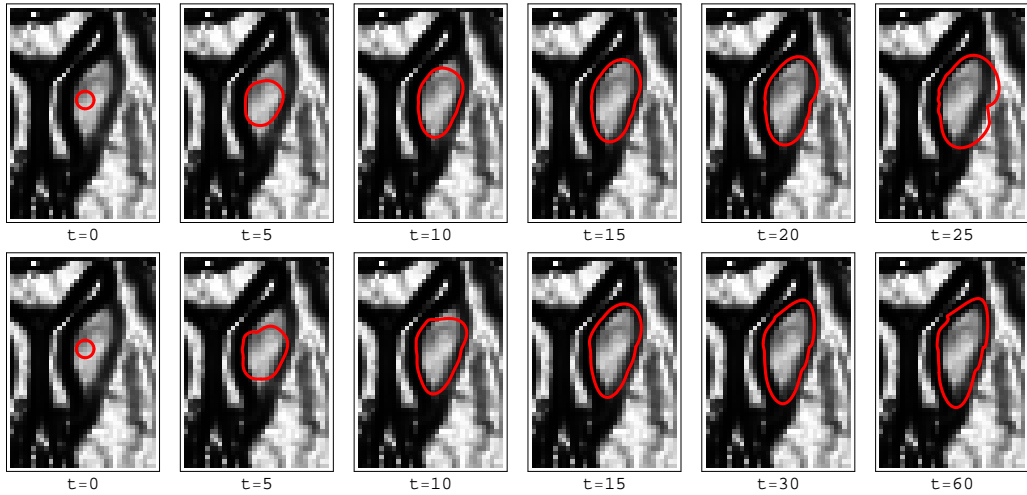

Fig. 2. Examples of edge-based contour evolution before (top) and after (bottom) adding the advection term. Without advection, the contour leaks past the boundaries of the caudate nucleus because the external force is non-negative.

Fig. 3. The plot on the left gives an example of three smooth threshold functions with different values of the smoothness parameter κ. To the right of the plot are the input grayscale image and the feature images corresponding to the three thresholds.

Fig. 4. Active contour evolution using the feature image based on region competition. The propagation force acts outwards over the 'foreground' region (red) and inwards over the 'background' region (blue), causing the active contour to reach equilibrium at the boundary of the regions.

47

a.   b.

Fig. 5. **a.** SNAP user interface shows three orthogonal views of a volumetric image, linked by a common cursor (light blue crosshairs). A fourth panel is used to view the segmented structures in three dimensions. **b.** Alternatively, SNAP can be focused on a single slice, and zoom facilities are provided for segmenting high resolution images.
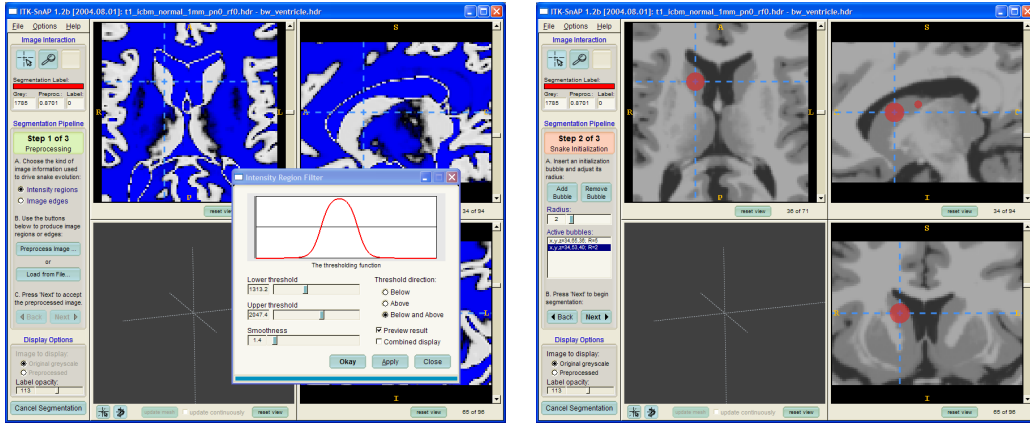

Fig. 6. An example of using cut-plane operations to relabel a segmentation into three components (left lateral ventricle, right lateral ventricle and third ventricle.)

a.   b.

Fig. 7. **a.** User interface for feature image specification: the user is setting the values of the smooth threshold parameters, and the feature image is displayed in the orthogonal slice views using a color map.**b.** User interface for active contour initialization: the user has placed two spherical bubbles in the caudate nucleus.
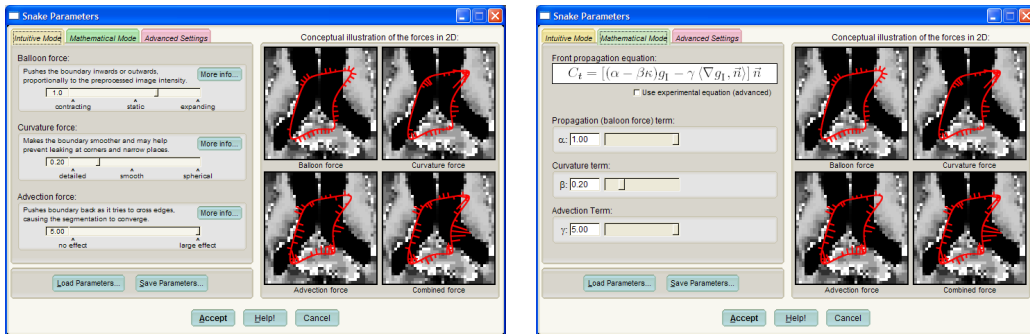

Fig. 8. User interface for specifying contour evolution parameters, including the relative weights of the forces acting on the contour. The intuitive interface is shown on the left and the mathematical interface on the right. The parameter specification window also shows how the forces interrelate in a two-dimensional example.
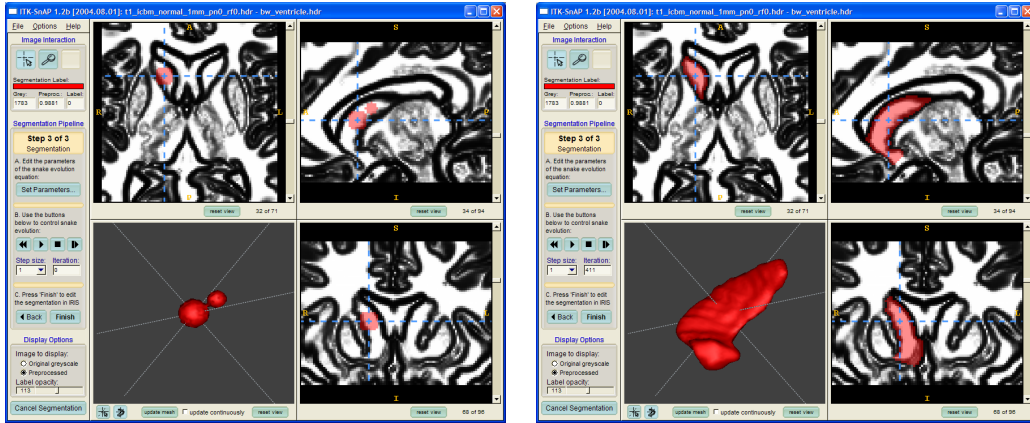
Fig. 9. The user interface for contour evolution. The image on the left shows SNAP before evolution is run, and the image on the right is taken after the contour runs for a few seconds. In this example, an edge based feature image is used.
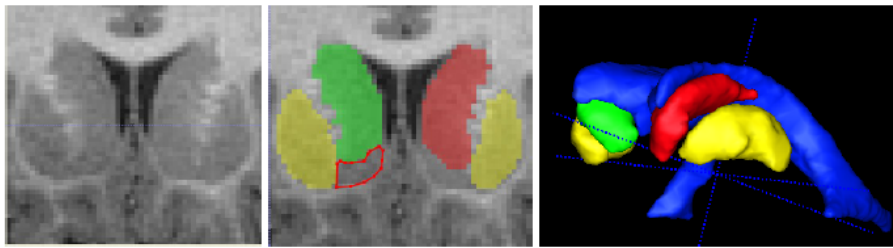

Fig. 10. Two and three-dimensional views of the caudate nucleus. Coronal slice of the caudate: Original T1-weighted MRI (left), and overlay of segmented structures (middle). Right and left caudate are shown shaded in green and red; left and right putamen are sketched in yellow, laterally exterior to the caudates. The nucleus accumbens is sketched in red outline. Note the lack of contrast at the boundary between the caudate and the nucleus accumbens, and the fine-scale cell bridges between the caudate and the putamen. At right is a 3D view of the caudate and putamen relative to the lateral ventricles.
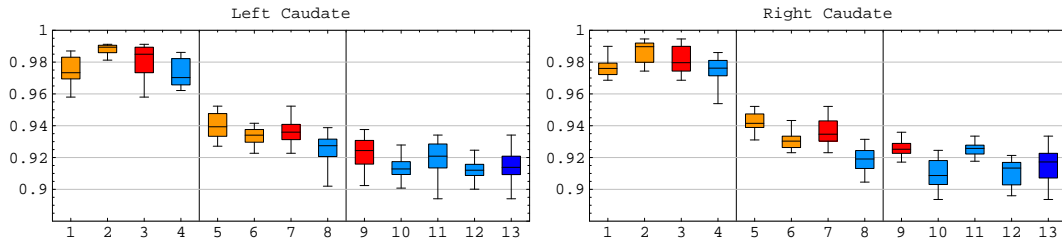
Fig. 11. Box and whisker plots showing order statistics (minimum, 25% quantile, median, 75% quantile, maximum) of overlaps between pairs of segmentations of the caudate nucleus. The horizontal axis represents 13 different categories of pairwise comparisons that are listed in Table 3, and the vertical axis plots the Dice Similarity Coefficient (DCS). Columns 1-4 are SNAP-to-SNAP comparisons, columns 5-8 are manual-to-manual comparisons, and columns 9-13 are mixed-method comparisons. Orange boxes represent intrarater comparisons for specific raters; red boxes stand for intrarater comparisons pooled over available raters; light blue boxes are inter-rater comparisons for specific pairs of raters; and dark blue boxes are interrater comparisons pooled over available rater pairs.
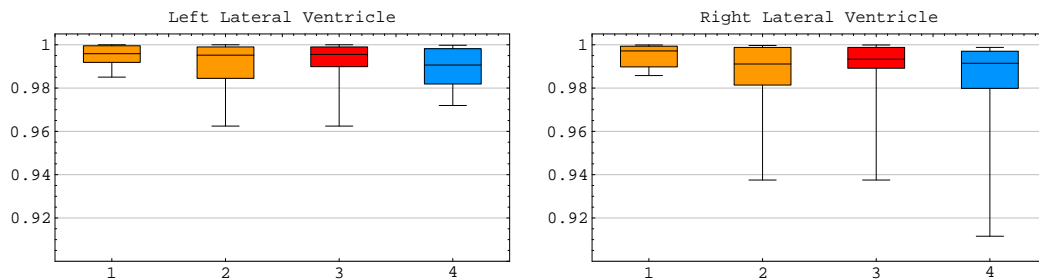

Fig. 12. Box and whisker plots of overlaps between pairs of segmentations of the lateral ventricles nucleus. All columns represent SNAP-to-SNAP comparisons; columns 1 and 2 represent intrarater comparisons for specific raters; column 3 plots intrarater comparisons pooled over the two raters; and column 4 shows interrater comparisons.