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 and Guido Gerig b, c

a Penn Image Computing and Science Laboratory (PICSL), Department of Radiology, University of Pennsylvania, PA 19104-6274, USA
b Departments of Computer Science and Psychiatry, University of North Carolina, NC 27599-3290, USA
c Neurodevelopmental Disorders Research Center, University of North Carolina, NC 27599-3367, USA

Received 29 January 2005; revised 21 December 2005; accepted 24 January 2006
Available online 20 March 2006

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-rater and 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.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Computational anatomy; Image segmentation; Caudate nucleus; 3D active contour models; Open source software; Validation; Anatomical objects

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-docu-
mented 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 6 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

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 Section 2. A brief summary of active contour segmentation and level set methods appears in Section 3.1. Section 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 Section 4. Finally, Section 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.

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 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 intrarater reliability of 0.84. Other studies report higher intrarater 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 et al., 1995; Alsabti et al., 1998; Van Leemput et al., 1999a,b) with high accuracy, but they cannot 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., in press; 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 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 to 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 selection easier for non-expert users.

Materials and methods

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

¹ SNAP binaries are available for download at www.itksnap.org; source code is managed at www.itk.org.
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):

\[
\frac{\partial}{\partial t} C(t, u, v) = F \mathbf{N},
\]

(1)

\( \mathbf{N} \) 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 in 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.

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

\[
F = \beta g_t + \gamma \left( \nabla g_t \cdot \mathbf{N} \right),
\]

(2)

where \( g_t \) is the speed function derived from the gradient magnitude of the input image \( I \), \( \kappa \) is the mean curvature of the contour, and \( \alpha, \beta, \gamma \) 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:

\[
g_t(x) = \frac{1}{1 + (\text{NGM}_t(x)/\nu)^{\alpha}}
\]

\[
\text{NGM}_t(x) = \frac{\| \nabla (G_\sigma * I) \|}{\max_t \| \nabla (G_\sigma * I) \|}
\]

(3)

where \( \text{NGM}_t \) is the normalized gradient magnitude of \( I \); \( G_\sigma * I \) denotes convolution of \( I \) with the isotropic Gaussian kernel with aperture \( \sigma \); and \( \nu \) and \( \lambda \) 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 counter-balanced by the so-called advection force \( \gamma (\nabla g_t \cdot \mathbf{N}) \), 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

\[
F = \alpha (P_{\text{obj}} - P_{\text{bg}}) + \beta \kappa.
\]

(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

---

**Fig. 1.** An illustration of the parameters \( \kappa \) and \( \lambda \) 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.
as the zeroth level set of some function $\phi$, which is defined at every voxel in the input image. The relationship $N = \nabla \phi / \| \nabla \phi \|$ is used to rewrite the PDE (1) as a PDE in $\phi$:

$$\frac{\partial}{\partial t} \phi(x,t) = F \nabla \phi.$$  \hspace{1cm} (5)

Typically, such equations are not solved over the entire domain of $\phi$ 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).

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 postprocessing 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.

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 cannot be represented, but it allows users to view

---

**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 $\kappa$. To the right of the plot are the input grayscale image and the feature images corresponding to the three thresholds.
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 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 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.

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 $P_{\text{obj}} - P_{\text{bg}}$ 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 filters 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 $\alpha$, $\beta$, and $\gamma$ 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):

$$F = \alpha h^a - \beta h^b \kappa - \gamma h^c \left( \nabla h \cdot \mathbf{N} \right).$$  (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 is 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 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 subvoxel accuracy of the segmentation is compromised. Before merging, the user has the option to export the segmentation as a real-valued image.

Results

The new SNAP tool, with its combination of user-guided 3D active contour segmentation and postprocessing 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 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.

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 2. 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.
Gray-level MRI data

Caudate segmentation uses high-resolution T1-weighted MRI with voxel size 0.78 × 0.78 × 1.5 mm³. 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 1 mm³. 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.

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.

Caudate segmentation

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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Caudate volumes (in mm³) from the validation study comparing SNAP reliability to manual segmentation

<table>
<thead>
<tr>
<th>Case</th>
<th>Right caudate volumes</th>
<th>Left caudate volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>gr1a</td>
<td>3720</td>
<td>3738</td>
</tr>
<tr>
<td>gr1b</td>
<td>3713</td>
<td>3780</td>
</tr>
<tr>
<td>gr1c</td>
<td>3790</td>
<td>3786</td>
</tr>
<tr>
<td>Mean (gr1)</td>
<td>3741</td>
<td>3768</td>
</tr>
<tr>
<td>gr2a</td>
<td>4290</td>
<td>4353</td>
</tr>
<tr>
<td>gr2b</td>
<td>4229</td>
<td>4247</td>
</tr>
<tr>
<td>gr2c</td>
<td>4233</td>
<td>4232</td>
</tr>
<tr>
<td>Mean (gr2)</td>
<td>4251</td>
<td>4277</td>
</tr>
<tr>
<td>gr3a</td>
<td>4211</td>
<td>4250</td>
</tr>
<tr>
<td>gr3b</td>
<td>4299</td>
<td>4155</td>
</tr>
<tr>
<td>gr3c</td>
<td>4264</td>
<td>4257</td>
</tr>
<tr>
<td>Mean (gr3)</td>
<td>4255</td>
<td>4221</td>
</tr>
<tr>
<td>gr4a</td>
<td>4091</td>
<td>4105</td>
</tr>
<tr>
<td>gr4b</td>
<td>4151</td>
<td>4150</td>
</tr>
<tr>
<td>gr4c</td>
<td>4081</td>
<td>4149</td>
</tr>
<tr>
<td>Mean (gr4)</td>
<td>4108</td>
<td>4134</td>
</tr>
<tr>
<td>gr5a</td>
<td>4112</td>
<td>4197</td>
</tr>
<tr>
<td>gr5b</td>
<td>4165</td>
<td>4226</td>
</tr>
<tr>
<td>gr5c</td>
<td>4191</td>
<td>4237</td>
</tr>
<tr>
<td>Mean (gr5)</td>
<td>4156</td>
<td>4220</td>
</tr>
</tbody>
</table>

Table 2

Intrarater and interrater reliability of caudate segmentation

<table>
<thead>
<tr>
<th>Validation type</th>
<th>Side</th>
<th>Intra A</th>
<th>Intra B</th>
<th>Intra AB</th>
<th>Inter AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra- and interrater manual</td>
<td>Right</td>
<td>0.963</td>
<td>0.845</td>
<td>0.902</td>
<td>0.916</td>
</tr>
<tr>
<td>(A Manual, B Manual)</td>
<td>Left</td>
<td>0.970</td>
<td>0.954</td>
<td>0.961</td>
<td>0.967</td>
</tr>
<tr>
<td>Manual vs. SNAP</td>
<td>Right</td>
<td>0.963</td>
<td>0.967</td>
<td>0.964</td>
<td>0.967</td>
</tr>
<tr>
<td>(A Manual, B SNAP)</td>
<td>Left</td>
<td>0.970</td>
<td>0.969</td>
<td>0.969</td>
<td>0.907</td>
</tr>
<tr>
<td>Intra- and interrater SNAP</td>
<td>Right</td>
<td>0.967</td>
<td>0.958</td>
<td>0.962</td>
<td>0.958</td>
</tr>
<tr>
<td>(A SNAP, B SNAP)</td>
<td>Left</td>
<td>0.969</td>
<td>0.990</td>
<td>0.978</td>
<td>0.961</td>
</tr>
</tbody>
</table>

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 h to 30 min, while the training period to establish reliability was several months for the manual method and significantly shorter for SNAP.
Table 3
Overlap statistics between pairs of caudate segmentations, categorized by different methods and raters

<table>
<thead>
<tr>
<th>Category</th>
<th>$\mu_{\text{Left}}$</th>
<th>$\sigma_{\text{Left}}$</th>
<th>$\mu_{\text{Right}}$</th>
<th>$\sigma_{\text{Right}}$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SNAP A vs. SNAP A</td>
<td>97.5</td>
<td>0.836</td>
<td>97.7</td>
<td>0.598</td>
<td>15</td>
</tr>
<tr>
<td>2. SNAP B vs. SNAP B</td>
<td>98.8</td>
<td>0.291</td>
<td>98.7</td>
<td>0.679</td>
<td>15</td>
</tr>
<tr>
<td>3. SNAP intrater average</td>
<td>98.1</td>
<td>0.931</td>
<td>98.2</td>
<td>0.823</td>
<td>30</td>
</tr>
<tr>
<td>4. SNAP A vs. SNAP B</td>
<td>97.3</td>
<td>0.810</td>
<td>97.5</td>
<td>0.686</td>
<td>45</td>
</tr>
<tr>
<td>5. Manual A vs. Manual A</td>
<td>94.0</td>
<td>0.802</td>
<td>94.2</td>
<td>0.617</td>
<td>15</td>
</tr>
<tr>
<td>6. Manual C vs. Manual C</td>
<td>93.4</td>
<td>0.568</td>
<td>93.1</td>
<td>0.504</td>
<td>15</td>
</tr>
<tr>
<td>7. Manual intrater average</td>
<td>93.7</td>
<td>0.752</td>
<td>93.6</td>
<td>0.783</td>
<td>30</td>
</tr>
<tr>
<td>8. Manual A vs. Manual C</td>
<td>92.5</td>
<td>0.913</td>
<td>91.9</td>
<td>0.697</td>
<td>45</td>
</tr>
<tr>
<td>9. SNAP A vs. Manual A</td>
<td>92.2</td>
<td>0.991</td>
<td>92.6</td>
<td>0.515</td>
<td>45</td>
</tr>
<tr>
<td>10. SNAP A vs. Manual C</td>
<td>91.3</td>
<td>0.602</td>
<td>91.0</td>
<td>0.893</td>
<td>45</td>
</tr>
<tr>
<td>11. SNAP B vs. Manual A</td>
<td>91.9</td>
<td>1.11</td>
<td>92.5</td>
<td>0.421</td>
<td>45</td>
</tr>
<tr>
<td>12. SNAP B vs. Manual C</td>
<td>91.2</td>
<td>0.635</td>
<td>91.1</td>
<td>0.733</td>
<td>45</td>
</tr>
<tr>
<td>13. SNAP vs. Manual C</td>
<td>91.5</td>
<td>0.860</td>
<td>91.5</td>
<td>0.999</td>
<td>135</td>
</tr>
</tbody>
</table>

Letters A, B, and C refer to individual raters. Overlap values are given in percent, i.e., overlap ($S_1, S_2$) = 100%*$DSC (S_1, S_2)$. The number of pairs in each category is given in the last column ($n$).

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.

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 intramethod reliability and intermethod validity. However, this is to be compared with the significantly reduced segmentation time and short rater training time of SNAP.

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 \(\binom{12}{2} = 6\) ordered pairs. For each pair, we count the number of 2 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 \(S\) and \(S'\) as the Dice Similarity Coefficient (DSC):

\[
DSC(S, S') = \frac{2 \text{Vol}(S \cap S')}{{\text{Vol}(S)} + \text{Vol}(S')}
\]

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

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 postoperative prostate segmentation. Variance components in the ANOVA model include the method (\(M \in \{\text{SNAP, Manual}\}\)), case (i.e., subject; \(C \in \{1, 2, 3, 4, 5\}\)), anatomy \(A \in \{\text{l. caudate, r. caudate}\}\), 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:

\[
LDSC(S, S') = \ln \left( \frac{DSC(S, S')}{1 - DSC(S, S')} \right)
\]

Our ANOVA results are listed in Table 4. We conclude that SNAP segmentations are significantly more reproducible than manual segmentation (\(F = 1066, P < 0.001\)). 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.

### 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 segmentation, we present the results of a

<table>
<thead>
<tr>
<th>Validation type</th>
<th>Side</th>
<th>Intra A</th>
<th>Intra B</th>
<th>Intra AB</th>
<th>Inter AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-interater SNAP Right</td>
<td>0.9942</td>
<td>0.9999</td>
<td>0.9970</td>
<td>0.9917</td>
<td></td>
</tr>
<tr>
<td>Intra-interater SNAP Left</td>
<td>0.9977</td>
<td>0.9987</td>
<td>0.9976</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reliability was measured based on 3 replications of 5 test datasets by two raters.
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 postprocessing, 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 min, 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.

**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 intramethod 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 counting 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 laboratories, 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

---

**Table 7**

<table>
<thead>
<tr>
<th>Category</th>
<th>μ&lt;sub&gt;Left&lt;/sub&gt;</th>
<th>σ&lt;sub&gt;Left&lt;/sub&gt;</th>
<th>μ&lt;sub&gt;Right&lt;/sub&gt;</th>
<th>σ&lt;sub&gt;Right&lt;/sub&gt;</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SNAP A vs. SNAP A</td>
<td>99.5</td>
<td>0.481</td>
<td>99.5</td>
<td>0.482</td>
<td>15</td>
</tr>
<tr>
<td>2. SNAP B vs. SNAP B</td>
<td>99.0</td>
<td>1.16</td>
<td>98.6</td>
<td>1.69</td>
<td>15</td>
</tr>
<tr>
<td>3. SNAP intrarater average</td>
<td>99.3</td>
<td>0.898</td>
<td>99.1</td>
<td>1.31</td>
<td>30</td>
</tr>
<tr>
<td>4. SNAP A vs. SNAP B</td>
<td>98.9</td>
<td>0.914</td>
<td>98.3</td>
<td>2.20</td>
<td>45</td>
</tr>
</tbody>
</table>

Overlap values are given as percentages, i.e., overlap \( S_1(S_2) = 100\% \times \text{DSC} (S_1, S_2) \).
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 semiautomated 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.

Acknowledgments

The integration of the SNAP tool with ITK was performed by Cognitica Corporation under NIH/NLM PO 467-MZ-202446-1. The validation study is supported by the NIH/NIBIB P01 EB002779, NIH Conte Center MH064065, and UNC Neurodevelopmental Disorders Research Center, Developmental Neuroimaging Core. The MRI images of infants and expert manual segmentations are funded by NIH ROI MH61696 and NIMH MH 64580 (PI: Joseph Piven). Manual segmentations for the caudate study were done by Michael Graves and Todd Mathews; SNAP caudate segmentation was performed by Rachel Smith and Michael Graves; Rachel Smith and Carolyn Kylstra were raters for the SNAP ventricle segmentation.


References


