Machine Learning for Brain Image Segmentation

Jonathan Morra, Zhuowen Tu, Arthur Toga, Paul Thompson

Laboratory of Neuro Imaging, Department of Neurology,

University of California, Los Angeles

Abstract

In this chapter, we review a variety of algorithms developed by different groups for automatically segmenting structures in medical images, such as brain MRI scans. Some of the simpler methods, based on active contours, deformable image registration, and anisotropic Markov Random Fields, have known weaknesses, which can be largely overcome by learning methods that better encode knowledge on anatomical variability. We show how the anatomical segmentation problem may be re-cast in a Bayesian framework. We then present several different learning techniques increasing in complexity until we derive two algorithms recently proposed by the authors. We show how these automated algorithms are validated empirically, by comparison with segmentations by experts, which serve as independent ground truth, and in terms of their power to detect disease effects in Alzheimer's disease. We show how these methods can be used to investigate factors that influence disease progression in databases of thousands of images. Finally we indicate some promising directions for future work.

Keywords

Learning, Segmentation, AdaBoost, Support Vector Machines, Registration, Feature Selection, Shape Analysis, Disease Modeling

Introduction

Automated analysis of brain scans is increasingly important as the cost of acquiring a brain scan decreases, and the frequency of their use increases. Drug trials and genetic studies often collect hundreds or thousands of images, and efficient algorithms are increasingly needed to compute morphometric statistics. Automated segmentation has been successfully applied to magnetic resonance images (MRI), which are used clinically to examine disease effects. Research studies of large-scale image databases now survey thousands of images at once. These population-based image analyses have discovered how diseases spread in the living brain over time (Thompson, et al. 2003), which medications best resist brain changes in disease (Jack, et al. 2008b; Thompson, et al. 2008), and have discovered specific genes that protect the brain from illness (Hua, et al. 2008), or increase the risk for disease (Leow, et al. 2008; Morra, et al. 2008b; Morra, et al. 2008c). All of these studies have been accelerated by learning approaches that identify and analyze features in brain images automatically (Fischl, et al. 2002; Grenander and Miller 1998).

MRI scans can be automatically analyzed using a sequence of several steps, including intensity normalization, registration to a common template, segmentation of specific substructures, and statistical analysis. In this chapter, we will focus on current trends for segmenting brain structures on MRI, focusing specifically on learning methods. Most of these approaches are somewhat generic, and have been used to segment images of the heart, liver, lungs, and other organs. They are also applicable in principle to other types of biomedical images, such as computed tomography or histology (Pitiot, et al. 2005).

In MRI studies, automated segmentations have been used to compute volumetric measures or shape statistics for specific brain regions, in studies of Alzheimer's Disease (Apostolova, et al. 2007; Clare, et al. 2003; Csernansky, et al. 2000; Morra, et al. 2008a; Morra, et al. 2008b; Morra, et al. 2008c), epilepsy (Lin, et al. 2005), childhood development (Gogtay, et al. 2006), autism (Nicolson, et al. 2006), drug-related degeneration in methamphetamine users (Thompson, Hayashi, Simon et al., 2004), and effects of lithium treatment in bipolar illness (Bearden, et al.

2007). Figure 1 shows an example of a subject's brain MRI segmented both by hand and automatically.

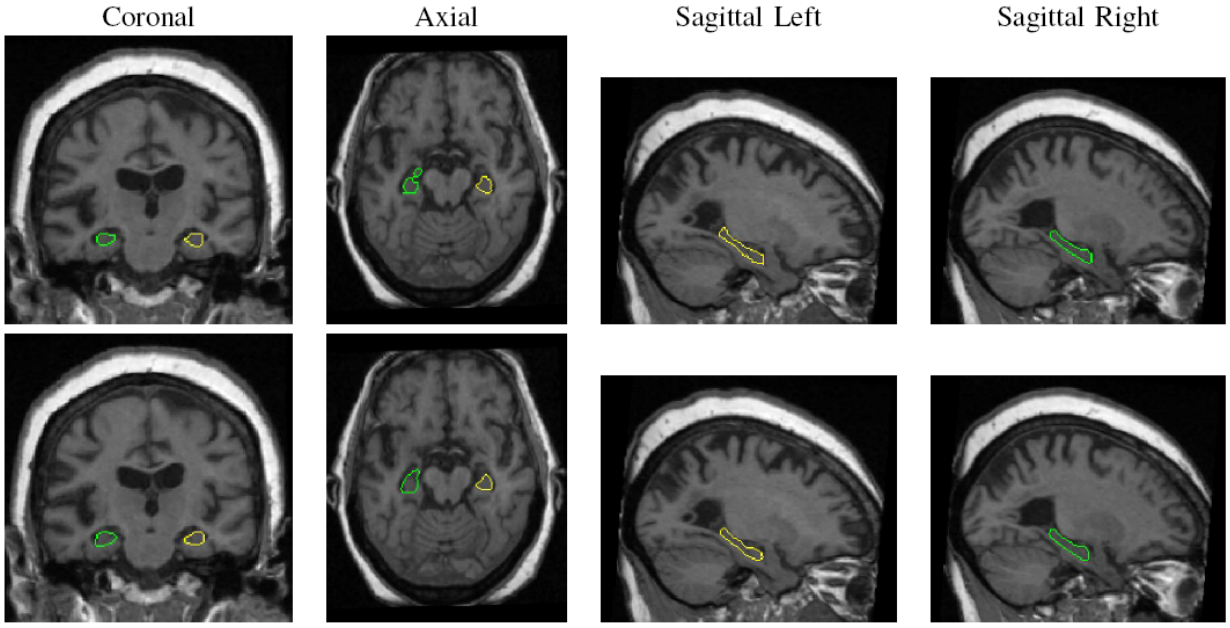

Figure 1: An example of a human brain MRI scan with the hippocampus segmented by hand in the first row, and also segmented automatically by one of the authors' algorithms, Ada-SVM, on the second row. The images labeled *Sagittal Left and Right* show sagittal slices through the hippocampus in the left and right brain hemispheres (there is one hippocampus on each side of the brain).

Anatomical segmentation is a key step in many of these imaging projects, but most studies still rely on manual segmentations by experts, who delineate each region of interest (ROI) in consecutive sections of each subject's 3D MRI scan (Apostolova, et al. 2006; Schuff, et al. 2008). This is time consuming, especially in very large studies. For instance, the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Jack, et al. 2008a) is a longitudinal study of 800 subjects scanned five times. Assuming it takes about 2 hours to manually segment the hippocampus from an MRI, then segmenting the hippocampus for all subjects in ADNI would take 2 hours x 2 hippocampi per individual x 800 subjects x 5 time points = 16,000 man-hours for just the hippocampus in this study; clearly this process needs to be automated.

The goal of this chapter is to give an overview of the general principles of image segmentation based on learning. We introduce various methods, increasing in complexity, finally describing a state-of-the-art segmentation algorithm that overcomes several limitations of prior methods. Throughout, we discuss validations that evaluate the accuracy and reproducibility of the segmentations. Finally, we highlight some directions for future research.

Background

Most subcortical segmentation algorithms may be grouped into two broad classes: those that rely on low-level intensity based information – also called *appearance-based models* – and those that incorporate higher-order shape information about the objects in the image. Deformable templates, or atlases, for example, are canonical images that contain detailed 3D labelings of brain structures. An automated image registration algorithm can deform this labeled template to match a new image, transforming the labels in the template onto the new image (Collins, et al. 1994). Many atlasing efforts have been coupled with nonlinear image registration techniques, that use elastic or fluid transformations of images to reshape an atlas template to match new images (Christensen, et al. 1993; Toga and Thompson 2000). Morphometric statistics can then be computed from the deformed labels. Additional statistics and maps can be derived from the deformation field used to align the atlas to the new image (e.g., tensor-based morphometry (Ashburner, et al. 1998; Hua, et al. 2008; Thompson and Toga 2000)).

Alternatively, this atlas deformation approach can be used in reverse. Instead, all subjects' MRI scans are non-rigidly aligned to a common template, or atlas, where all ROIs have already been segmented. This process, called spatial normalization or image warping (Toga 1999), allows regional statistics to be derived for other measures such as functional activation or metabolism, depending on the type of image aligned to the atlas. The accuracy with which a new image may be registered to a common atlas, depends on the registration model chosen to deform the image (e.g., elastic, fluid), the similarity measure defined between the images to optimize their

alignment (e.g., cross-correlation, mutual information, etc.), and the number of degrees of freedom in the transformation model (see (Klein, et al. 2008), for a comparison of 15 such approaches; approaches with more degrees of freedom tend to perform best).

Miller et al. (Miller 2004), for example, developed a complex process called Large Deformation Diffeomorphic Metric Mapping (LDDMM), which deforms individual subjects' brain MRI scans to a common template, and then deforms the corresponding ROIs back to the original image space. The approach is based on the metric pattern theory of (Grenander 1976; Grenander and Miller 1998). The vector-valued flow field, which is applied to the deforming template, is computed by solving a partial differential equation derived from continuum mechanics, assuming that the image is embedded in a deforming fluid. This ensures that the mapping is diffeomorphic (i.e., the deforming image does not tear or fold), and the resulting maps that are provably one-to-one, smooth, and have smooth inverses (Christensen, et al. 1996; Christensen, et al. 1995). Because the labeling error of these registration-based approaches depends on the accuracy of the registration method, some variants of this basic approach have used multiple registrations, or even multiple labeled atlases, to segment a single image, prior to combining the multiple segmentations using voting or other statistical methods (Chou, et al. 2008a; Chou, et al. 2008b; Lepore, et al. 2008; Lepore, et al. 2007). The tactic of 'targetless' spatial normalization (Kochunov, et al. 2005), in which all images in a database are deformed to all others, has also been proposed to minimize errors associated with single registrations. It also removes statistical bias associated with prescribing a single individual as the anatomical reference. Even so, all such deformable atlas methods assume that the image to be labeled is a smoothly deformed version of the template; this assumption is problematic when tumors or other pathologies are present, and it is also not ideal for labeling sulcal features of the cortex, as not all sulci are consistent in their incidence and topology across subjects.

3D deformable atlas methods were originally based on simpler deformable template formulations, such as active contours or 'snakes,' in which moving curves are attracted to structure boundaries in 2D images (Davatzikos and Prince 1996; Yushkevich, et al. 2006). In

active contour approaches, a user initializes an approximate boundary on a subject's image and an energy minimization algorithm drives the curve to surround the ROI. Over the years, these methods were generalized to segment 3D surfaces with spherical topology such as the cortex (Cohen and Cohen 1992), using governing equations that drive a moving surface into correspondence with a target boundary in the image, often specified by an intensity isovalue. Deformable template methods based on parameterized curves and surfaces were also generalized to include level sets, permitting topological changes in templates if needed (McInerney and Terzopoulos 2000; Osher and Sethian 1988). The governing equations for these templates initially penalized high curvature regions, sometimes leading to models that were artificially smooth. Recent innovations, such as gradient vector flow, have allowed them to segment more convoluted surfaces in medical images, including the cortex (Han, et al. 2004). Other approaches have initialized numerous templates all at once in an image, combining the results through genetic or evolutionary programming (Pitiot, et al. 2002b).

Deformable template methods usually make strong assumptions about the topology of the objects to be segmented in the image. This is both a strength and a limitation. A complementary segmentation approach uses low-level computer vision methods to classify each voxel in the image into one of several classes, often as a pre-processing step for other methods that aggregate these labeled classes into an object. This may be considered a "bottom-up" approach, as it classifies each voxel first, without necessarily referring to the global structure of the image. Bayesian tissue classifiers, for example, are widely used to classify an MRI image into distinct tissue classes such as gray and white matter and cerebro-spinal fluid (CSF). These tissue types have characteristic intensities in T1-weighted MR images, but they are corrupted by noise and partial volume effects, so that each tissue class's intensity mean and variance parameters need to be estimated from individual images. Tissue types in standard MR images are also easier to differentiate by considering Bayesian prior information on the expected locations and intensities of the signals, after transforming images to a standard anatomical space (such as normalization to an atlas). Wells et al. (Wells, et al. 1996) proposed a Bayesian tissue classification for brain MRI driven by the expectation maximization (EM) algorithm. This method was extended by Shattuck et al. (Shattuck, et al. 2001) to better accommodate voxels containing mixtures of tissue

classes (partial volumed voxels), and to better adjust the image for shading artifacts and intensity inhomogeneities associated with the non-uniform sensitivity of the MR scanner. If these common artifacts are not adjusted in advance - or iteratively during the classification (as in an EM algorithm) - serious misclassifications of tissue can occur (for example, the cortical gray matter may appear to have zero thickness in some regions of the image).

Many learning methods have been used to analyze intensity based information in images, including support vector machines (SVM) (Morra, et al. 2007; Powell, et al. 2008), artificial neural networks (ANN) (Pitiot, et al. 2004; Pitiot, et al. 2002a), and principal components analysis (PCA) (Golland, et al. 2005). Additionally, Fischl et al. (Fischl, et al. 2002) have integrated both appearance and shape-based information into a publicly available software suite called FreeSurfer which, among other things, performs subcortical segmentation by first registering new subjects' images to a common template. It then uses an anisotropic Markov Random Field to learn the likely stereotaxic positions and intensities for different anatomical structures. Domain specific knowledge, about which subcortical structures are likely to be adjacent, is encoded into the Bayesian prior distribution of the segmentation algorithm. A related approach called BrainVisa (Mangin, et al. 2004) performs a tissue classification of the image, and then performs a 3D skeletonization of the extracted cerebro-spinal fluid component. A topological analysis of the junctions between voxels in this branching structure is used to create an attributed relational graph (Mangin, et al. 1994) that models deep sulcal surfaces as a network of interconnected 3D sheets. These sheets are then labeled using an artificial neural network that uses pre-compiled statistics on which features are likely to occur in each region of the 3D standard template space.

Given this plethora of algorithms, it is perhaps surprising that none has proven effective in all situations or for all structures. Semi-automatic methods (such as snakes) become ineffective as the number of subjects grows large, and registration-based methods are limited by the accuracy of the registration, the subjects used to create the template, and by the structures delineated on the template (for instance, a registration method would be ineffective in finding multiple

sclerosis lesions, tumors or pathologies that may be distributed throughout the brain). Some learning applications have also been somewhat limited when they incorporate little or no shape information. The user is then forced to select important information for segmentation *a priori* (this could potentially include intensity information, position information, information from various filters, etc.). Even so, the most useful information may differ for different structures, suggesting the need for adaptive learning algorithms that can identify the most relevant information for segmenting a structure in a principled way.

Learning for Segmentation

Basic Procedure

One goal of image segmentation algorithms that label brain substructures is to assign a label to each voxel in the image assigning it to one of possibly many classes (e.g., hippocampus, caudate, ventricles, etc.). For the remainder of this chapter, we focus on the two-class classification problem (e.g., "is this voxel part of the hippocampus or not"), but most methods here can be extended to multi-class classification relatively easily.

Supervised learning approaches require two distinct groups of images, one for training and one for testing. Training images are images that have been manually segmented by experts; these are used to learn patterns associated with a particular ROI. Testing images are an independent set of images not included in the training set. These are used to validate how well the patterns were learned, by labeling structures automatically. To compare an algorithm's performance with ground truth defined by an expert, the testing images should also be manually segmented, although obviously (after validation) new datasets would not need to be manually segmented. One of the potential pitfalls of learning algorithms is over-fitting, where a model memorizes the

training images instead of creating a general model. For this reason, a separate group of images must be used for validation.

Limitations

Not every study can benefit from learning based methods. First, these methods may not be ideal for small studies where absolute precision is needed. In a small study, to create a training set most of the images may already be segmented by hand. There is almost always a trade-off between the low efficiency and high quality of manual segmentations, versus the high efficiency of automated segmentations, but the need to hand-trace a training set. A second potential limitation is poor training ROI delineation. Since the goal of learning methods is to learn patterns associated with the training images, if these patterns are inexactly or irregularly captured by the expert raters, the patterns will not be learned well and the resulting model will suffer, though unsupervised and weakly supervised methods can help with incomplete or inconsistent data. This problem is usually addressed by developing a standardized protocol for tracing anatomy, with explicit rules to ensure that inter-rater reliability is high. These protocols may be extremely complex in the case of cortical surface anatomy and may require neuroanatomical training that is expensive and time-consuming to obtain, over and above the time needed to actually segment the images (see, e.g., Hayashi et al., 2002; http://www.loni.ucla.edu/~khayashi/Public/medial_surface/). Since different manual raters can trace the same training set, it is common to compute measures of inter-rater and intra-rater reliability. A reasonable target for the learning algorithm is that it should agree with manual raters at least as well as human raters tend to agree with each other (see (Morra, et al. 2008e), for this type of validation). This is in some respects analogous to a Turing test, in which the algorithm's performance should not be easily distinguishable from that of a group of human raters.

Preprocessing

Preprocessing involves any of a number of processes that help the segmentation algorithm to produce a more accurate model. Ideally these should all be fully automatic, so that the entire segmentation process involves no user interaction. Common preprocessing algorithms include registration of images to a common coordinate space, intensity normalization, and bias field correction. Registration involves aligning a given target image to a common template image, and, as noted earlier, this may involve a linear or nonlinear (warping) transformation. If all images in a study are first aligned to a common template, a specific ROI is then already in approximately the same region of the coordinate space for all subjects, which makes learning patterns easier and reduces the search space for a particular ROI. For the algorithms described here, simple linear registration has proved sufficient, but extra power may be gained through non-linear registration (although at considerable computational cost).

Intensity normalization entails removing artifacts introduced by the scanner and rescaling all the image intensities so that ROIs have approximately the same intensities across subjects. On standard T1-weighted MRI, relative signal intensities of gray and white matter and CSF are in the same order (gray matter being darker and white matter being lighter), but there is no absolute scale for the signal intensity units, unless specialized MRI protocols are used (e.g., relaxometry or parametric imaging). As such, intensity rescaling is usually beneficial; it may be either linear or non-linear, and may involve histogram normalization or equalization. Finally, bias field correction involves removing a specific bias introduced by the MRI machine. When a scan is performed, even voxels in the same tissue class may have spatially-varying intensities. A bias field corrector attempts to remove these spatial dependencies. Initial approaches to this problem often involved scanning a phantom (i.e., a geometrical test object) to model the field as a multiplicative correction factor. More recently, entropy-based methods have been proposed that automatically sharpen the intensity histogram to maximize its entropy (Boyes, et al. 2008; Shattuck, et al. 2001; Sled, et al. 1998)

Features

Most learning approaches for image segmentation combine several image features to form a decision rule that assigns each voxel to a specific structure or tissue class. Generally, a feature is any piece of information present in every scan (both training and testing scans). Examples of features include image intensities, *x*, *y*, *z* positions, and the values returned by various neighborhood filters such as mean or median filters, standard deviation filters, Haar filters, wavelets, etc. This list is not exhaustive but provides an overview of some types of features the authors have used in their own segmentation algorithms, and that have been chosen from a larger list by adaptive feature selection methods, such as AdaBoost, which minimize classification error. Features may also consist of domain-specific knowledge such as rules, or normative statistics on shape, intensity, topology, boundary regularity, connectivity, convexity, etc. When choosing features, the goal is to have the features encompass all the information necessary to segment an ROI.

Bayesian Problem Formulation

Bayes' rule provides a useful mathematical representation to formulate the segmentation problem. Assuming that we have *N* examples and *M* features, we can define $X \in (\vec{x}_1 \ldots \vec{x}_N)$ as the features where each $\vec{x}_i$ is a vector of length *M* and $Y \in (y_1 \ldots y_N)$ are the training labels with each $y_i \in \{-1,+1\}$. We then wish to solve for P(*Y*|*X*) which according to Bayes' law is P(*Y*|*X*) = P(*X*|*Y*) P(*X*) / P(*Y*). Formulating this as a maximization problem, we want *Y** = argmax P(*X*|*Y*) P(*X*). However, this problem is very difficult because of its extremely high dimensionality (with many potential features at each voxel). It can be complex to find a numerical solution using standard optimization methods.

Formulating the problem in this way combines the strengths of each of the two main schools of segmentation - those that focus on appearance models, which mainly aim to model the posterior

probability P(X|Y), and those that focus on shape information, which mainly encode features into the prior probability, P(X). We note that even registration models for feature labeling may be considered as Bayesian labeling models with *maximum a posteriori* (MAP) solutions: the intensity-based cost function serves as the likelihood term, and the regularizer (such as the elastic or fluid energy) serves as a Gibbs prior encoding the plausibility of different image deformations. The connection between Bayesian theory and registration is further presented in (Gee and Bajcsy 1998; Gee, et al. 1995; Thompson and Toga 2000).

Linear Discriminant Analysis

One very simple attempt at creating a decision rule in feature space is based on Fisher's linear discriminant analysis (LDA). In LDA, it is assumed that a decision rule can be formulated by projecting the data from M dimensional feature space onto a one-dimensional line and then thresholding. To do this, we first divide the feature vectors into two classes D_1 and D_2 containing the positive and negative examples (i.e. known members of each of the two classes). Next, we define the sample mean as the average of all feature vectors for each class such that $\vec{m}_i = \frac{1}{n_i} \sum_{\vec{x} \in D_i} \vec{x}$ where i is either plus one or minus one. Next, we define the scatter matrix $S_w = \sum_{i=1}^{2} \sum_{\vec{x} \in D_i} (\vec{x} - \vec{m}_i)(\vec{x} - \vec{m}_2)^t$. Finally, we can define the best one-dimensional metric for classification as $\vec{w} = S_w^{-1}(\vec{m}_1 - \vec{m}_2)$, and determine the optimal threshold empirically. For a more detailed analysis, including derivations, refer to Duda et al. (Duda, et al. 2001).

Problems with LDA include the fact that the classifier must have a linear form, the classifier's performance degrades when the two groups to be classified are not perfectly separable in feature space, and when the training set gets very large the minimum error rate in feature space is not achievable. These last two problems are especially important in medical image segmentation because of the large number of examples in an image and the high complexity of the images.

Support Vector Machines

Many of the pitfalls of LDA can be overcome by a more sophisticated learning technique, support vector machines (SVMs). SVMs look for support vectors to separate the data in the feature space. To create an SVM, first we must define a few terms including $\vec{w}$, the separating hyperplane, *b*, a scalar bias term so that the hyperplane is not forced to go through the origin, z_i, the slack variables which are examples that are on the wrong side of $\vec{w}$, and *C*, which is a user defined parameter to control the trade-off between the number of slack variables and maximizing the margin. The margin is the distance from $\vec{w}$ to the nearest example in feature space. Ideally this should be as large as possible in order to increase the generalization ability of the resulting classifier. The SVM formulation is then as follows:

$$\min \frac{1}{2}\|\vec{w}\|_2 + C\sum_{i=1}^{N} z_i$$
$$\text{subject to } y_i(\vec{w}\cdot\vec{x}_i - b) \geq 1 - z_i$$

$\frac{1}{\|\vec{w}\|_2}$ is the margin of the hyperplane according to the $l^2 - norm$ (based on the sum of squared distances). This problem may also be solved by formulating a dual problem and maximizing it as follows:

$$\max \sum_{i=1}^{N} \alpha_i - \sum_{i,j} \alpha_i \alpha_j y_i y_j \vec{x}_i^T \vec{x}_j$$
$$\text{subject to } \alpha_i \geq 0$$

Here α_i is a dual representation of the hyperplane such that $\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i$, and the final classification of a new example is $\text{class}(\vec{x}) = \text{sign}(\vec{w}\cdot\vec{x} + b) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i \vec{x}_i \cdot \vec{x} + b\right)$. Typically

quadratic programming is used to solve this problem. In the dual formulation of SVM, every time an $\vec{x}$ appears, another one appears right next to it (in both the maximization problem and the classification formula). This allows the introduction of the "kernel trick", which involves replacing every $\vec{x} \cdot \vec{x}$ by $K(\vec{x}, \vec{x})$ where K is any general kernel function. Some examples of kernels include polynomial kernels, radial basis functions, and sigmoid kernels. This may be viewed either as creating a non-linear classifier, or as projecting the feature space into a very high-dimensional space and using a linear classifier there. For a more detailed analysis, including derivations, refer to Vapnik et al. (Vapnik 2000).

In addition to returning a class assignment, SVM also returns a "confidence" value (or margin), which is the distance from the separating hyperplane to a given example. Thresholding this confidence at values different from b will produce a more conservative or aggressive classifier. Additionally, this "confidence" may be converted to a pseudo-probability using the logit transform $p = \frac{1}{1 - e^{-x}}$ which transforms (-∞, +∞) to (0, 1). Therefore, SVMs can easily be altered to output a probability instead of a decision rule.

Though the SVM is considered as an off-the-shelve classifier, it has many limitations. First, kernel selection can be a tricky task for SVMs. Usually this is done with a "guess and check" method, or by visualizing the data and trying to see which kernel fits best. Second (and more importantly), the feature space may be too large to store in memory. Since images can contain potentially many features at each voxel it is usually not possible to run SVM on the full feature space.

Feature Selection

Selecting important features from a potentially large pool is difficult, and on the surface seems to require expert knowledge, which may vary for different studies or different structures. This becomes unmanageable as the number of ROIs and different imaging studies increases. Clearly, this problem would be more tractable if feature selection were automated.

One solution, that is easy to implement, is principal component analysis (PCA), which performs dimensionality reduction (Duda, et al. 2001). Specifically, PCA uses eigenvalue decomposition of the feature space and selects features whose variance is the largest. This projection algorithm is often used as a visualization tool to see if data tend to lie in a lower-dimensional space than that in which they are originally defined. The main problem with using PCA as a dimensionality reduction tool for classification is that PCA has no notion of classes. An easy to visualize counter-example is in two-dimensional space, where the two classes are completely separable and lie in two long straight cigar-shaped objects. PCA would project the data along the direction of the largest variance; however, a projection in the perpendicular direction would be better for classification purposes. Despite this limitation, PCA has still been used with SVM to classify medical image data (Golland, et al. 2005).

AdaBoost

AdaBoost is a learning algorithm that addresses the problem of feature selection for classification. As a type of meta-algorithm, AdaBoost uses a weighted vote of weak learners to form a strong learner. A weak learner is any learning technique that performs better than pure chance in classifying data to classes. Some examples of weak learners are decision trees, naïve Bayes classifiers, LDA, and SVMs. AdaBoost selects weak learners from a pool of candidate weak learners one at a time and assigns a weight to each of them based on their error when classifying the training data. The final classification of an example, the strong learner, is then a weighted vote of the weak learners. **Figure 2** gives an overview of the AdaBoost algorithm.

Given: N labeled training examples (x_i, y_i) with $y_i \in \{-1,+1\}$ and $x_i \in X$, and an initial (possibly uniform) distribution of weights $D_1(i)$ over the examples.

For $t = 1 \ldots T$:

- Train a weak classifier $h_t : X \rightarrow \{-1, +1\}$ using distribution D_t
- Calculate the error of h_t: $e_t = \sum_{i=1}^{N} D_t(i)\mathbf{1}(y_i \neq h_t(x_i))$
- $\alpha_t = -\frac{1}{2}\log\left(\frac{e_t}{1-e_t}\right)$
- $D_{t+1}(i) = D_t(i)\exp(-\alpha_t y_i h_t(x_i))/Z_t$ where $Z_t = 2\sqrt{e_t(1-e_t)}$ is a normalization factor

Output the strong classifier $H(x) = \text{sign}(f(x))$, where $f(x) = \frac{\sum_{i=1}^{T}\alpha_t h_t(x)}{\sum_{i=1}^{T}\alpha_t}$

Figure 2: AdaBoost algorithm. 1 is an indicator function returning 1 if its input expression is true and 0 otherwise.

AdaBoost does not directly select features, it selects weak learners. However, if we impose a one-to-one relationship between a feature and a weak learner, then AdaBoost directly selects features. In all of the authors' experiments, a weak learner is a decision tree of depth zero, also known as a *decision stump* (Morra, et al. 2007). A decision stump is a classifier that classifies all examples above a threshold within a particular feature as either positive or negative, and below a threshold the opposite. As it is based on creating a one-to-one relationship between a feature and a weak learner, a decision stump can be very quickly calculated. This is advantageous because the vast majority of computational time is spent finding the best weak learner at each iteration of AdaBoost.

In addition to being a feature selection algorithm, AdaBoost has some very useful properties including known bounds on its training and testing error. It is also provable that adding more weak learners continues to decrease the testing error. See a review by Schapire et Al. (Schapire,

et al. 1998) for a comprehensive review of AdaBoost properties. Similarly to SVM, AdaBoost may be modified to output a probability by running $f(x)$ through the logit function.

Layered Classifiers

Because of the nature of image segmentation (finding a structure in an overall image) there are almost always more negative than positive examples, which leads to a "biased" training set, in the sense that one type of example is over-represented. Additionally, many examples (such as parts of the image outside the brain) are usually easily to classify, whereas boundary voxels are much more difficult to classify. One way to exploit this fact is to use a *pruning tree*. A pruning tree involves running a classification algorithm and then discarding those examples that are clearly negative. This may be achieved by transforming the output of a classifier to a probability and thresholding it at a very low value, such as 0.1. All examples with a probability greater than the threshold are passed to another iteration of the algorithm and the process is repeated. By re-running the classification algorithm on successively more difficult examples, a different decision rule may be used for each layer of the tree.

An extension to the pruning tree is the probabilistic tree or when used in conjunction with AdaBoost, the probabilistic boosting tree (PBT; (Tu 2005)). The probabilistic tree involves separating both positive and negative examples into two separate classes instead of just retraining on all the clearly non-negative examples. However the probabilistic tree has a soft separating criterion ε, a user defined parameter. If a given classifier is unsure of an example, its probability is between $0.5 + \varepsilon$ and $0.5 - \varepsilon$, then it is passed to both of the child classifiers. This allows examples that are difficult to classify to be classified by more than one classifier. When testing, an example follows the same rules regarding passing to children and its overall probability is a combination of both classifier's outputs. See Tu (Tu 2005) for a comprehensive explanation.

AdaSVM

Since AdaBoost chooses features incrementally, it incrementally minimizes an error function, but only with respect to features already chosen. However, SVM minimizes a similar error function over all features simultaneously. Therefore, we could design an algorithm that uses AdaBoost to select features and SVM to classify examples. This should theoretically outperform AdaBoost and SVM with manually selected features (Morra, et al. 2007). The only change we must make to AdaBoost to make it compatible with AdaSVM is that AdaBoost must select weak learners without replacement. This is because having the same feature appear twice in an SVM formulation is redundant, but in traditional AdaBoost the same feature may be selected more than once with a different threshold, therefore making it a different weak learner.

AdaSVM may also be seamlessly integrated with the probabilistic tree. Instead of having ε be a user-defined parameter, we can automatically estimate an optimal ε using the margin of both AdaBoost and SVM. SVM has a natural margin built into its formulation, namely $\frac{1}{\|\vec{w}\|_2}$.

AdaBoost has also been shown to be a margin maximization algorithm, in the $l^1 - norm$ $\frac{1}{\|\vec{\alpha}\|_1}$.

The probabilistic tree can now have a dynamic ε, changing at each node (Morra, et al. 2007).

Auto Context Model

AdaBoost and AdaSVM present powerful approaches to the feature selection problem, but they are still only appearance-based methods, focusing exclusively on estimating the posterior probability. A model that also incorporates an accurate prior probability is likely to outperform

any of the methods mentioned so far. Instead of using a different method to estimate the prior probability and then pre-multiplying by it, we will instead formulate a model that estimates both the posterior and prior together, the likelihood, $P(Y|X)$.

To estimate the likelihood, we will augment the feature pool from just features based on the intensity to those also based on contextual information. Contextual features are features that encode information such as "if voxel x has a high probability of being in the ROI, then x's neighbors should also have a high probability of being in the same ROI." We do this by introducing another "image," the probability map **P**. **P** contains the probability of each voxel being in the ROI. We can then incrementally update **P**, at each iteration, to achieve the auto-context model described in **Figure 3.**

Given: N labeled training images $S = \{(Y_j, X_j), j = 1, \ldots, m\}$: For each image X_j, construct probability maps $\mathbf{P}_j^{(0)}$, with a distribution (possibly uniform on all the labels.

For $t = 1 \ldots T$:

- Make a training set $S_t = \{(Y_j, X_j(N_i), \mathbf{P}_j^{(t-1)}(N_i)), j = 1, \ldots, m, i = 1, \ldots, n\}$
- Train a classifier on both image and context features extracted from $X_j(N_i)$ and $\mathbf{P}_j^{(t-1)}(N_i)$
- Use the trained classifier to compute new classification maps $\mathbf{P}_j^{(t)}$ for each training image X_j

The algorithm outputs a series of trained classifiers $p^{(t)}(Y_i|X(N_i), \mathbf{P}^{(t-1)}(N_i))$

Figure 3: The auto context algorithm. N_i is a general neighborhood filter. Testing is done in a very similar manner except the classifier is in the testing mode, and the final output is $\mathbf{P}^{(T)}$.

The reason that the auto context model estimates the likelihood is that it is formulating its decision rule based on both intensity-based information and neighborhood filters on the probability map, which contain contextual and shape information. No specific classification algorithm has to be used to make the auto context model effective, but AdaBoost provides a

natural way to do it because the only thing to change is the feature pool. By augmenting the feature pool and allowing AdaBoost to select either contextual or intensity features, it will incorporate both sources of information seamlessly.

Also, because all that the auto context model is doing is updating the feature pool at each iteration, it is guaranteed to never increase the training error (based on the fact that AdaBoost's error is provably non-increasing (Schapire, et al. 1998)). If these context-based features prove useless, any classification algorithm will just ignore them, therefore keeping the error the same at each iteration. However, if these context-based features prove informative, the error will decrease. For most algorithms the testing error can be bounded as a function of the training error, so the overall error is guaranteed to never increase.

A reasonable stopping criterion may also be formulated by comparing $\mathbf{P}^{(t)}$ and $\mathbf{P}^{(t-1)}$. If the probability map does not change much between iterations, then new intensity-based information is not being incorporated and therefore the model is not changing much and training can be stopped (Morra, et al. 2008d; Morra, et al. 2008e).

Validation and Medical Application

In any learning study, it is important to validate the algorithm. There are many different ways to validate an algorithm, so here we will focus on two specific goals. First, it is desirable to have segmentations that agree with manual tracings defined by an expert. This agreement may be quantified by volumetric measurements or distance metrics. For comparisons with manual segmentations, the authors commonly use the definitions of Figure 4, which are widely employed in medical image segmentation studies (Morra, et al. 2007; Morra, et al. 2008d; Morra, et al. 2008e).

- Precision $= \frac{A \cap B}{B}$
- Recall $= \frac{A \cap B}{A}$
- Relative Overlap $= \frac{A \cap B}{A \cup B}$
- Similarity Index $= \frac{A \cap B}{\left(\frac{A+B}{2}\right)}$
- $H_1 = \max_{a \in A}(\min_{b \in B}(d(a,b)))$
- $H_2 = \max_{b \in B}(\min_{a \in A}(d(b,a)))$
- Hausdorff $= \frac{H_1 + H_2}{2}$
- Mean $= avg_{a \in A}(\min_{b \in B}(d(a,b)))$

Figure 4: Commonly used error metrics to evaluate the difference between manual and automatic segmentations. Define A, the manually segmented ROI, B, the automatically segmented ROI and, *d(a,b)*, the Euclidean distance between points *a* and *b*. The intersection of A and B just means that we count the voxels belonging to A and B. We note that the Hausdorff distance here is slightly different from the traditional definition; it has been modified so as to symmetrize the measurement.

Second, when analyzing brain morphometry, it is desirable to have an automated segmentation algorithm that detects disease-related effects as powerfully as possible. This is not necessarily the same as agreeing with manual raters, as an algorithm that exaggerates the disease effects may be just as useful for a disease study (or even more useful) than manual segmentations. To assess disease classification, one could look for known effects and see if the automated segmentation correctly predicts these known relationships. For instance, it is known that in Alzheimer's disease (AD), hippocampal volume declines. It is also known that in AD, hippocampal morphology is statistically correlated with various cognitive test scores, with smaller hippocampal volumes being associated with poorer cognitive performance. After these relationships have been established, a particular automated segmentation algorithm can then be used to test predicted relationships or discover new and unsuspected correlates of brain changes.

In a study by the authors involving AdaSVM (Morra, et al. 2007), we wished to show that AdaSVM was outperforming both AdaBoost and the widely-used package FreeSurfer (Fischl, et al. 2002) for hippocampal segmentation. We compared the results of all these methods to manual segmentations. The study consisted of 30 training subjects (15 normal healthy elderly subjects and 15 age-matched AD patients), and 80 testing subjects (40 normal healthy elderly, and 40 age matched AD patients). Table 1 compares manual segmentations with automated segmentations obtained using probabilistic tree AdaSVM, PBT, and FreeSurfer. **Figure 5** shows an analysis of disease effects for the same images and methods.

	AdaSVM				**AdaBoost**				**FreeSurfer**	
	Left		*Right*		*Left*		*Right*		*Left*	*Right*
	Train	*Test*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*	*Test*	*Test*
Precision	**0.821**	**0.785**	**0.844**	**0.802**	0.792	0.771	0.777	0.760	0.716	0.737
Recall	**0.868**	**0.851**	**0.848**	**0.848**	0.841	0.828	0.827	0.839	0.743	0.732
R.O.	**0.728**	**0.691**	**0.732**	**0.701**	0.687	0.665	0.666	0.663	0.572	0.577
S.I.	**0.841**	**0.814**	**0.845**	**0.822**	0.813	0.795	0.797	0.795	0.726	0.729
Hausdorff	**4.04**	**4.34**	**4.41**	**4.63**	4.64	4.98	5.20	4.83	4.97	4.99
Mean	**0.019**	**0.029**	**0.018**	**0.034**	0.024	0.028	0.027	0.041	0.075	0.065

Table 1: Precision, recall, relative overlap (R.O.), similarity index (S.I.), symmetrized Hausdorff distance, and mean distance are reported for training and testing data from 3 segmentation algorithms, probabilistic tree with AdaSVM, probabilistic boosting tree (AdaBoost), and FreeSurfer for hippocampal segmentation. Distance measures are expressed in millimeters. The best values are obtained by AdaSVM, and are highlighted in bold font; in the first 4 rows, higher numbers indicate better performance, but the in bottom two rows, lower numbers are better.

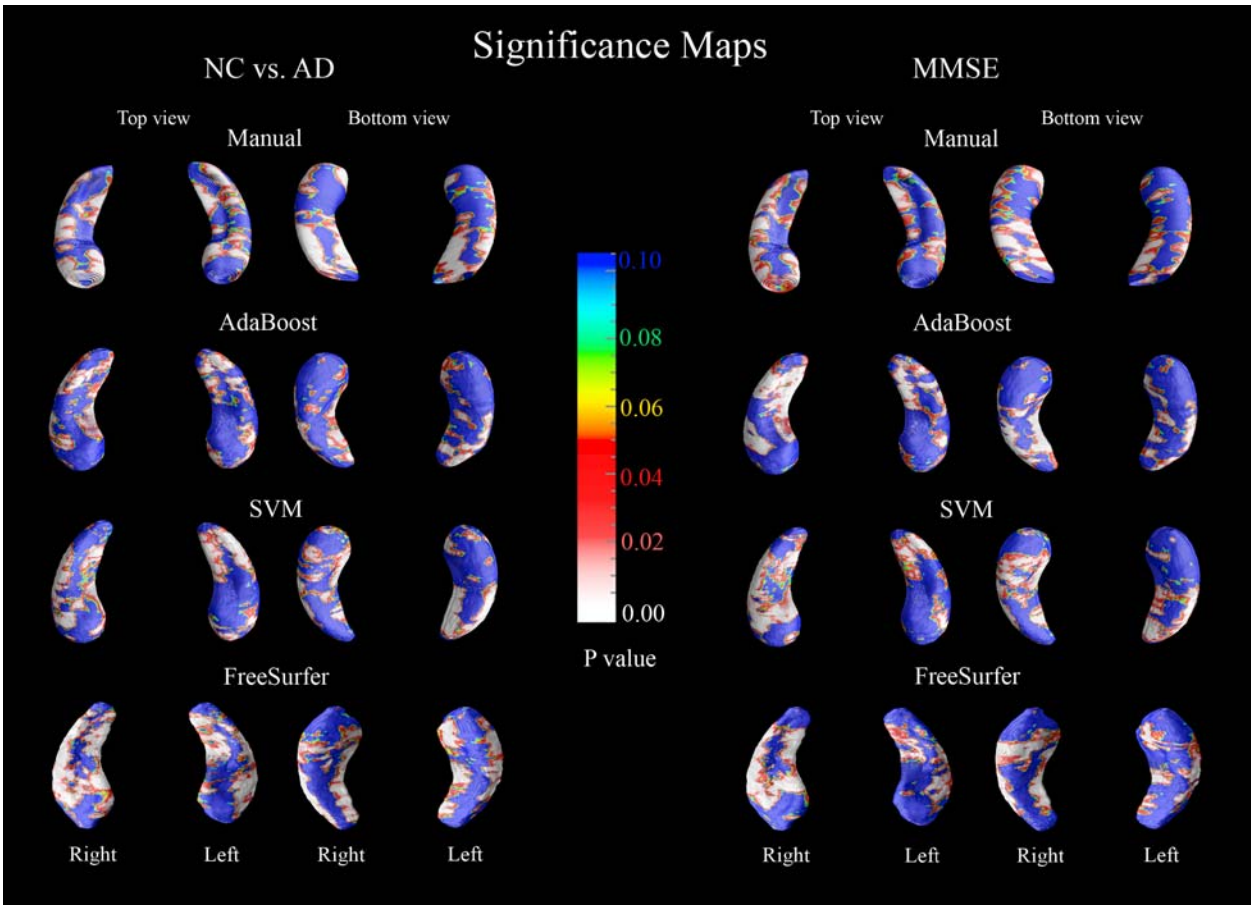

Figure 5: Significance maps (*p*-maps) based on manual, AdaSVM, AdaBoost, and FreeSurfer hippocampal segmentations. Based on the automated segmentations (binary maps), a parametric mesh is fitted to the hippocampus in each subject, and the size of the hippocampus at each surface point is measured as the distance between that surface point and a central curve threading down the center of the structure (see(Thompson, et al. 2004); also see (Styner, et al. 2005) for related work on *M-reps*). Significance maps involve performing a Student's *t*-test at each point with the null hypothesis being that there is no correlation between a covariate and the hippocampal size (radial distance) at a given point. Therefore many low *p*-values suggest that there is a statistical correlation between a given covariate and radial atrophy. Mini-mental state exam score (MMSE) is a common clinical test that measures cognitive function and is frequently used in AD studies. See Thompson et al. (Thompson, et al. 2004) for an in depth review of these surface-based *p*-maps; see (Csernansky, et al. 1998), for related work.

It is also interesting to know how many training subjects are required to achieve adequate segmentation performance. Manually segmenting a training set can be time-consuming (but not as time-consuming as hand-segmenting the whole database), so it is useful to know how the labeling error declines when different numbers of training brains are used. **Figure 6** shows that after 20 brains are included in the training set, the hippocampal segmentation error levels off, on an independent testing set.

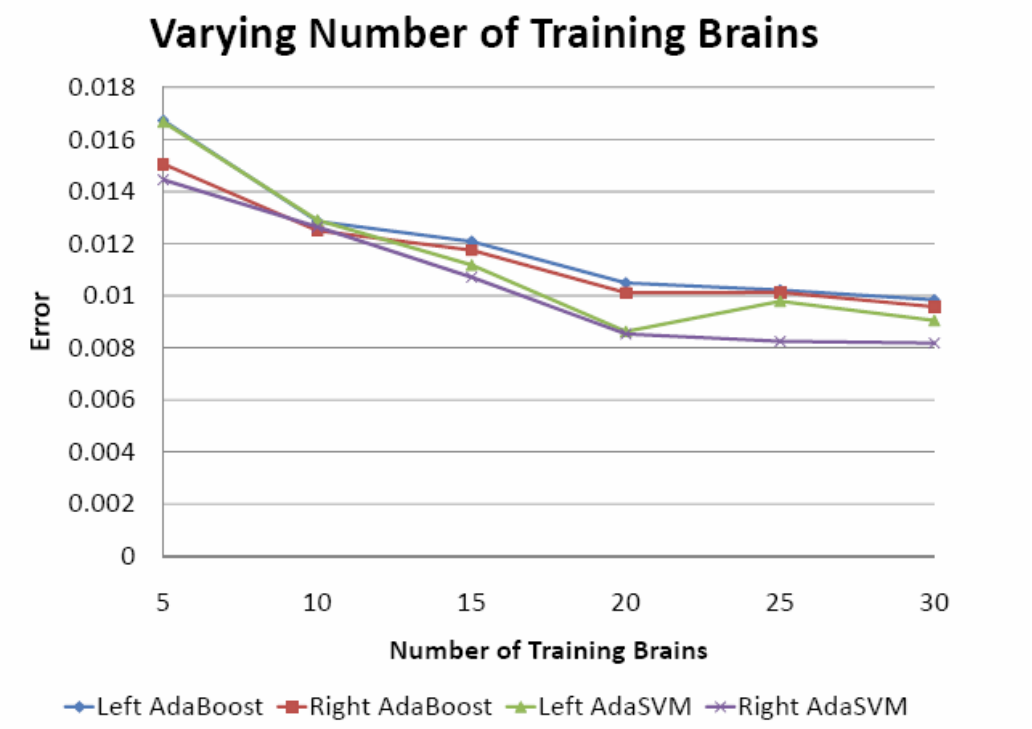

Figure 6: The effect of varying the number of images in the training set versus the testing error (number of incorrectly classified examples divided by the number of total voxels) between automated and gold standard manual hippocampal segmentations. Values are obtained for 5, 10, 15, 20, 25, and 30 brains in the training set. The curves level off after 20 brains indicating diminishing returns when training the classifier on more than 20 brains. The same trend is shown using both AdaBoost and a variant of AdaBoost that uses support vector machines (SVMs; *see main text*). Left and right indicate results for the left and right hippocampi (there is one hippocampus in each brain hemisphere).

One final analysis involves parameter optimization for the auto context model. **Figure 7** shows how two different error metrics change as the number of auto context model iterations increases (Morra, et al. 2008d). The error metrics level off after a few iterations, suggesting that at this point the classifier is just returning the probability map and is not adding new intensity-based information.

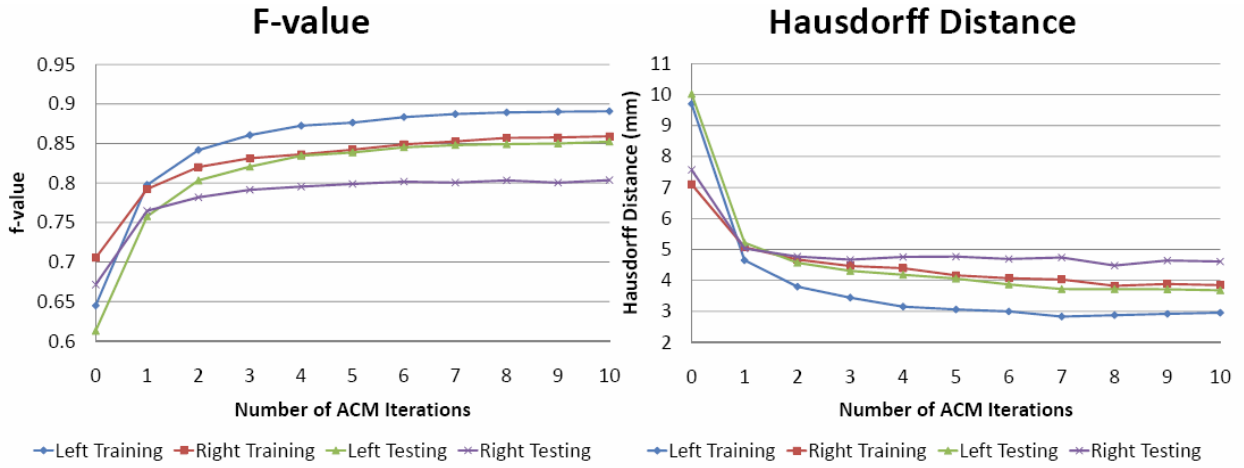

Figure 7: Effects of varying the number of iterations of the auto context model (ACM) on the Hausdorff distance and the *f*-value, defined as the average of precision and recall. All other error metrics showed a similar pattern. This data is derived by evaluating automated hippocampal segmentations against independently defined ground truth segmentations.

Future Research Directions

A key goal of future research in learning for brain image segmentation is selecting robust features. Ideal features should be independent of scanning parameters, intensity information, the registration algorithm used, etc. A universal model could then be created for each structure without the need to retrain the classifier for each new study, and would greatly increase the usability of any of the above algorithms.

Another open research question concerns online learning. Longitudinal studies of the brain (such as the ADNI initiative) commonly acquire data over a long period of time – often up to 10 years – and depending on the storage ability and computational resources of a given system it might be impractical to retrain a new model to incorporate new training data. Online learning can update a given classifier without observing previously learned data. This would make model updating much easier. Many online learning algorithms exist for other models and applications, but finding one specific to this problem is still an open research question.

Conclusion

In this chapter, we showed why automated segmentation is useful in medical imaging, and presented a variety of methods for performing it. We built up a hierarchy of learning approaches, culminating in a method that integrates both appearance and context-based information and benefits from the advantages of each. We presented some results that the authors have gathered in their own research, focusing on parameter selection for these models, and comparisons with independently defined ground truth. Finally, we provided future areas of research where innovations are likely. These developments are already having a major impact on the pace of discovery in biomedical science, integrating information from thousands of images to understand disease, drug and gene effects in large populations.

Acknowledgments

Algorithm development for this study was funded by the NIA, NIBIB, the National Library of Medicine, and the National Center for Research Resources (AG016570, EB01651, LM05639, RR019771 to P.T.), and NCRR P41 Resource Grant RR013642 (to A.W.T.).

References

Apostolova LG, Akopyan GG, Partiali N, Steiner CA, Dutton RA, Hayashi KM, Dinov ID, Toga AW, Cummings JL, Thompson PM. (2007): Structural correlates of apathy in Alzheimer's disease. Dement Geriatr Cogn Disord 24(2):91-7.

Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM. (2006): Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. Arch Neurol 63(5):693-9.

Ashburner J, Hutton C, Frackowiak R, Johnsrude I, Price C, Friston K. (1998): Identifying global anatomical differences: deformation-based morphometry. Hum Brain Mapp 6(5-6):348-57.

Bearden CE, Thompson PM, Dutton RA, Frey BN, Peluso MA, Nicoletti M, Dierschke N, Hayashi KM, Klunder AD, Glahn DC and others. (2007): Three-Dimensional Mapping of Hippocampal Anatomy in Unmedicated and Lithium-Treated Patients with Bipolar Disorder. Neuropsychopharmacology.

Boyes RG, Gunter JL, Frost C, Janke AL, Yeatman T, Hill DL, Bernstein MA, Thompson PM, Weiner MW, Schuff N and others. (2008): Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. Neuroimage 39(4):1752-62.

Chou YY, Lepore N, Avedissian C, Madsen S, Parikshak N, Hua X, Trojanowski JQ, Shaw L, Weiner M, Toga A and others. (2008a): Mapping Correlations between Ventricular Expansion, and CSF Amyloid & Tau Biomarkers in 240 Subjects with Alzheimer's Disease, Mild Cognitive Impairment and Elderly Controls. NeuroImage in press.

Chou YY, Lepore N, de Zubicaray GI, Carmichael OT, Becker JT, Toga AW, Thompson PM. (2008b): Automated ventricular mapping with multi-atlas fluid image alignment reveals genetic effects in Alzheimer's disease. Neuroimage 40(2):615-30.

Christensen GE, Rabbitt RD, Miller MI. 1993. Deformable Neuroanatomy Textbook based on Viscous Fluid Mechanics. 27th Ann. Conf. on Inf. Sciences and Systems. p 211-216.

Christensen GE, Rabbitt RD, Miller MI. (1996): Deformable templates using large deformation kinematics. IEEE Trans Image Process 5(10):1435-47.

Christensen GE, Rabbitt RD, Miller MI, Joshi S, Grenander U, Coogan TA, Van Esses DC. 1995. Topological Properties of Smooth Anatomic Maps. In: Bizais Y, Barillot C, Di Paola R, editors. Information Processing in Medical Imaging. p 101-112.

Clare L, Woods RT, Moniz Cook ED, Orrell M, Spector A. (2003): Cognitive rehabilitation and cognitive training for early-stage Alzheimer's disease and vascular dementia. Cochrane Database Syst Rev(4):CD003260.

Cohen ID, Cohen I. 1992. Models for 3D Medical Images using Finite Elements & Balloons. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Comput. Soc. Press.

Collins DL, Neelin P, Peters TM, Evans AC. (1994): Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18(2):192-205.

Csernansky JG, Joshi S, Wang L, Haller JW, Gado M, Miller JP, Grenander U, Miller MI. (1998): Hippocampal morphometry in schizophrenia by high dimensional brain mapping. Proc Natl Acad Sci U S A 95(19):11406-11.

Csernansky JG, Wang L, Joshi S, Miller JP, Gado M, Kido D, McKeel D, Morris JC, Miller MI. (2000): Early DAT is distinguished from aging by high-dimensional mapping of the hippocampus. Dementia of the Alzheimer type. Neurology 55(11):1636-43.

Davatzikos C, Prince JL. 1996. Convexity Analysis of Active Contour Problems. Proceedings of CVPR. San Francisico: IEEE.

Duda D, Hart P, Stork D. 2001. Pattern Classification: Wiley-Interscience. 654 p.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S and others. (2002): Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33(3):341-55.

Gee JC, Bajcsy R. 1998. Matching: Continuum-Mechanical and Probabilistic Analysis. In: Toga A, editor. Brain Warping: Academic Press.
Gee JC, LeBriquer L, Barillot C, Haynor DR, Bajcsy R. 1995. Bayesian Approach to the Brain Image Matching Problem. Report 95-08.
Gogtay N, Nugent TF, 3rd, Herman DH, Ordonez A, Greenstein D, Hayashi KM, Clasen L, Toga AW, Giedd JN, Rapoport JL and others. (2006): Dynamic mapping of normal human hippocampal development. Hippocampus 16(8):664-72.
Golland P, Grimson WE, Shenton ME, Kikinis R. (2005): Detection and analysis of statistical differences in anatomical shape. Med Image Anal 9(1):69-86.
Grenander U. (1976): Pattern Synthesis: Lectures in Pattern Theory. Applied Math Sci. 13.
Grenander U, Miller MI. (1998): Computational anatomy: an emerging discipline. Q. Appl. Math. LVI(4):617-694.
Han X, Pham DL, Tosun D, Rettmann ME, Xu C, Prince JL. (2004): CRUISE: cortical reconstruction using implicit surface evolution. Neuroimage 23(3):997-1012.
Hua X, Leow AD, Parikshak N, Lee S, Chiang MC, Toga AW, Jack CR, Jr., Weiner MW, Thompson PM. (2008): Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects. Neuroimage.
Jack CR, Jr., Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, J LW, Ward C and others. (2008a): The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 27(4):685-91.
Jack CR, Jr., Petersen RC, Grundman M, Jin S, Gamst A, Ward CP, Sencakova D, Doody RS, Thal LJ. (2008b): Longitudinal MRI findings from the vitamin E and donepezil treatment study for MCI. Neurobiol Aging 29(9):1285-95.
Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Hellier P, Hyun PSJ and others. (2008): Evaluation of 15 nonlinear Deformation Algorithms Applied to Human Brain MRI Registtration. NeuroImage in press.
Kochunov P, Lancaster J, Hardies J, Thompson PM, Woods RP, Cody JD, Hale DE, Laird A, Fox PT. (2005): Mapping structural differences of the corpus callosum in individuals with 18q deletions using targetless regional spatial normalization. Hum Brain Mapp 24(4):325-31.
Leow A, Yanovsky I, Parikshak N, Hua X, Lee S, Toga A, Jack C, Bernstein M, Britson P, Ward C and others. (2008): Alzheimer's Disease Neuroimaging Initiative: A One-year Follow up Study Correlating Degenerative Rates, Biomarkers and Cognition. NeuroImage in press.
Lepore N, Brun C, Chou YY, Lee AD, Barysheva M, Pennec X, McMahon KL, Meredith M, De Zubicaray G, Wright M and others. Multi-Atlas tensor-based morphometry and its application to a genetic study of 92 twins; 2008; New York, NY.
Lepore N, Brun C, Pennec X, Chou YY, Lopez OL, Aizenstein HJ, Becker JT, Toga AW, Thompson PM. (2007): Mean template for tensor-based morphometry using deformation tensors. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv 10(Pt 2):826-33.
Lin JJ, Salamon N, Dutton RA, Lee AD, Geaga JA, Hayashi KM, Toga AW, Engel J, Jr., Thompson PM. (2005): Three-dimensional preoperative maps of hippocampal atrophy predict surgical outcomes in temporal lobe epilepsy. Neurology 65(7):1094-7.

Mangin JF, Frouin V, Bloch I, Bendriem B, Lopez-Krahe J. (1994): Fast nonsupervised 3D registration of PET and MR images of the brain. J Cereb Blood Flow Metab 14(5):749-62.

Mangin JF, Riviere D, Cachia A, Duchesnay E, Cointepas Y, Papadopoulos-Orfanos D, Scifo P, Ochiai T, Brunelle F, Regis J. (2004): A framework to study the cortical folding patterns. Neuroimage 23 Suppl 1:S129-38.

McInerney T, Terzopoulos D. (2000): T-snakes: topology adaptive snakes. Med Image Anal 4(2):73-91.

Miller MI. (2004): Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. Neuroimage 23 Suppl 1:S19-33.

Morra J, Tu Z, Apostolova LG, Green A, Avedissian C, Madsen S, Parikshak N, Hua X, Toga A, Jack C and others. (2008a): Automated 3D Mapping of Hippocampal Atrophy and its Clinical Correlates in 400 Subjects with Alzheimer's Disease, Mild Cognitive Impairment, and Elderly Controls. Human Brain Mapping submitted.

Morra J, Tu Z, Apostolova LG, Green A, Avedissian C, Madsen S, Parikshak N, Hua X, Toga A, Jack C and others. (2008b): Automated Mapping of Hippocampal Atrophy in 1-Year Repeat MRI Data in 490 Subjects with Alzheimer's Disease, Mild Cognitive Impairment, and Elderly Controls. Neuroimage submitted.

Morra J, Tu Z, Apostolova LG, Green A, Avedissian C, Madsen S, Parikshak N, Hua X, Toga A, Jack C and others. 2008c. Mapping Hippocampal Degeneration in 400 Subjects with a Novel Automated Segmentation Approach. International Symposium on Biomedical Imaging. Paris, France: IEEE.

Morra J, Tu Z, Apostolova LG, Green A, Toga A, Thompson P. (2007): Comparison of AdaBoost and support vector machines for detecting Alzheimer's disease through automated hippocampal segmentation. IEEE Trans Med Imaging submitted.

Morra J, Tu Z, Apostolova LG, Green A, Toga A, Thompson P. Automatic Subcortical Segmentation Using a Contextual Model. In: Metaxas D, Axel L, Fichtinger G, Szekely G, editors; 2008d; New York, New York. Springer. p 194-201.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR, Jr. and others. (2008e): Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. Neuroimage.

Nicolson R, DeVito TJ, Vidal CN, Sui Y, Hayashi KM, Drost DJ, Williamson PC, Rajakumar N, Toga AW, Thompson PM. (2006): Detection and mapping of hippocampal abnormalities in autism. Psychiatry Res 148(1):11-21.

Osher S, Sethian JA. (1988): Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. J. Comput. Phys. 79:12-49.

Pitiot A, Delingette H, Thompson P. 2005. Automated Image Segmentation: Issues and Applications. In: Leondes C, editor. Medical Imaging Systems: Technology and Applications: World Scientific.

Pitiot A, Delingette H, Thompson PM, Ayache N. (2004): Expert knowledge-guided segmentation system for brain MRI. Neuroimage 23 Suppl 1:S85-96.

Pitiot A, Toga A, Ayache N, Thompson P. 2002a. Texture-Based MRI Segmentation with a Two-Stage Hybrid Neural Classifier. EEE 2002 World Congress on Computational Intelligence and Neural Nets. Honolulu, HI.

Pitiot A, Toga AW, Thompson PM. (2002b): Adaptive elastic segmentation of brain MRI via shape-model-guided evolutionary programming. IEEE Trans Med Imaging 21(8):910-23.
Powell S, Magnotta VA, Johnson H, Jammalamadaka VK, Pierson R, Andreasen NC. (2008): Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. Neuroimage 39(1):238-47.
Schapire RE, Freund Y, Bartlett JW. (1998): Boosting the margin: A new explanation for the effectiveness of voting methods. The Annals of Statistics 26(5):1651-1686.
Schuff N, Woerner N, Boreta L, Komfeld, Shaw L, Trojanowski JQ, Thompson P, Jack C, Weiner M, Consortium A. (2008): Progression of Hippocampal Decline in Alzheimer's Disease and Mild Cognitive Impairment in Relation to ApoE Status and CSF Biomarkers: An MRI Study of ADNI. BRAIN in press.
Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, Leahy RM. (2001): Magnetic resonance image tissue classification using a partial volume model. Neuroimage 13(5):856-76.
Sled JG, Zijdenbos AP, Evans AC. (1998): A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17(1):87-97.
Styner M, Lieberman JA, McClure RK, Weinberger DR, Jones DW, Gerig G. (2005): Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. Proc Natl Acad Sci U S A 102(13):4872-7.
Thompson P, Toga A. 2000. Elastic Image Registration and Pathology Detection. In: Bankman I, Rangayyan R, Evans A, Woods R, Fishman E, Huang H, editors. Handbook of Medical Image Processing: Academic Press.
Thompson PM, Bartzokis G, Hayashi KM, Klunder AD, Lu PH, Edwards N, Hong MS, Yu M, Geaga JA, Toga AW and others. (2008): Time-Lapse Mapping of Cortical Changes in Schizophrenia with Different Treatments. Cereb Cortex.
Thompson PM, Hayashi KM, de Zubicaray G, Janke AL, Rose SE, Semple J, Herman D, Hong MS, Dittmer SS, Doddrell DM and others. (2003): Dynamics of gray matter loss in Alzheimer's disease. J Neurosci 23(3):994-1005.
Thompson PM, Hayashi KM, De Zubicaray GI, Janke AL, Rose SE, Semple J, Hong MS, Herman DH, Gravano D, Doddrell DM and others. (2004): Mapping hippocampal and ventricular change in Alzheimer disease. Neuroimage 22(4):1754-66.
Toga A. 1999. Brain Warping: Academic Press.
Toga A, Thompson P. 2000. Brain Atlases and Image Registration. In: Bankman I, Rangayyan R, Evans A, Woods R, Fishman E, Huang H, editors. Handbook of Medical Image Processing: Academic Press.
Tu Z. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering; 2005. p 1589-1596 Vol. 2.
Vapnik VN. 2000. The nature of statistical learning theory. New York: Springer. xix, 314 p. p.
Wells WM, III, Grimson WEL, Kikinis R, Jolesz FA. (1996): Adaptive segmentation of MRI data. Medical Imaging, IEEE Transactions on 15(4):429-442.
Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. (2006): User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116-28.