Automated Landmark Localization in Diffusion MR Brain Scans using Regression Forests

Colin J. Brown
School of Computing Science
Simon Fraser University
Burnaby, BC V5A 1S6
cjbrown@sfu.ca

Abstract

Localization of anatomical landmarks in magnetic resonance (MR) images of the human brain is an important first step for a variety of analysis tasks. In particular, many image registration algorithms are based on or can be validated by relative landmark positions. In this paper, we present a method for automatic localization of anatomical landmarks in diffusion MR images. We show that it is possible to uniquely characterize specific anatomical structures in the brain using a set of voxel level and local haar-like features. A regression forest is trained on a set of hand labelled images and, on a new test image, outputs the probability of a given landmark at each voxel. We test our method on a set of 11 brain images each labelled with 7 anatomical landmarks.

1 Introduction

Even when done by experts, anatomical landmark labelling in 3-D medical images is prone to error [6]. Precise localization of landmarks in MR images of the brain is especially difficult and time consuming. However, accurate landmarks can be useful, both for aiding registration between two or more images [6], or for validating competing registration techniques [3]. Another important application is that of providing reliable seed points to tractography algorithms which delineate white matter fibers in diffusion MR images [1].

Here we present a method for automatic localization of anatomical landmarks in diffusion tensor MR (DTMR) images. Our approach is based on the idea that certain contextual image features can uniquely identify each selected anatomical landmark and that these features can be learned by training a regressor on a set of hand-labeled data. The presented method is similar to those presented by Tu [8] and Morra [5]. Instead of segmenting structural MR volumes, here we want to find specific landmarks in DTMR volumes and instead of using boosting trees, here we use regression forests. The use of regression forests (and random forests) have had recent success in identifying the positions of other anatomical structures, including the abdominal organs[2] and vertebral bodies[7]. Regression forests are an excellent choice of model when a large space of possible features are available because the training algorithm will automatically decide which features are most useful. Furthermore, decision-tree-based regressors, in general, are fast, parallelizable and able to learn highly complex mappings. To our knowledge, this is the first attempt to localize landmarks in DTMR images of the brain using a learning based approach.

In a standard DTMR image, each voxel maps to a 3 by 3 tensor representing the directions of fluid diffusion which, in the brain, correlates to the orientation of fiberous white matter tissue [1]. Whereas structural MR images are typically higher resolution, they only convey one scalar value

per voxel. In contrast, a DTMR image is typically lower resolution but contains a tensor with 6 degrees of freedom at each voxel. Thus, we can obtain more information from fewer voxels. For this reason, a major difference between this work and previous, similar works is in the choice of features.

2 Method

Given a DTMR image of the brain, we would like to locate a set of pre-determined, anatomically significant landmarks. To do this we use a set of training data to train one regression forest for each landmark. The learned regression forest is then applied to the unseen image to get an output image which indicates the likelihood of the landmark being located at each voxel. The position of the maximum likelihood voxel is selected as the location of the landmark in the test image.

2.1 Training Data

The ground-truth location of each landmark in each training image is represented as an image of probabilities. In particular, a multivariate gaussian distribution is centered on the best estimated position of each anatomical landmark. The advantage of representing the landmarks in this way is twofold: First, this soft labelling encodes the inherent uncertainty in hand-labelling known to exist (as discussed above). Secondly, the distribution of non-zero probability voxels provides more positive samples to be used by the regression forest.

2.2 Regression Forest

A regression forest is an ensemble of weak learners. Each weak learner is a binary regression tree, built from a distinct subset of random image features, F, (discussed below). The training of each regression tree is similar to that of training a single decision tree. Beginning at the root node, we have a full set of training points (voxels), each with a vector of feature values and a single target value (set by the ground-truth labelling). The feature is selected that splits the data points best. Of the remaining data points, S, at a given node some points go to the left, some to the right and this process is repeated recursively until a maximum tree depth is reached or an insuficient number of points are remaining to split on. The output value at a given leaf is set as the mean target value over all data points at that leaf. At each node, the best feature, in the subset of features available to this tree, is selected based on its maximum information gain, G.

$$max_{f \in F}[G(S, f)] \tag{1}$$

For each feature, this maximum information gain is found by scanning the set of remaining training point feature values for the best split point. For S_f, the set of values of f and S_t, the associated set of target values available at the current node, and entropy, $E(S) = -\sum_{x \in S} P(x) log P(x)$, we have:

$$G(S, f) = E(S) - \min_{vSplit \in S_f} \left(\frac{|S_{vSplit}|}{|S|} E(S_{tSplit}) + \frac{|S| - |S_{vSplit}|}{|S|} E(S_t \setminus S_{tSplit}) \right) \tag{2}$$

$$S_{vSplit} = \{v_i \in S_v | v_i < vSplit\} \tag{3}$$

$$S_{tSplit} = \{t_i \in S_t | v_i \in S_{vSplit}\} \tag{4}$$

The set of labelled points to train on is selected in attempt to balance sensitivity, specificity and computation time. By training on every voxel, we would have far more target values close to zero (i.e. negative-like samples) than target values close to one (i.e. positive-like samples). Furthermore, selection of each feature would become very expensive due to the number of points to scan over. Instead, training points are first sampled such that each target value is above some threshold. This corresponds to a spherical region around the landmark. A number of random points are then sampled from the image outside of this sphere. The ratio of negative-like points to positive-like points used here is 10:1. Alternatively, one could sample points from the gaussian distribution that created the initial labelling image. However, we want a uniform sampling over the image of

negative-like points and sampling from the gaussian would not produce this.

Each individual regression tree is trained separately with its own set of randomly selected features. Regression trees are combined into the random forest simply by averaging the resulting classifier outputs. This process ensures that the weakness of no single tree causes the entire regressor to fail. In the case of soft labelled training data, the averaged regression tree output on a new test image is an image of probabilities denoting the likelihood of the landmark being at that location.

2.3 Image Features

Similar to those methods presented by Tu [8] and Morra [5] we use Haar-like features to capture local structure at each voxel. A Haar-like feature encodes this local information by computing a difference of summed-intensities in two regions nearby the target voxel. So, any given Haar-like feature is uniquely identified by the shape and size of the two regions and their offset from the target voxel. The shape of these regions is typically cuboidal, however more complex, non-contiguous shapes such as checker patterns, stripes, etc. are also used. When selecting a random Haar-like feature the system also requires a parameter to determine the size of the the Haar window, the region of furthest allowable offset.

Haar-like feature over larger volumes are less sensitive to minor variation between images, but also less accurate. For higher accuracy we require small scale features but by making features too small, there is a risk of not capturing enough information to uniquely identify the signature structure at a landmark. In scalar images, the intensity value at a given landmark is unlikely to be unique in the image. However, the tensor at a landmark in a DTMR image has 6 degrees of freedom and so is much more likely to be unique. Thus, to improve accuracy, we want to ensure that we have some small scale features. Here, in addition to the randomly selected Haar-like features, each tree is given 5 features (one for each channel, discussed below) which simply evaluate the image at the voxel.

Each selected feature is associated with an image channel. Each image channel is a scalar image, computed from the original DTMR image. Here we compute 5 distinct channels, each providing different but complementary image information. The first channel is the fractional anisotropy (FA) of the DTMR image. This channel will be 0 at voxels where diffusion is perfectly isotropic and 1 where diffusion is occurring only in one direction. The next three channels encode the 3 components of the principal diffusion vector. The principal diffusion vector is computed by calculating the largest eigenvector of the tensor matrix. The final channel yields the magnitude of the image gradient.

2.4 Output Locations

The output of the regression forest on a new test image is a probability image (see figure 1b). Here, we simply find the maximum likelihood voxel over the whole image. However in future works, it would be possible to use this probability image, for example, as an input to another model with learned high-level knowledge about the relationships between landmark positions or to generate an interpolated probability function and calculate the landmark position with sub-pixel accuracy.

3 Results

The ability of this method to autonomously locate a set of landmarks is demonstrated on a real data set. Errors are given as Euclidean distance in voxels (where $1voxel \simeq 1.7mm$).

3.1 Data

We test our approach on a set of 11 DTMR brain images from the John Hopkins MILBA MRI In Vivo Human Database[4]. Each image has dimensions 64x96x64 and has been hand labelled by the author with 7 anatomical landmarks: The front tip of the Genu, the left and rightmost points of the Splenium, left and right points on the middle cerebellar peduncle (MCP), and extreme points on left and right far-reaching major mid-brain fiber tracts (MFT). The locations of these points are shown in figure 1a. Unfortunately, any error in hand labelling these training points propagates to the output of the automatic localization on a new image. Thus, in future work, an expert labelling will be sought.

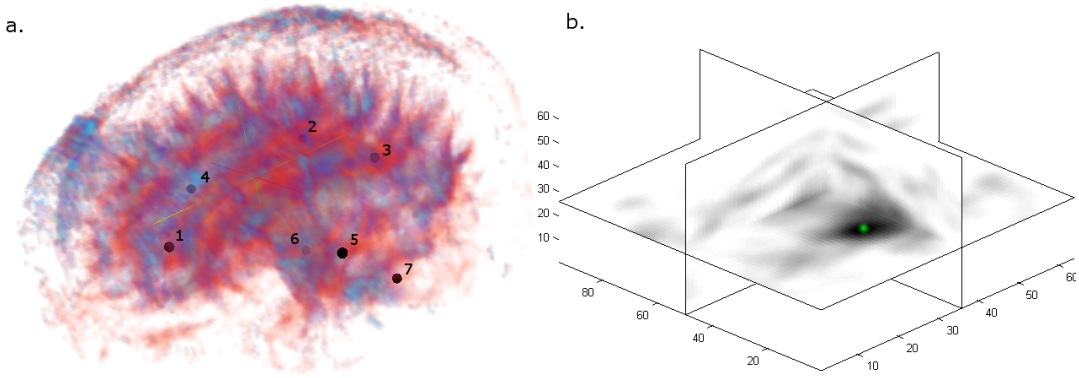

Figure 1: a) FA volume render of a brain with landmarks: 1.) Genu 2.) R. Splenium 3.) L. Splenium 4.) R. MFT 5.) L. MFT 6.) R. MCP 7.) L. MCP b) Example output of regressor in grayscale overlaid with ground-truth soft target in green. Darker values correspond to higher probabilities.

3.2 Parameters

In training, approximately 500 points are sampled from each image. In our regression forest we use 4 regression trees, each with a maximum height of 10 nodes or a maximum of 512 leaves. Each tree is given 5 identical small-scale features and 15 random Haar-like features to select from, each with a Haar-window size of 10x10x10 voxels.

3.3 Experiments

The results of training on 10 images and testing on the 11th is shown in figure 2. In figure 2 a) and b), random Haar-like features were used and the mean test error was 5.736 voxels. Running leave-one-out cross validation gives a mean error of 5.29 over all 11 images. To improve the regressor, 25 rounds of cross-validation were run, each with a set of random features. From this test, the best performing feature sets from all 25 rounds were stored. Using these features and repeating the experiment from figure 2 a) and b) generates the results in figure 2 c) and d). On this same test image, the test error was 3.9252. Certain landmarks are clearly easier to uniquely identify than others. For instance, error on the landmark at the tip of the genu changed little when using the best found features, however, the right MFT landmark was located with much higher accuracy. Running leave-one-out cross validation with the selected best features, the mean error is reduced to 4.05 voxels or 6.8mm.

3.4 Conclusions

We have shown that, using regression forests, it is possible to learn the unique structure of specific anatomical landmarks in DTMR images of the brain. While the results achieved here are an

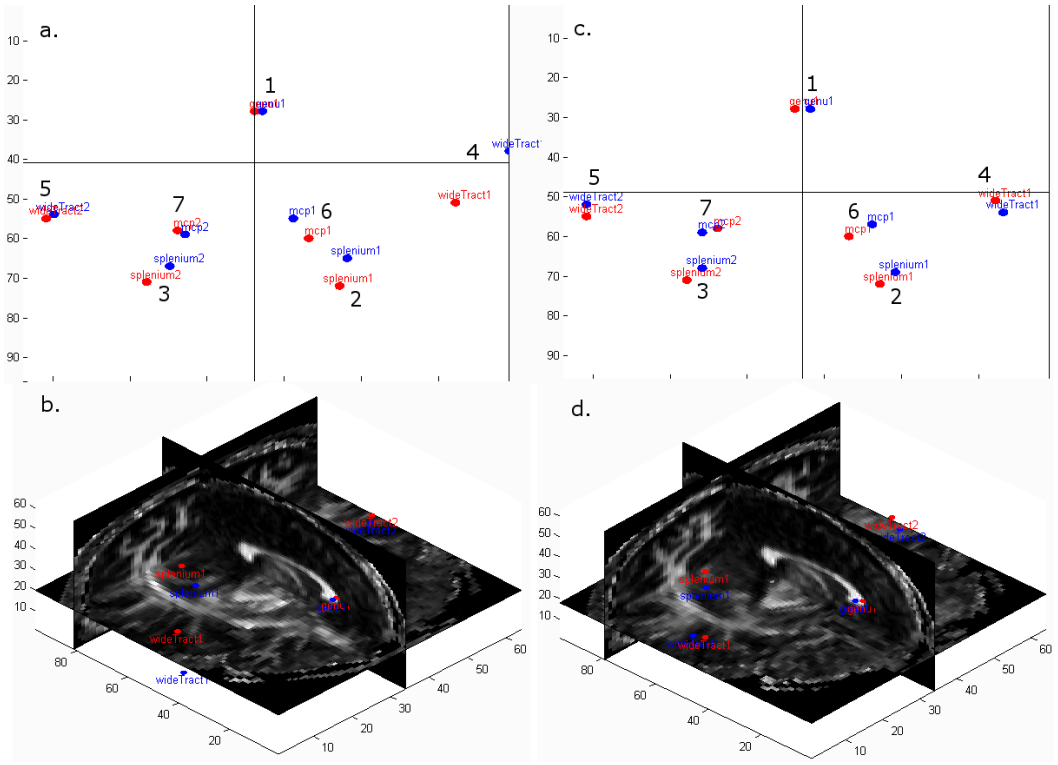

Figure 2: Red points are ground-truth, blue points are method output. a) and b) show axial and slice-plane view of regressor applied to a test image using randomly selected features. In c) and d) regressor uses best features discovered from 25 cross-validation rounds. Numbers correspond to same landmarks as numbered in figure 1.

important first step, there are many open avenues for improvement. For example, the probabilistic output of the regression forest could be used as input to a high level relational model over landmark positions. One simple possibility for improvement would be to run more rounds of cross-validation to try to find a better set of features. Also, many other features and channels could be explored. A major challenge for this approach is the number of input parameters. Different numbers and sizes of trees, Haar-like feature parameters, were all selected heuristically since a full grid search over this space would be infeasible.

Acknowledgments

Thanks to Andrew Top for the regression tree code. Also, thanks to Brian Booth for some of the code involved in loading and manipulating diffusion tensor images and for some lessons in brain anatomy. Finally, thanks to Ghassan Hamarneh for guidance and inspiration.

References

[1] R. Bammer, B. Acar, and M.E. Moseley. In vivo mr tractography using diffusion imaging. *European journal of radiology*, 45(3):223–234, 2003.

[2] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu. Regression forests for efficient anatomy detection and localization in ct studies. *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pages 106–117, 2011.

[3] I.D. Grachev, D. Berdichevsky, S.L. Rauch, S. Heckers, D.N. Kennedy, V.S. Caviness, and N.M. Alpert. A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. *NeuroImage*, 9(2):250, 1999.

[4] S. Mori. John Hopkins Medical Institute: Laboratory of Brain Anatomical MRI, in vivo human database. http://lbam.med.jhmi.edu/, accessed February 2009.

[5] J. Morra, Z. Tu, L. Apostolova, A. Green, A. Toga, and P. Thompson. Automatic subcortical segmentation using a contextual model. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 194–201, 2008.

[6] K. Rohr, H.S. Stiehl, R. Sprengel, T.M. Buzug, J. Weese, and MH Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *Medical Imaging, IEEE Transactions on*, 20(6):526–534, 2001.

[7] S. Schmidt, J. Kappes, M. Bergtholdt, V. Pekar, S. Dries, D. Bystrov, and C. Schnörr. Spine detection and labeling using a parts-based graphical model. In *Information Processing in Medical Imaging*, pages 122–133. Springer, 2007.

[8] Z. Tu, K.L. Narr, P. Dollár, I. Dinov, P.M. Thompson, and A.W. Toga. Brain anatomical structure segmentation by hybrid discriminative/generative models. *Medical Imaging, IEEE Transactions on*, 27(4):495–508, 2008.