NeuroImage 46 (2009) 726–738

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy

P. Aljabar [a,*], R.A. Heckemann [b], A. Hammers [c], J.V. Hajnal [b], D. Rueckert [a]

[a] *Visual Information Processing Group, Department of Computing, Imperial College London, 180 Queen's Gate, London, SW7 2AZ, UK*
[b] *Imaging Sciences Department, MRC Clinical Sciences Centre, Imperial College London, UK*
[c] *Division of Neuroscience and Mental Health, MRC Clinical Sciences Centre, Imperial College London, UK*

ARTICLE INFO

Article history:
Received 13 November 2008
Revised 22 January 2009
Accepted 7 February 2009
Available online 23 February 2009

Keywords:
MRI
Segmentation
Selection
Atlases
Registration

ABSTRACT

Quantitative research in neuroimaging often relies on anatomical segmentation of human brain MR images. Recent multi-atlas based approaches provide highly accurate structural segmentations of the brain by propagating manual delineations from multiple atlases in a database to a query subject and combining them. The atlas databases which can be used for these purposes are growing steadily. We present a framework to address the consequent problems of scale in multi-atlas segmentation. We show that selecting a custom subset of atlases for each query subject provides more accurate subcortical segmentations than those given by non-selective combination of random atlas subsets. Using a database of 275 atlases, we tested an image-based similarity criterion as well as a demographic criterion (age) in a leave-one-out cross-validation study. Using a custom ranking of the database for each subject, we combined a varying number n of atlases from the top of the ranked list. The resulting segmentations were compared with manual reference segmentations using Dice overlap. Image-based selection provided better segmentations than random subsets (mean Dice overlap 0.854 vs. 0.811 for the estimated optimal subset size, $n=20$). Age-based selection resulted in a similar marked improvement. We conclude that selecting atlases from large databases for atlas-based brain image segmentation improves the accuracy of the segmentations achieved. We show that image similarity is a suitable selection criterion and give results based on selecting atlases by age that demonstrate the value of meta-information for selection.

Introduction

Magnetic resonance (MR) imaging of the brain has established itself as an essential diagnostic method in neurology research and clinical practice. Quantitative studies often rely on the capability to label or segment regions of the brain that have distinctive structural or functional properties. This enables comparisons within and between subjects for determining how such regions are affected by physiological and pathological processes as well as therapeutic measures. Such studies benefit from increasing numbers of MR images becoming publicly available for use in research. This availability has made the creation and maintenance of MR image databases incorporating structural segmentations (manual or otherwise) more feasible. Good examples are the Internet Brain Segmentation Repository[1] and the *LONI Probabilistic Brain Atlas* (Shattuck et al., 2008). An obvious application of this work is the use of expert annotations in the form of prior information to assist in providing automatic segmentations of query or unseen images.

An atlas, in the context of this work, is defined as the pairing of a structural MR scan and a corresponding manual segmentation. Given an atlas, a segmentation for an unseen query subject can be estimated using image registration. The atlas MR image can be registered to the query image, yielding a transformation which allows the atlas segmentation to be transformed and treated as a segmentation estimate for the query subject. Within this process, commonly called atlas-based segmentation, the atlas that is propagated can represent a single segmented individual (Iosifescu et al., 1997; Svarer et al., 2005; D'Haese et al., 2003). The propagation of the atlas might also form a step within a larger framework. For example, probabilistic or 'soft' atlases may be propagated and treated as priors in a Bayesian framework within a further segmentation step (Murgasova et al., 2006).

Sources of error in atlas-based segmentations include registration error, the possibility that the atlas used is anatomically unrepresentative of the query image to be segmented (for example if there are topological differences) or existence of labelling errors in the atlas segmentation, something that cannot be overcome by accurate registration.

If a database of atlases is available, multiple segmentations from a group of atlases can be propagated to the query. After propagation, they can be treated as separate classifiers and fused to form a single consensus segmentation estimate. The main benefit of the multi-atlas

---

* Corresponding author.
E-mail address: paul.aljabar@imperial.ac.uk (P. Aljabar).

[1] www.cma.mgh.harvard.edu/ibsr.

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.02.018

segmentation approach is that the effect of errors associated with any single atlas propagation can be reduced in the process of combination. Multi-atlas segmentation has been shown to be effective in comparison with other atlas-based approaches (Rohlfing et al., 2004a) and for the task of segmenting structures in the human brain (Heckemann et al., 2006a; Klein and Hirsch, 2005). Relevant work has also been carried out on the methods used for combining classifiers within a multi-atlas segmentation framework (Warfield et al., 2004; Rohlfing et al., 2004b).

Multi-atlas segmentation faces various issues, however, if the number of atlases in the database becomes large. On a practical level, if every atlas is registered with the query image, the computational cost of segmentation increases linearly with the size of the database. More importantly, it is possible that the population represented by the atlases is heterogeneous, for example in terms of age, morphology or pathology. In this case, for a given query, certain subjects in the database may be more appropriate as candidate segmentations than others. Propagating and combining only these subjects' atlases is likely to produce a better segmentation estimate than one that draws on the full atlas database.

These considerations provide a motivation for the selection of atlases that are appropriate for a given query image, and this work presents an investigation of a practical strategy for such a selection approach within the context of multi-atlas segmentation. Rohlfing et al. (2004a) and Wu et al. (2007) investigated the optimal selection of a single template during atlas-based segmentation. Our work contrasts with this in that we select multiple atlases for subsequent propagation and fusion. We present the results of a series of experiments to assess the performance of atlas selection using a database consisting of 275 MR images and accompanying manual subcortical segmentations. Automated segmentation is carried out based on ranking and selecting atlases from a database according to criteria that are expected to predict their suitability for segmenting a given target. To test this, the accuracy of the resulting segmentations is measured using leave-one-out cross-validation and compared with the accuracy of segmentations derived from combining random sets of atlases. We also investigate different criteria for ranking the atlases and the effect of selecting and combining increasing numbers of atlases from a ranked set.

Over the mainly subcortical structures studied, a mean Dice overlap of 0.854 was obtained using selection. This compares with a reference value of 0.811 obtained by fusing random sets of atlases. For individual structures, selection provides typical Dice accuracy gains of 0.02 to 0.05 over random sets with the biggest improvement of 0.12 being shown by segmentations of the caudate nucleus.

In this paper, methods for multi-atlas segmentation and selection are initially described. This is followed by descriptions of the experiments to assess the effectiveness of atlas selection and their results which are discussed in the final section. Part of the research presented in this study appeared previously in a conference paper (Aljabar et al., 2007).

Methods

We describe multi-atlas segmentation along with the motivation and possible strategies for atlas selection. These strategies can be based on image information within the atlases or on subject-specific meta-information.

Background: multi-atlas segmentation

Atlases within a database can be registered to a query image, and their segmentations can be transformed and subsequently fused or combined to provide a consensus segmentation estimate for the query. Sometimes described as classifier fusion or label fusion, this method is illustrated schematically in Fig. 1. This multi-atlas approach to segmentation reduces the effect of errors associated with individual propagated atlases. For example, a registration error for a particular propagated atlas is less likely to affect the final segmentation when combined with other atlases. The proportion of errors incurred during propagation that are independent are those that are averaged out when multiple atlases are combined (Heckemann et al., 2006a). As well as a gain in accuracy, Heckemann et al. (2006a) also demonstrate that precision improves as more atlases are combined.

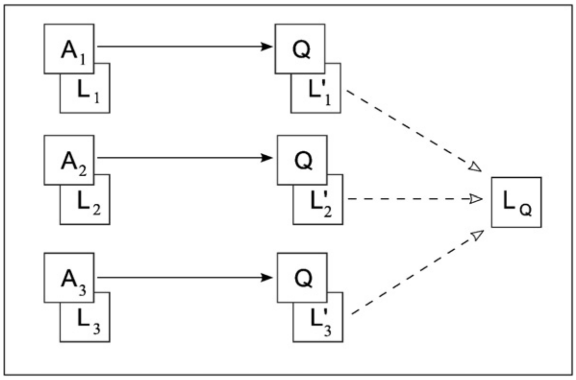

Fig. 1. Schematic illustration of multi-atlas segmentation. A set of atlas anatomical images A_i are registered to the query anatomy Q. The resulting transformations are used to transform the corresponding atlas segmentations L_i to the query. The transformed segmentations L'_i are then combined to create an estimate of the query segmentation L_Q.

The fusion of the propagated segmentations (or classifiers) takes place at the voxel level and can be achieved in different ways. In what is probably the simplest approach, the atlas segmentations are transformed using nearest-neighbour interpolation so that they each provide a discrete or 'hard' labelling for each voxel. The final label assigned to a voxel can then be decided by 'majority vote'.

More sophisticated methods for the combination or fusion of the segmentations are also available. For example it is possible to use a linear interpolator when transforming individual labels in order to obtain a probabilistic or 'soft' estimate for the label from each segmentation. This can be used to generate an array of values (p_{ij}) for a given voxel where p_{ij} represents the confidence level or probability of the voxel being assigned label i by the jth segmentation. A number of different rules can be used to generate a consensus estimate based on such data, and a good overview of these can be found in Kittler et al. (1998).

A notable example of producing consensus segmentations in the context of medical image processing is the STAPLE framework presented by Warfield et al. (2004). The STAPLE approach uses Expectation Maximisation to iterate between the estimation of the 'true' consensus segmentation and the estimation of reliability parameters for each of the raters (which in this work are represented by propagated segmentations). The reliability parameters are based on the sensitivity and specificity of each rater and are used to weight their contributions when generating the consensus estimate. The current consensus estimate can, in turn, be used to measure the reliability of the raters and this forms the basis of the EM iterations.

The use of majority voting for each voxel has, however, been shown to be effective in a number of contexts. Rohlfing et al. (2004a) used a database of images of bee brains to show that fusing segmentations using majority voting is robust and accurate compared with, for example, the propagation of an average shape atlas, or of an individual atlas, selected according to its similarity to the query image. The vote rule has also been shown to perform well relative to other fusion approaches in a more general pattern recognition context (Kittler et al., 1998).

In the context of human brain image segmentation, we have previously presented a series of experiments to investigate the precision and accuracy of structural multi-atlas segmentation using