Medical Image Analysis 14 (2010) 39–49

Contents lists available at ScienceDirect

Medical Image Analysis

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

Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus

Eva M. van Rikxoort *[,1], Ivana Isgum, Yulia Arzhaeva [2], Marius Staring [3], Stefan Klein [4], Max A. Viergever, Josien P.W. Pluim, Bram van Ginneken

Image Sciences Institute, University Medical Center Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 22 October 2008
Received in revised form 2 October 2009
Accepted 6 October 2009
Available online 13 October 2009

Keywords:
Segmentation
Multi-atlas
Atlas selection
Local
Stopping criterion

ABSTRACT

Atlas-based segmentation is a powerful generic technique for automatic delineation of structures in volumetric images. Several studies have shown that multi-atlas segmentation methods outperform schemes that use only a single atlas, but running multiple registrations on volumetric data is time-consuming. Moreover, for many scans or regions within scans, a large number of atlases may not be required to achieve good segmentation performance and may even deteriorate the results. It would therefore be worthwhile to include the decision which and how many atlases to use for a particular target scan in the segmentation process. To this end, we propose two generally applicable multi-atlas segmentation methods, adaptive multi-atlas segmentation (AMAS) and adaptive local multi-atlas segmentation (ALMAS). AMAS automatically selects the most appropriate atlases for a target image and automatically stops registering atlases when no further improvement is expected. ALMAS takes this concept one step further by *locally* deciding how many and which atlases are needed to segment a target image. The methods employ a computationally cheap atlas selection strategy, an automatic stopping criterion, and a technique to locally inspect registration results and determine how much improvement can be expected from further registrations.

AMAS and ALMAS were applied to segmentation of the heart in computed tomography scans of the chest and compared to a conventional multi-atlas method (MAS). The results show that ALMAS achieves the same performance as MAS at a much lower computational cost. When the available segmentation time is fixed, both AMAS and ALMAS perform significantly better than MAS. In addition, AMAS was applied to an online segmentation challenge for delineation of the caudate nucleus in brain MRI scans where it achieved the best score of all results submitted to date.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Atlas-based segmentation uses registration to achieve segmentation and this has proven to be a powerful and successful concept (Rohlfing et al., 2005). Atlas-based methods start by registering an anatomical image from an atlas[5] with a target image to be segmented. To obtain a segmentation of the target image, the manual labeling of the atlas is transformed using the mapping determined during the registration; this process is called label propagation. The approach has several major advantages compared to other generic segmentation techniques. First of all, simplicity: only a registration method and a number of pre-segmented data sets are required. There is no need for landmarking and complex training procedures. Secondly, its general applicability: a wide range of segmentation tasks can be solved by this method. Thirdly, for several registration methods software is freely available.

Atlas-based segmentation is essentially a prototype based method. A critical underlying assumption is that it is possible to find a deformation that aligns the atlas with the target in such a way that label propagation lines up the objects of interest. Clearly this will not always be possible. Insufficient similarity between the atlas and the target image often results in local mismatches, which in turn leads to segmentation errors (Crum et al., 2003). Success therefore depends on the choice of a prototype: some atlases will work better than others. Two classical ways to address this funda-

* Corresponding author. Address: UCLA Thoracic Imaging Research Group, 924 Westwood Blvd., Suite 650, Los Angeles, CA 90024, United States.
E-mail address: evanrikxoort@mednet.ucla.edu (E.M. van Rikxoort).
[1] Eva M. van Rikxoort is currently employed at the Department of Radiological Sciences, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, CA, United States.
[2] Yulia Arzhaeva is currently employed at CSIRO Mathematical and Information Sciences, Australia.
[3] Marius Staring is currently employed at Leiden University Medical Centre, Department of Radiology, Division of Image Processing, Leiden, The Netherlands.
[4] Stefan Klein is currently employed at the Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands.
[5] The term atlas in this paper refers to the pair of an anatomical image and a manual labeling.

1361-8415/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2009.10.001

mental problem are blending the results obtained by different prototypes and prototype selection. Typically a combination of both is used. Several groups have attempted to improve upon single atlas-based segmentation by registering multiple atlases and blending the propagated labels, for example by averaging (Rohlfing et al., 2005, 2004; Heckemann et al., 2006). This prevents errors, as long as the majority of the labels is correct. This concept is similar to classifier combination strategies as often applied in pattern recognition methods (Jain et al., 2000). Several studies have shown that multi-atlas segmentation outperforms methods that use a single atlas (Rohlfing et al., 2005, 2004; Heckemann et al., 2006).

Although better than single atlas methods, blindly applying the same set of atlases to every target image (that is, prototype blending, but no selection) has several drawbacks. The main drawback is the computational complexity. Since atlases and target images are taken from different patients, an affine transformation is usually not sufficient and non-rigid registration is needed. Non-rigid registration of high resolution images is time-consuming. In practice a large set of atlases is needed to represent the variety of data encountered in a real application, which makes the method unsuitable for routine clinical use. The second drawback is that if the image to be segmented shows substantial local deviations from the majority of atlases (and many images will show such deviations in at least some locations), the method will still produce errors, i.e. the majority is not always right. Related to this second drawback is the fact that registering an inappropriate atlas (whether the atlas is fundamentally different from a target image or the registration method is not able to find a good transformation is not relevant to the argument) can thus make the final segmentation worse. The effect is seen when segmentation performance is plotted as a function of the number of atlases used when the atlases are ordered, on, for example, segmentation similarity (see e.g. Fig. 4 in Aljabar et al. (2007)). Initially, adding atlases improves performance but at some point, adding more atlases can deteriorate the result, albeit usually very slowly. The same phenomenon is well known from feature selection in pattern recognition (Jain and Zongker, 1997).

Based on those drawbacks three improvements for multi-atlas segmentation methods can be identified. The first improvement is to automatically select appropriate atlases for the image to be segmented (prototype selection). Secondly, when an algorithm for selecting appropriate atlases is available, a stopping criterion can be introduced to automatically stop registering scans when no further improvement is expected. Introducing an automatic stopping criterion can substantially speed up the segmentation and make multi-atlas segmentation suitable for clinical use. A third possibility to improve the segmentation results is to inspect the segmentation results *locally*. Since the atlas and target images usually show local deviations, it might be advantageous to select the most appropriate atlas locally instead of for the whole image. Local inspection and local registration have two positive effects: if the segmentation of the target image is correct except for a small part it is more efficient to only register a small part of an additional atlas image to that part. Secondly, an atlas might be locally similar to the target image but not to other parts of the image. Registering the complete atlas in this case would give local improvements but might deteriorate the results elsewhere in the image.

We integrate these improvements in a new framework for multi-atlas segmentation. Two new methods are introduced. The first one is adaptive multi-atlas segmentation (AMAS). Instead of registering all atlases to the target image, as is done in a conventional multi-atlas method (MAS), AMAS automatically selects the most appropriate atlases for a target image and automatically stops registering atlases when no further improvement is expected. The second method we propose takes this concept one step further by performing registration updates locally. We propose a framework that starts with a single atlas registration and subsequently locally inspects the likelihood that the segmentation is correct. If not, blocks from additional atlases are selected that are expected to be locally similar to the target and thus may improve the segmentation most. The process is iterative and uses different atlas blocks for different locations. Updating stops automatically when no further improvement for a block is expected. We have named this method adaptive local multi-atlas segmentation (ALMAS). Fig. 1 illustrates the ALMAS algorithm with a synthetic data example.

To illustrate the effectiveness and general applicability of AMAS and ALMAS, we applied it to two segmentation tasks. The first one is the segmentation of the heart from non-contrast, non ECG gated volumetric chest computed tomography (CT) scans, which are often performed in for example screening trials. Segmentation of the heart in such scans is challenging due to the lack of contrast between the heart and surrounding tissues along large parts of the heart border. Segmentation of the heart in such scans allows for automated assessment of cardiac abnormalities such as coronary calcification. The results will be compared to a standard multi-atlas segmentation method. The second task is the segmentation of the caudate nucleus from magnetic resonance (MR) brain images. The caudate nucleus is a periventricular gray matter structure that is involved in sensory-motor control, cognition, language, emotion and other important brain functions. It has a rather homogeneous intensity in T1-weighted MR images and it is difficult to segment because it is attached to other structures with similar intensities at multiple locations. Atlases and test data were taken from an ongoing online contest for caudate segmentation (van Ginneken et al., 2007) which allows us to compare our performance to a large number of other state-of-the-art methods, including other atlas-based schemes.

This paper is organized as follows: Section 2 provides an overview of previous work on atlas-based segmentation. In Section 3 AMAS and ALMAS will be detailed. Section 4 describes the experiments that were performed for the heart and caudate segmentation. Next, Section 5 provides the results of those experiments. Finally, Section 6 discusses the results and draws conclusions.

2. Atlas-based segmentation: previous work

Atlas-based segmentation has been successfully applied to a number of applications. For example, Heckemann et al. (2006) segmented 67 brain structures using a non-rigid registration approach and label fusion of 29 atlases. Rohlfing et al. (2004) compared several atlas-based segmentation techniques for the segmentation of structures in bee brains, and Klein et al. (2008) applied a non-rigid registration approach to segment the prostate in 3D MR images. In this section we will first briefly discuss the literature that shows the superiority of a multi-atlas approach to a single atlas approach. Then we will discuss previous work on atlas selection strategies and motivate our choice for local registrations and a stopping criterion, based on findings in the literature.

Several studies have shown that multi-atlas segmentation outperforms methods that use a single atlas. For example, Rohlfing et al. (2004) compared four approaches for atlas-based segmentation: registration of an individual atlas, registration of an average-shape atlas image, registration of the most similar image from the database of individual atlases, and registration of all atlases in the database. The multi-atlas approach was significantly better than the other approaches.

Heckemann et al. (2006) compared three atlas-based approaches for the segmentation of brain structures using non-rigid registrations. They showed that registering all atlases and fusing the labels performed better than registering an individual atlas

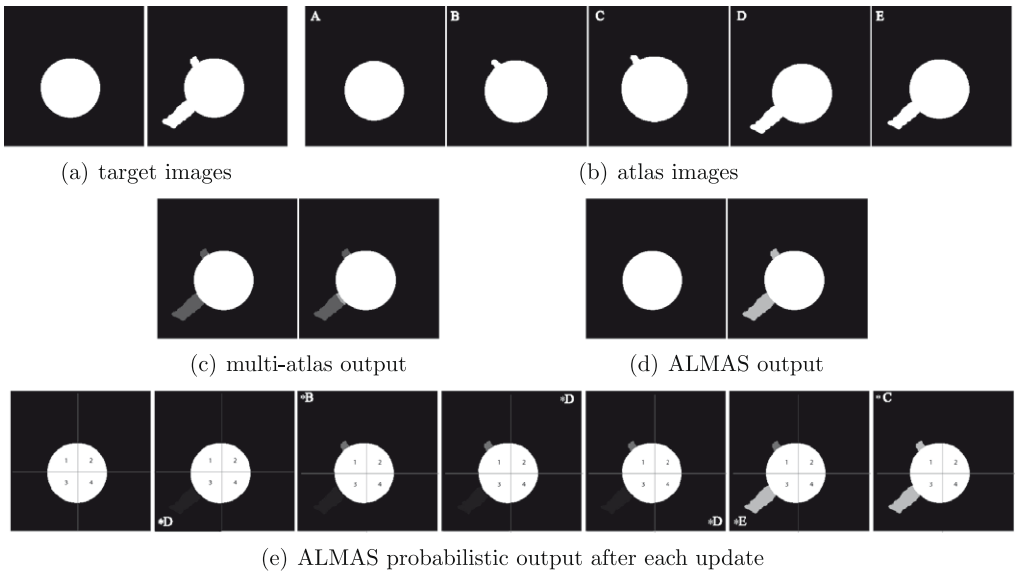

Fig. 1. Illustration of the concept of ALMAS on synthetic data. The results shown are generated by the actual algorithm. In this example there are (a) two target images to be segmented: a circle and a circle with a large and a small protuberance. (b) Shows the five atlases (A . . .E). Note that there is no single atlas that resembles the second target image. (c) With multi-atlas segmentation, all atlases are registered and the propagated labels are averaged. This leads to a probabilistic segmentation that is similar for both target images, in which the circle has a probability of 1 and the two protuberances have a probability of 0.4. As a result, thresholding this segmentation would either resemble the second target image (if a low threshold were used) or the first (when the usual threshold of 0.5 would be used). (d) Results of ALMAS. For both target images, the central circle has a probability of 1. For the second target image, the two protuberances each have a probability of 0.67. Thresholding at 0.5 would give a correct segmentation for both target images. ALMAS achieves this by first determining the optimal reference image among the atlases. This is determined to be the first atlas image. This atlas is registered to each target image. This reference atlas is subdivided in parts, 4 blocks in this case. Next, parts of atlases are registered to parts of the target image. At each step, ALMAS identifies the part in the target image that is expected to benefit most from an additional registration and it identifies the most appropriate atlas part to use. (e) Illustrates the sequence of updates for the second target image. Probabilistic output after every iteration is shown, the updated block is indicated with an asterisk and the atlas used for updating. First the circle atlas is registered. Next, blocks are updated separately. Steps three and four are performed since the stopping criterion of ALMAS is only applicable when at least 2 atlases have been registered, see equation 3. After 6 blocks ALMAS terminates automatically because further updates are not considered necessary.

or indirectly propagating a single source through the remaining atlases.

Wu et al. (2007), Klein et al. (2008), Rohlfing et al. (2004), and Aljabar et al. (2007) have investigated strategies for atlas selection which we will discuss below. To our knowledge, no previous work is available on automatically setting a stopping criterion and locally registering different atlases when more information is deemed beneficial.

Wu et al. (2007) designed an optimum template selection method for atlas-based segmentation for several regions of interest in brain images. The rationale behind their approach is that atlas-based segmentation methods of MR brain images typically use a single atlas. Anatomical variations in the images makes this a sub-optimal method, therefore they proposed to register a set of atlases (templates) and for defined ROIs select the most optimal atlas, based on normalized mutual information (NMI) non-rigid registration. This approach showed significantly better results than using a single atlas for the complete brain. The main difference from the approach presented in this paper is that the selection is performed after non-rigidly registering each atlas to the target image, so an improvement in performance is accomplished but no improvement in speed. Moreover, only a single atlas is selected for each region.

Klein et al. (2008) applied a similar approach as Wu et al. (2007) for the segmentation of the prostate in MR images. After registering all atlases, a selection of atlases was made using similarity after registration based on the NMI measure. The results showed that selecting around 50% of all available atlases gave the best results for several experiments.

Rohlfing et al. (2004) presented four approaches to select the most similar atlas from the database of atlases for the segmentation of 22 brain structures in confocal microscopy of bee brains. The first approach is the NMI between the target image and the atlas image after affine registration. The atlas image with the highest NMI after affine registration is selected to be the most similar atlas. This approach is considerably less time-consuming than the approaches applied by Wu et al. (2007) and Klein et al. (2008). The second approach they applied is the same as Wu et al. (2007) and Klein et al. (2008), the value of NMI after non-rigid registration. The last two approaches were based on the magnitude of the deformations after non-rigid registration; the atlas with the smallest average or maximum deformation is selected to be most similar. Rohlfing et al. (2004) showed that selection of the best atlas using the NMI after non-rigid registration performed best. However, the results were still significantly worse than the results of a multi-atlas strategy. Note that the goal was to select only one best atlas, and atlas selection was not applied in a multi-atlas setting.

Aljabar et al. (2007) applied four methods for atlas selection based on similarity of atlas and target image for the segmentation of brain structures in MR images. The similarity measures used were sum of squared differences, cross correlation, mutual information and NMI. The similarity was calculated in an ROI that was defined after registering all atlases to a single subject simulated image. The authors showed that all selection methods except using the sum of squared differences performed better than a normal multi-atlas approach in which random atlases were selected. The main difference with the approach presented in this paper is that all images were first registered to a single subject where the ROIs were defined, and next, all operations were performed in this space.

No previous work is available on using an automatic stopping criterion for multi-atlas segmentation. However, there are studies that indicate this would be helpful. For example, Hammers et al. (2003) developed a model for the asymptotic level of accuracy that can be achieved as the number of atlases increases. They showed,

for the application of segmenting structures in the brain, that when using more than fifteen atlases the improvement in accuracy when adding more atlases was very limited. Aljabar et al. (2007) showed that in an atlas selection scheme, the overlap values increase rapidly for the first ten to twenty most similar atlases. Using more atlases slightly deteriorated results. This indicates that when using an atlas selection scheme, the best result can be obtained by automatically stopping when no more appropriate atlases are available. A stopping criterion could simply be determined by setting a fixed number of scans to be registered. However, the optimum number of atlases differed per scan and therefore a stopping criterion that is specific for a target scan can be expected to yield better results.

The only previous work that touches upon the concept of local registrations is the approach by Wu et al. (2007) in which a different atlas was used for different regions of interest. Their results showed that this yields better results than using one complete single atlas. This suggests that locally choosing different atlases can improve the results of a segmentation by registration approach.

3. Method

This section describes three automatic segmentation methods: MAS, AMAS, and ALMAS. The goal of these methods is to produce a segmentation S for a given target image T that accurately defines the object to be segmented. All methods are capable of producing probabilistic segmentations that can be thresholded to produce binary segmentations (1 inside the object, 0 in the background). For all methods it is assumed that a set of n atlases $A_i, i = 1, \ldots, n$ with corresponding manually segmented binary images S_i is available.

3.1. Registration

In registration, an atlas image is transformed to the target image. The methods we propose are generally applicable and any registration method can be plugged in. AMAS and ALMAS require two registration methods to be available: a fast (computationally cheap) method which is used in atlas selection, and an accurate (computationally expensive) method which transforms an atlas image to a target image with high accuracy. Details about the specific registration methods we used in our experiments are given in Section 4. Throughout this section we will refer to these two registration methods as 'fast registration' and 'accurate registration'.

3.2. Multi-atlas segmentation (MAS)

We first formulate standard multi-atlas segmentation (MAS) as a reference method. This method was used in e.g. Rohlfing et al. (2004, 2006). In MAS, all n atlases are registered to T using the accurate registration resulting in n transformations u_i from A_i to T. u_i is used to transform S_i to T. We denote the transformed image as $u_i(S_i)$. As a result, for each voxel in T, n opinions exist about its label. The labels are combined by averaging all opinions for each voxel. After n registered atlases, S is defined as:

$$S = \frac{1}{n}\sum_{i=1}^{n} u_i(S_i) \tag{1}$$

To obtain a binary segmentation S is thresholded, typically at 0.5.

3.3. Atlas selection

Atlas selection is needed for AMAS and ALMAS. During the registration, a spatial mapping is determined which transforms the atlas image to the target image. In the ideal case the transformed atlas image would be equal to the target image and the difference between them would be a zero image. In reality, registration does not perfectly align the two images and local misalignment occurs. Based on this observation, we propose to employ difference images to decide which atlas image is most similar to the target image.

The rationale behind our approach for atlas selection is that the atlas image that is most similar to the target image is expected to provide the best segmentation accuracy. The optimal way to determine this is after applying the accurate registration. However, this would lead to a framework that has computation times comparable to MAS. Therefore we quantify the similarity between each atlas image and the target image by the difference image after fast registration. The atlas image selected to be most similar is the image with the lowest mean absolute difference to the target image.

Formally the atlas selection is described as follows. Let ω_i be the transformation from the fast registration between A_i and T. We define $D(\omega_i(A_i), T)$ as the absolute voxelwise difference image between $\omega_i(A_i)$ and T. The atlas selection function f is defined as:

$$f = \arg\min_{i \notin \Omega} \text{mean}(D(\omega_i(A_i), T)) \tag{2}$$

where Ω indicates the set of previously selected atlases.

3.4. Stopping criterion

As explained in Section 1, avoiding unnecessary registrations speeds up multi-atlas segmentation and may increase performance. However, it is not obvious when the optimal performance has been achieved for a certain image. Therefore we use a replacement criterion: we quantify, per voxel, the disagreement between the available propagated labels. After every accurate registration, we compute the percentage p of voxels that may change segmentation label when one additional set of propagated labels were available. When p is low, the segmentation result will not change substantially by adding additional information and therefore no further improvement is expected. Although this approach does not guarantee that the best possible segmentation result is obtained, previous work has shown that when plotting segmentation accuracy as a function of the number of atlases, this curve tends to be flat around the optimum. So, a stable segmentation (low p) equals a flat curve, which typically indicates an optimal accuracy. This complies with the intuitive notion that when experts disagree, it makes sense to obtain advice from additional experts; when experts agree this is not necessary.

We define voxels for which disagreement exists as voxels where $\frac{m}{2}$ out of m propagated labels agree. If m is odd we average the number of voxels where floor$(\frac{m}{2})$ and ceil$(\frac{m}{2})$ labels agree. For example, if the number of available labels is five, $\frac{m}{2}$ equals 2.5, therefore, we average the number of voxels where two and three labels agree. For these voxels adding one more opinion can change the segmentation label, assuming a threshold of 0.5 to convert probabilities into binary labels. The stopping criterion is now defined as a threshold θ on p. For a multi-label segmentation problem, p is defined as a threshold on the percentages of voxels that can flip label, either to the background or to another structure.

In Fig. 2a the segmentation accuracy in terms of overlap for the segmentation of the heart using MAS is plotted against the number of atlases used. In Fig. 2b the corresponding values of p are shown for each number of atlases used. It can be seen that the segmentation accuracy does not improve much after using approximately eight atlases. Similarly, p declines steeply until approximately eight atlases have been used, after that, p stabilizes, although a small decline is still present when the number of atlases is increased.

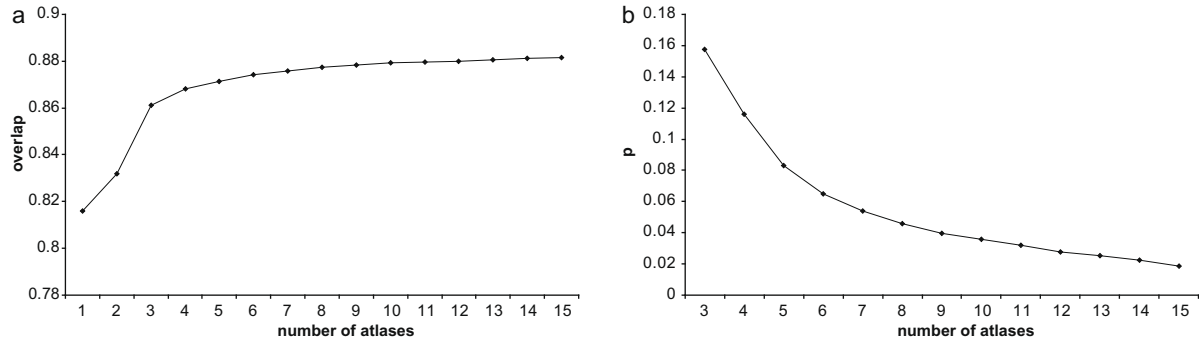

Fig. 2. This figure illustrates the stopping criterion p for segmentation of the heart in chest CT scans. In (a), the overlap as a function of the number of registered atlases is plotted when MAS is used for segmentation. (b) shows the corresponding values of p for each number of atlases used.

3.5. Adaptive multi-atlas-based segmentation (AMAS)

Based on the atlas selection and stopping criterion as defined above an adaptive multi-atlas segmentation strategy (AMAS) can be defined. In AMAS, instead of registering all available atlases as in MAS, only the most similar atlases are registered until the stopping criterion is reached. This results in fewer atlases being registered, and therefore a decrease in computational complexity. In addition, by automatically selecting the most appropriate atlases from a database better segmentation accuracy can be obtained. Let m be a counter for the number of accurate registrations that have been performed, and let S^m indicate the segmentation after m registrations. AMAS can now be stated as follows:

Input: T, A_i, S_i for $i = 1,\ldots,n$
Parameter: θ
For $i = 1,\ldots,n$ calculate ω_i. Set $p = 1, m = 0$.
While ($p > \theta$ and $m < n$) {
  Select A_i using f
  Compute u_i
  Increment m, update S^m with $u_i(S_i)$
  Compute p
}
S is now given by S^m

3.6. Adaptive local multi-atlas-based segmentation (ALMAS)

Registration usually does not perfectly align two images and local misalignment occurs. Based on this observation we propose to decide *locally* how many and which atlases to use. This strategy was implemented as follows. ALMAS starts with a preparation phase that only needs to be performed once. In this preparation phase, from the total set of n atlases, one reference atlas A_r is selected as the atlas that gives the best segmentation accuracy on the other $n-1$ atlases in a single atlas setting. Next, A_r is subdivided into z parts. We indicate these parts with a subscript j. A_r is registered to every A_i and the parts A_{rj} are propagated to define A_{ij} in every atlas. In this way each set of parts j corresponds to roughly the same anatomy. To prevent inconsistencies in the final segmentation around the border of two parts, slight overlap between the parts is used. This overlap is an extra insurance that approximately the same anatomy is present in the different images. The overlapping part is only used during registration, for updating the segmentation result after an additional block has been registered, only the actual block is used.

When the preparation phase is finished, target images can be segmented. The segmentation of T starts by computing u_r and $u_r(S_r)$. Based on $u_r(S_r)$ the parts T_j are defined in T by propagation. S^1 is initialized by $u_r(S_r)$. Next, an iterative process is started that updates the different parts of T_j, S_j separately.

Since we are now using parts, we need to determine at any stage of the algorithm which part to update. The number of voxels in a part about which disagreement exist, p_j, is a natural candidate for this purpose, but initially the p_j values for every part T_j are equal. Therefore, initially a different block update selection rule is used. This alternative rule is in effect as long as there are still parts for which $m = 1$ (i.e. only the reference atlas segmentation is available for that part). If that is the case, the part j to be updated is selected as the part with the largest average difference between block j of the reference atlas (A_rj) and block j in the target image (T_j) : $D(u_{rj}(A_{rj}), T_j)$. Once for all parts m is at least 2, the part with the highest p_j is selected to be updated. In summary, a combined rule g is defined as:

$$g = \begin{cases} \arg\max_j \text{mean}(D(u_{rj}(A_{rj}), T_j)), & \text{if } m = 1 \text{ for any part} \\ \arg\max_j p_j, & \text{otherwise} \end{cases} \quad (3)$$

Once the part to be updated has been determined, the most suitable part for updating it must be determined, and we can use the atlas selection criterion (now applied part-wise) for this purpose. The stopping criterion is also applied per part. So applying ALMAS to an unseen target image involves the following steps:

Input: T, A_r, A_{ij}, S_{ij} for $i = 1,\ldots,n$, $j = 1,\ldots,z$
Parameters: θ
Compute u_r, set $S^1 = u_r(S_r)$
For $j = 1,\ldots,z$ compute T_j, S_j and set $p_j = 1$, $m_j = 1$
For $i = 1,\ldots,n$
  For $j = 1,\ldots,z$
    Compute ω_{ij}
While ($\exists\, j$ for which $p_j > \theta$ and $m_j < n_j$) {
  Select j using g
  Select A_{ij} using f
  Compute u_{ij}
  Increment m_j, update S_j^m with $u_{ij}(S_{ij})$
  Compute p_j
}
S is now determined by combining all S_j^m for $j = 1,\ldots,z$

4. Experiments

4.1. Registration

In all experiments presented here, elastix[6] version 3.9 was used for registration. This framework formulates the registration problem as an optimization problem in which the similarity between

[6] elastix can be downloaded from http://elastix.isi.uu.nl.

the target and atlas image is maximized. The framework is largely based on techniques described in Thévenaz et al. (2000), Rueckert et al. (1999) and Mattes et al. (2003). Many parameters in the algorithm can be chosen. For the cost function the negative Mutual Information was used, which is implemented in `elastix` according to Thévenaz et al. (2000). First, images were roughly aligned with an affine transformation. This step comprised the fast registration algorithm. After that a non-rigid registration modeled by B-splines was performed, and this constituted the accurate registration algorithm. For the optimization of the cost function, an iterative stochastic gradient descent optimizer Klein et al. (2007) was used. In each iteration a step towards the minimum is taken, and this direction is based on the derivative of the cost function to the transformation parameters. To speed up the registration, the derivative is calculated on a small randomly chosen subset of samples in each iteration. To avoid local minima, a multi-resolution strategy was taken. Detailed parameter settings vary per application and are given below.

4.2. Heart segmentation in thoracic CT scans

Twenty-nine CT scans of the thorax were randomly taken from a lung cancer screening trial with low dose CT (30 mAs at 120 kV for patients weighing $\leqslant$80 kg and 30 mAs at 140 kV for those weighing over 80 kg). Data was acquired in spiral mode with 16×0.75 mm collimation. No contrast material was injected. Axial images of 1.0 mm thickness at 0.7 mm increment were reconstructed. All scans were reconstructed with a 512×512 matrix, yielding an axial resolution between 0.6 and 0.8 mm. The scans were randomly divided into two sets, a set of 15 atlases and a set of 14 target images. The heart was segmented by a medical student who was trained and supervised by a radiologist. Manual delineation of the heart was performed in axial slices by clicking points on the border; between two points a straight line was automatically drawn. This straight line could be changed by adding more points, or moving existing points. In addition, scrolling through the whole scan in all directions was possible while segmenting a slice. Complete segmentation of the heart in a single scan took about 90 min.

All 29 images were down-sampled by a factor of two in each dimension in order to reduce computation time of registration. Pilot experiments showed that using the full resolution data increased computation times by an order of six and hardly improved the results. For MAS and AMAS the settings of the registration parameters were as follows. For the affine registration four resolutions were used, in each of which 1000 iterations of the stochastic gradient descent optimizer were performed. For the non-rigid B-spline registration six resolutions were used. The B-spline grid spacing used in these resolutions was 128, 64, 32, 16, 8 and 4 voxels, respectively. The optimizer performed 256 iterations in the first three resolutions, 512 iterations in the fourth and fifth resolution and 1024 iterations in the last resolution. In each iteration, 3000 random samples were used to calculate the derivative of the cost function. For both affine and non-rigid registration, 32 histogram bins were used. For the other settings the default values of the registration program were applied. With these settings, one accurate registration takes about fifteen minutes on a standard high-end PC (Intel Core (TM)2, 2.40 GHz). A fast registration takes 25 s. For ALMAS, the registration parameters were altered based on the size of the parts. The number of samples to calculate the derivative of the cost function was scaled proportionally to the size of the part that was registered. The scaling was performed in such a way that the accurate registration of all parts separately took about fifteen minutes on a standard high-end PC.

Ten experiments were performed on the 14 target images applying different settings for MAS, AMAS and ALMAS. MAS was applied using all atlas images. In addition, three different settings were applied for ALMAS, AMAS and MAS, resulting in nine experiments. Since a setting in which all atlases are used is computationally expensive, three experiments were performed in which the segmentation times allowed were limited to 30 min, 45 min, and 1 h, respectively. To accomplish this for MAS, two, three and four atlases were used. For AMAS and ALMAS, the value for θ was manipulated to accomplish this. For AMAS θ was set 15%, 12%, and 10%, respectively, and for ALMAS θ was set to 25%, 20%, and 8% for each block. Since in MAS random atlases are selected, the algorithm was run 20 times randomly selecting the atlases that were used. All results reported for MAS with a limited number of atlases were averaged over 20 runs. For ALMAS, after registration of A_r and propagation of its labels, eight blocks were defined by dividing the volume of the resulting segmentation in two in each direction, so that each block contained exactly $\frac{1}{8}$ volume of the heart, and was guaranteed to contain the border of the heart. An overlap of 14 mm was taken into account for neighboring blocks. To obtain segmentation of the images in their original size, the segmentation results were super-sampled to the original resolution.

4.3. Caudate nucleus segmentation in brain MR images

Thirty-nine MRI images of the brain were downloaded from `http://www.cause07.org`. This web site provides the online continuation of a segmentation contest held at the 2007 conference on Medical Image Computing and Computer Assisted Intervention (van Ginneken et al., 2007). The site provides a set of training images with manual segmentations of the caudate nucleus as well as a set of target images for which no manual segmentations are provided. The performance on this test data can be obtained by submitting the results to the same web site. This way it is ensured that all submitted results are evaluated in exactly the same way and can therefore be directly compared.

All MRI images were scanned with an Inversion Recovery Prepped Spoiled Grass sequence on a variety of scanners (GE, Siemens, Phillips, mostly 1.5 Tesla). Some scans were acquired in axial direction, whereas others in coronal direction. All scans were reoriented to axial RAI-orientation, but had not been aligned in any fashion.

The test data came from three sources:

- 14 MRI images from the Psychiatry Neuro imaging Laboratory at the Brigham and Women's Hospital Boston (BWH).
- Five pediatric (no older than two years) MRI images acquired at the UNC Neuro Image Analysis Laboratory, Chapel Hill (UNC Ped).
- Five MRI images from a Parkinson Disease study at the UNC Neuro Image Analysis Laboratory, Chapel Hill (UNC Eld).

The resolution of the data sets was $0.9375 \times 0.9375 \times 1.5$ mm for the BWH data, $1 \times 1 \times 1$ mm for the UNC Eld data sets, and $1.01562 \times 1.01562 \times 1.01562$ mm for the UNC Ped data. The diversity of the data allows an evaluation of the algorithms with respect to flexibility to pathology, age group, and signal-to-noise ratio.

For training, 15 additional BWH data sets with manual delineations of the caudate nucleus were made available, these scans were used as atlases. Since there was no pediatric and elderly data available for training, we used the test data as atlases in a leave-one-out fashion; the scan that was used as a target image was removed from the set of atlases. We used the manual segmentation of the UNC data from a radiological expert to be able to use these scans as atlas. So, for the caudate segmentation in total we have 24 target images $(14 + 5 + 5)$, 24 atlases $(15 + 5 + 5 - 1)$ for the UNC data and 25 atlases $(15 + 5 + 5)$ for the BWH data. Note that we did not use a different set of atlases to segment the adult (BWH), pediatric, and elderly cases; there are 24 atlases available for every tar-

Table 1
Results for the heart segmentation for the different atlas-based methods in terms of overlap. Mean overlap and standard deviation (SD) between the reference standard and the automatic methods are provided. The numbers behind MAS refer to the number of atlases used, the numbers behind ALMAS and AMAS refer to the setting of θ.

225 min			60 min			45 min			30 min		
Method	Mean	SD	Method	Mean	SD	Method	Mean	SD	Method	Mean	SD
MAS	0.8793	0.0307	ALMAS8	0.8757	0.0249	ALMAS20	0.8727	0.02441	ALMAS25	0.8715	0.0231
			AMAS10	0.8694	0.0332	AMAS12	0.8694	0.0336	AMAS15	0.8671	0.0312
			MAS4	0.8680	0.0292	MAS3	0.8613	0.0283	MAS2	0.8320	0.0512

Table 2
Results for the heart segmentation for the different atlas-based methods in terms of average surface distance. Mean and standard deviation (SD) between the reference standard and the automatic methods are provided. The numbers behind MAS refer to the number of atlases used, the numbers behind ALMAS and AMAS refer to the setting of θ.

225 min			60 min			45 min			30 min		
Method	Mean	SD	Method	Mean	SD	Method	Mean	SD	Method	Mean	SD
MAS	2.1796	0.6044	ALMAS8	2.2265	0.5323	ALMAS20	2.2736	0.4454	ALMAS25	2.2950	0.4737
			AMAS10	2.3202	0.6138	AMAS12	2.3236	0.6407	AMAS15	2.5084	0.4454
			MAS4	2.3709	0.5650	MAS3	2.4717	0.5822	MAS2	2.9285	0.7687

get image from the UNC data and 25 atlases for every target image from the BWH data.

Before processing the images, the voxel intensities in all scans were normalized. In this paper, a simple intensity normalization was applied to assure images from different scanners get a similar intensity distribution. The method was implemented as follows: The intensities of all voxels within the brain were ordered. Only 90% of the voxels in the brain were used to compute the mean and standard deviation, leaving out 5% of the highest and lowest intensity voxels to be robust against outliers. Next, the image intensities were translated and scaled to have zero mean and unit standard deviation on those 90% of the voxels.

Due to the different nature and size of the MRI images as compared to the CT images used for the heart segmentation task, slightly different registration settings were applied. For the affine registration five resolutions were used, in each of which 600 iterations of the stochastic gradient descent optimizer were performed. For the non-rigid B-spline registration six resolutions were used. The B-spline grid spacing used in these resolutions was 128, 64, 32, 16, 8 and 4 voxels, respectively. In each iteration, 2000 samples were used to calculate the derivative of the cost function. The optimizer performed 300 iterations in the first three resolutions, 400 iterations in the fourth and fifth resolution and 600 iterations in the last resolution. For both affine and non-rigid registrations 32 histogram bins were used. With these settings, one accurate registration takes about seven minutes on a standard high-end PC, and a fast registration takes 15 s.

AMAS was applied to the 24 target images. The rationale behind ALMAS is that the atlas and target image will show local deviations around the border of the structure to segment. However, when the structure is very small, like the caudate nucleus, this is not applicable: any local deviation is a total deviation. Therefore, ALMAS was not expected to yield better results than AMAS and was not applied. AMAS was applied with the segmentation time limited to around 30 min, and to achieve this, θ was set to 20%.

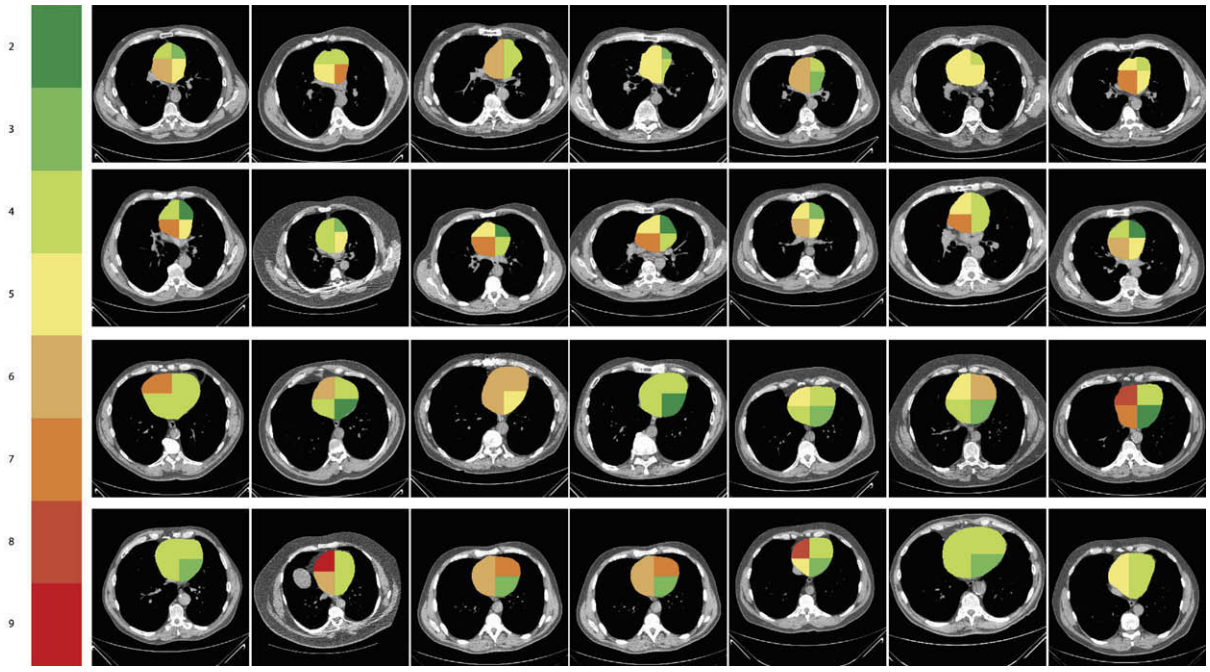

Fig. 3. Number of updates performed by ALMAS ($\theta = 8\%$) per block for each of the 14 test images. The top two rows show the blocks in the upper part of the heart for each test image, the last two rows show the blocks in the lower part of the heart. The color coding indicates how often each block was updated.

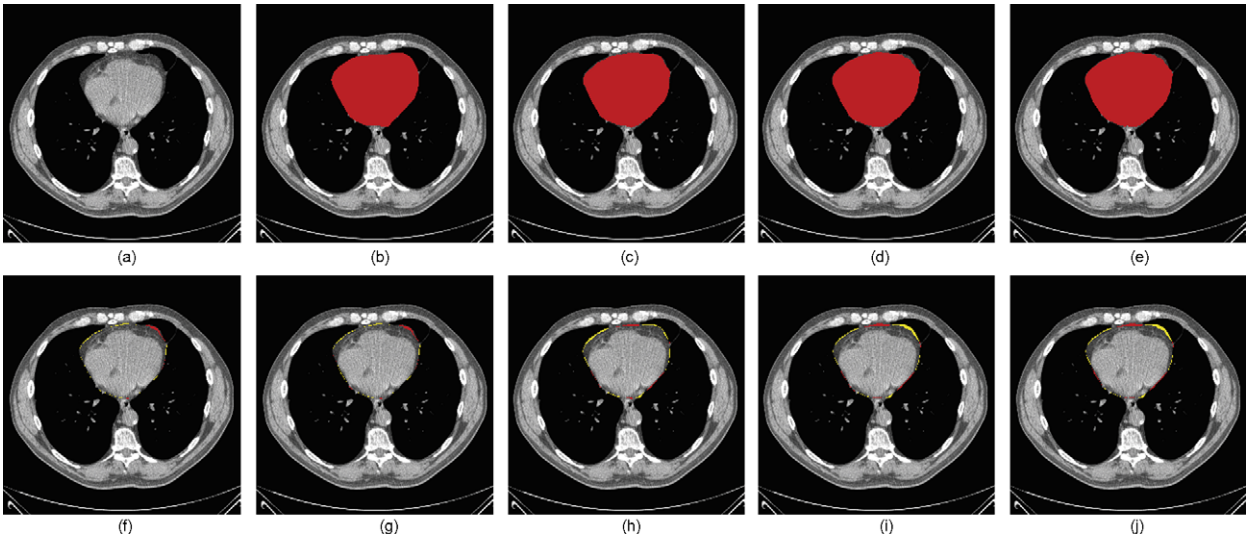

Fig. 4. Example of the results of segmentation of the heart for a slice of one scan. In (a) the original slice is shown, (b) shows the reference standard, in (c) the result of ALMAS8 (60 min) is shown, (d) shows the result of ALMAS20 (45 min), and (e) shows the result of ALMAS25 (30 min). In the second row difference images between the segmentations are shown, yellow indicates voxels that are in the first image but not in the second, red indicates the other way around. In (f) the difference image between ALMAS8 (60 min) and ALMAS20 (45 min) is shown. (g) shows the difference image between ALMAS8 (60 min) and ALMAS25 (30 min). (h), (i), and (j) show the difference images between the reference standard and ALMAS8 (60 min), ALMAS20 (45 min), and ALMAS25 (30 min), respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Results

5.1. Heart segmentation in chest CT scans

The performance of the different methods was evaluated by computing the volumetric overlap which is defined as the number of voxels in the intersection of automatic segmentation and reference standard divided by the number of voxels in the union. Table 1 gives for each method, grouped by required computation time, the average overlap on the target images and the standard deviation. In Table 2 for each method, the average surface distance and the standard deviation are provided.

Setting θ to 8%, ALMAS performed on average four updates per block (ranging from three to six average updates per block), which in terms of computation time is the same as performing four complete atlas registrations. Fig. 3 shows for each of the 14 target images the number of updates performed for each block by ALMAS with θ set to 8%. It can be seen that in each image, different blocks receive a different number of updates. In addition, Fig. 3 shows that each block gets different numbers of updates in different target images. From Fig. 3 it can be seen how often each version of g (Eq. 3) was applied: for each block that only received two updates, only the first version of g was applied. For all other blocks, the first version was applied twice, followed by the second version for all

Table 3
p-Values for the significance of the differences in overlap for the segmentation of the heart for the various methods. The p-values are from a two-tailed paired t-test. The numbers behind MAS refer to the number of atlases used, the numbers behind ALMAS and AMAS refer to the setting of θ.

	225 min	60 min			45 min			30 min		
	MAS	ALMAS8	AMAS10	MAS4	ALMAS20	AMAS12	MAS3	ALMAS25	AMAS15	MAS2
ALMAS8	0.15									
AMAS10	$p < 0.01$	0.05								
MAS4	$p < 0.01$	$p < 0.01$	0.7							
ALMAS20	0.1	0.2	0.5	0.42						
AMAS12	$p < 0.01$	0.07	0.82	0.97	0.46					
MAS3	$p < 0.01$	$p < 0.01$	0.1	$p < 0.01$	$p < 0.05$	0.11				
ALMAS25	0.14	0.27	0.69	0.61	0.79	0.61	0.17			
AMAS15	$p < 0.05$	0.1	0.81	0.86	0.31	0.86	0.54	0.42		
MAS2	$p < 0.01$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.01$	$p < 0.05$	$p < 0.05$	$p < 0.01$	$p < 0.01$	

Table 4
Results of AMAS and the other three best ranked atlas-based segmentation methods; average overlap (OV) and the total score (TS) are provided for each method for each data set as well as in total. The names of the methods are taken from the segmentation competition website; I3A (UZ) provides no reference to a publication, Uath-UNC refers to (Gouttard et al., 2007) and ISICAD refers to (van Rikxoort et al., 2007).

Method	AMAS		I3A(UZ)		Uath-UNC		ISICAD	
	OV	TS	OV	TS	OV	TS	OV	TS
UNC Ped	0.7462	84.14	0.6366	82.09	0.6286	74.00	0.4747	54.24
UNC Eld	0.6685	81.87	0.5740	73.27	0.5814	65.97	0.5041	66.61
BWH	0.6761	76.41	0.6359	71.51	0.7037	72.02	0.6979	78.28
Total	0.6891	79.16	0.6440	74.08	0.6626	71.17	0.6110	70.84

other updates. Fig. 4 shows an example output of ALMAS for all settings of θ for a slice of one target image.

To compare the performance of the different methods, two-tailed paired t-tests were performed. Table 3 lists the results. For all settings of θ, ALMAS is not significantly different from MAS using 15 complete atlas scans, which requires 225 min per target scan. When segmentation time is limited to 60, 45 or 30 min, ALMAS performs significantly better than MAS. When segmentation

Table 5
Selected atlas images, in each iteration, for each target image in the caudate segmentation task.

Target	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
UNC Ped 10	UNC Ped 14	UNC Ped 30	UNC Ped 15		
UNC Ped 14	UNC Ped 15	UNC Ped 30	UNC Ped 19	UNC Ped 10	
UNC Ped 15	UNC Ped 30	UNC Ped 14	UNC Ped 19		
UNC Ped 19	UNC Ped 14	UNC Ped 15	UNC Ped 30	UNC Ped 10	
UNC Ped 30	UNC Ped 15	UNC Ped 14	UNC Ped 10	UNC Ped 19	
UNC Eld 01	UNC Eld 26	UNC Eld 13	UNC Eld 12	UNC Eld 20	
UNC Eld 12	UNC Eld 13	UNC Eld 01	UNC Eld 26	UNC Eld 20	
UNC Eld 13	UNC Eld 01	UNC Eld 26	UNC Eld 12	UNC Eld 20	
UNC Eld 20	UNC Eld 12	UNC Eld 13	UNC Eld 26	UNC Eld 01	
UNC Eld 26	UNC Eld 13	UNC Eld 01	UNC Eld 12	UNC Eld 20	
BWH 16	BWH 13	BWH 04	BWH 11	BWH 06	
BWH 17	BWH 14	BWH 02	BWH 13	BWH 09	
BWH 18	BWH 13	BWH 14	BWH 08	BWH 12	
BWH 19	BWH 08	BWH 02	BWH 14	BWH 09	
BWH 20	BWH 13	BWH 14	BWH 04	BWH 10	BWH 09
BWH 21	BWH 13	BWH 11	BWH 04	BWH 02	
BWH 22	BWH 03	BWH 12	BWH 11	BWH 13	BWH 14
BWH 23	BWH 14	BWH 02	BWH 05	BWH 08	
BWH 24	BWH 01	BWH 05	BWH 09	BWH 02	
BWH 25	BWH 02	BWH 08	BWH 04	BWH 09	
BWH 26	BWH 02	BWH 09	BWH 05	BWH 08	
BWH 27	BWH 14	BWH 13	BWH 08	BWH 02	
BWH 28	BWH 14	BWH 02	BWH 05	BWH 13	
BWH 29	BWH 02	BWH 08	BWH 09	BWH 14	

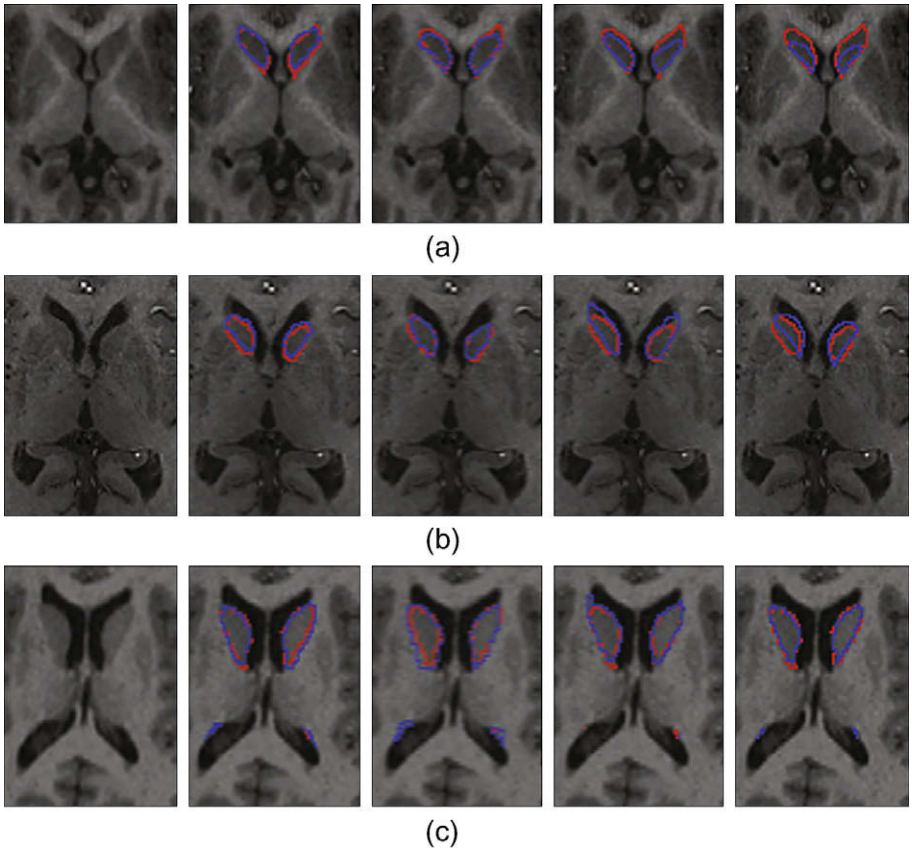

Fig. 5. Example output of AMAS and the three other highest ranked atlas-based segmentation methods for the caudate segmentation competition on (a) a scan from the UNC Ped data set; (b) a scan from the UNC Eld data set; (c) a scan from the BWH data set. The output of the automatic methods is given in red, the ground truth is in blue. Images were taken from http://www.cause07.org. The first column provides the original slice, the second column shows the result of AMAS, the third column shows the result produced by I3A(UZ), in the fourth column the result produced by Gouttard et al. (2007) is given, the last column shows the result of van Rikxoort et al. (2007). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Results of AMAS, MAS, and ISICAD (28-10-2007).

Method	AMAS		MAS		ISICAD	
	OV	TS	OV	TS	OV	TS
UNC Ped	0.7462	84.14	0.5574	63.07	0.4747	54.24
UNC Eld	0.6685	81.87	0.5680	70.66	0.5041	66.61
BWH	0.6761	76.41	0.5860	61.35	0.6979	78.28
Total	0.6891	79.16	0.5763	63.65	0.6110	70.84

time is restricted to 30 min AMAS also performs significantly better than MAS.

5.2. Caudate nucleus segmentation in brain MR images

The results reported here were obtained by submitting our segmentations to the web site of the segmentation contest (http://www.cause07.org) on October 07, 2008 under the team name ISICAD. The results are listed online with team name and date and it can be seen that the results of our method are substantially better than those of all previously submitted methods.

The segmentation contest uses five performance measures, and these are combined into a single overall score. This score is related to the result that could be expected if an independent human observer would perform the segmentation manually; a score of 100 points indicates a perfect result, a score of 90 points is typical for an independent human observer (van Ginneken et al., 2007). The overlap value, as used for the heart segmentation task, is one of the five measures. Table 4 lists the results (overlap and score) of AMAS as well as the results of the other three best ranked atlas-based methods from http://www.cause07.org.

Note that one of the atlas-based submissions is our original submission (van Rikxoort et al., 2007) to the segmentation contest workshop. In this case MAS was applied with the appropriate atlases manually selected for each data set.

Setting θ to 20%, AMAS performed on average 4 updates per scan. Table 5 lists the selected atlases for each target image. It can be seen that for each target image only atlases from the same group (adults in the BWH data, pediatrics and elderly) were selected by the automatic atlas selection strategy.

Fig. 5 illustrates the results of the four best atlas-based methods currently submitted to the segmentation competition for one scan of each source. AMAS performs consistently on the scans from all three sources, where the other methods show deviations from the reference in at least one of the data sets.

6. Discussion and conclusion

Medical imaging data exhibits tremendous variation. Patient data often contains abnormalities but even healthy, normal anatomy differs widely in young and old, male and female, slim and obese individuals. Moreover, imaging protocols and scanner characteristics vary between and even within institutions. This poses major challenges for developers of segmentation algorithms. It is common practice to develop and test methods on data from a single source. This leads to results that look good on paper (i.e. when tested on data from the same distribution from which the training data originated) but disappoint in practice. This is one of the main reasons why so few published methods make the transition from the research lab to the clinic.

The caudate segmentation task addressed in this work is a good example. In practice, most labs still manually segment brain structures, despite a plethora of published algorithms. Many of these publications acknowledge the limitation of their proposed method and even remark that a method could be applied to, say, pediatric scans, if retrained with suitable training data. The problem is that even if different training data, and thus multiple systems, were available, it is not trivial to decide for a particular target scan which version of the method to apply. In fact, a close look at the descriptions of some of the methods that participated in the online segmentation contest at http://www.cause07.org reveals that several of them, including our original contribution van Rikxoort et al. (2007), while claiming to be fully automatic, actually manually select different training sets to segment the different groups of test data. This makes them semi-automatic.

The AMAS framework proposed in this study has been designed to overcome the atlas selection problem. During the segmentation process, the atlases most suitable to segment the target image under consideration are automatically selected. Atlases that are different from the target image are not taken into account, but are still available for segmenting other test scans: AMAS adapts to each target scan individually. Although the technique we use to let AMAS make these decisions is fairly simple, results listed in Table 5 show that the system behaves as would be expected: only adult datasets were used to segment adult scans, pediatrics were used for pediatrics and elderly atlases for elderly target scans. Some scans, such as UNC Ped 14, were apparently representative of the test data and were selected often whereas others, such as BWH 06, were only chosen once, in a final iteration. So, by design, AMAS can deal with test data of a different nature, as long as the training data contains some scans representative of each test scan. AMAS vastly outperforms our previous multi-atlas segmentation scheme (van Rikxoort et al., 2007), that used the same registration framework, and does so at a substantially lower computational cost (30 min). The method in van Rikxoort et al. (2007) is the only one among the atlas-based methods listed in Table 4 that achieves a higher score on one of the three subgroups, namely the adult (BWH) cases. The reason for this is that van Rikxoort et al. (2007) used all 15 BWH training cases to segment the test cases, whereas ALMAS used at most five (Table 5). The results listed at http://www.cause07.org show that AMAS also outperforms a wide variety of other state-of-the-art techniques for brain structure segmentation, such as voxel classification, m-reps and active appearance models. An interesting topic for further research is to apply AMAS to other brain structures.

To show that the improved results of AMAS as compared to our previous method van Rikxoort et al. (2007) are not due to the addition of the training data for UNC Ped and UNC Eld, we generated the results of MAS using all 25 atlases used in this paper. The results are listed in Table 6. It can be seen that for the elderly and pediatric data, the results of the old method improve, but for the BWH data the results decline.

Just as there is no reason to assume that a particular training scan would be optimal for segmenting *any* test scan, there is no reason to believe that a scan that lines up well with a particular part of a test scan, would do so *everywhere*, throughout the complete volume. This led us to develop ALMAS, where the choice how many and which atlases to use is made *locally*. Although the differences in performance between AMAS and ALMAS for the heart segmentation task are not big, ALMAS seems to have a competitive advantage, especially when more updates are allowed ($p = 0.05$ for ALMAS8 versus AMAS10 in Table 3).

The idea behind ALMAS is that 'difficult' parts of the image will get more attention than 'easier' parts. For segmentation of the heart, the average number of updates per block ranged from three to six, with the lowest number of updates being two and the highest nine. This indicates that the hypothesis behind ALMAS works out in practice. This is illustrated in Fig. 3 which shows that in each target image for the segmentation of the heart, the number of updates differs per block, in addition, the number of updates for each block in the different images varies.

Lower computational costs to achieve the same performance is a reason to prefer AMAS and ALMAS over MAS. The experiments performed for the segmentation of the heart clearly showed that when segmentation time is limited, ALMAS is preferable to MAS. In addition, when segmentation time was not limited, MAS was not significantly better than ALMAS, where MAS takes 225 min per scan and ALMAS 60, 45 or 30 min. Note that AMAS and ALMAS register the most promising scans first, therefore they could be stopped at any moment to produce an optimal result for the allocated amount of computation time. The lower computational cost of AMAS and ALMAS as compared to MAS are based on sequential processing. In the case of parallel processing, computational complexity is a different issue. In MAS, atlases can be registered in parallel. One could parallelize AMAS and ALMAS to a certain extent by first registering x selected atlases in parallel, where x is the number of cores available.

AMAS and ALMAS should be interpreted as *frameworks* to apply atlas-based segmentation. Any registration method, atlas selection method, stopping criterion and strategy for dividing a scan into smaller blocks can be plugged in. The atlas selection as applied for heart and caudate segmentation is monomodal; only target images from the same image modality as the atlases can be used since otherwise the difference image is meaningless. For a multimodal application, the atlas selection could for example be based on NMI after affine registration. In our experiments for heart segmentation, a straightforward division of the heart into eight blocks of equal size was used. It might be better for certain applications to base the subdivision on anatomy. The number and size of the blocks is clearly an interesting variable in the framework. When very small blocks are defined, a locally affine registration might be sufficient which will further speed up the algorithm. On the other hand, blocks probably require some overlap to avoid visible transitions in the segmentation result and this will slow down the algorithm when very small blocks are used. When small blocks are defined, registration becomes more difficult since less information is available; the estimation of the mutual information using histograms might not be valid anymore. In those cases, another similarity measure, for example SSD, could be preferable.

Another open question is how AMAS and ALMAS will behave when the number of atlas scans is very large. We hypothesize that ALMAS will prove especially useful for tasks where target scans contain anatomical variations or pathologic abnormalities. For such tasks, many more atlases than were used here might be needed, to ensure that for each location in each target image, a few similar atlas blocks exist. Note that in this scenario, the computational requirements of MAS are prohibitive. But also AMAS and ALMAS may need computationally less expensive ways to select promising atlases. For example, a greedy selection scheme, that would avoid having to register every atlas to the target, which took around 25 s for the fast registration in the heart application.

In summary, a new adaptive local multi-atlas segmentation framework was proposed that selects optimal atlases and can do that locally if desired. In an experiment on heart segmentation in CT data, the method performed as well as a standard multi-atlas method that was computationally much more expensive. In an additional experiment on brain MRI data it was shown that the atlas selection method was able to automatically select the most appropriate atlas scans to segment data from children, adults and elderly. The method achieved better results than many other segmentation techniques.

References

Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J.V., Rueckert, D., 2007. Classifier selection strategies for label fusion using large atlas databases. In: Tenth International Conference on Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science, vol. 4791, pp. 523–531.

Crum, W.R., Griffin, L.D., Hill, D.L.G., Hawkes, D.J., 2003. Zen and the art of medical image registration: correspondence, homology, and quality. NeuroImage 20 (3), 1425–1437.

Gouttard, S., Styner, M., Joshi, S., Davis, B., Smith, R.G., Hazlett, H.C., Gerig, G., 2007. Subcortical structure segmentation using probabilistic atlas priors. In: Proceedings of MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge, pp. 37–46.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Human Brain Mapping 19 (4), 224–247.

Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33 (1), 115–126.

Jain, A.K., Zongker, D., 1997. Feature selection: evaluation, application and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (2), 153–158.

Jain, A.K., Duin, R.P.W., Mao, J., 2000. Statistical pattern recognition: a review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (1), 4–37.

Klein, S., Staring, M., Pluim, J.P.W., 2007. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. IEEE Transactions on Image Processing 16 (12), 2879–2890.

Klein, S., van der Heide, U.A., Lips, I.M., van Vulpen, M., Staring, M., Pluim, J.P.W., 2008. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Medical Physics 35 (4), 1407–1417.

Mattes, D., Haynor, D.R., Vesselle, H., Lewellen, T.K., Eubank, W., 2003. PET-CT image registration in the chest using free-form deformations. IEEE Transactions on Medical Imaging 22 (1), 120–128.

Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr., C.R., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage 21 (4), 1428–1442.

Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D.B., Maurer Jr., C.R., 2005. Quo vadis, atlas-based segmentation? In: The Handbook of Medical Image Analysis – Volume III: Registration Models. Kluwer Academic/Plenum Publishers, New York, NY, pp. 435–486.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18 (8), 712–721.

Thévenaz, P., Blu, T., Unser, M., 2000. Image interpolation and resampling. In: Handbook of Medical Imaging, Processing and Analysis. Academic Press, San Diego, CA, pp. 393–420.

van Ginneken, B., Heimann, T., Styner, M., 2007. 3D segmentation in the clinic: a grand challenge. In: Proceedings of MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge, pp. 7–15.

van Rikxoort, E.M., Arzhaeva, Y., van Ginneken, B., 2007. A multi-atlas approach to automatic segmentation of the caudate nucleus in MR brain images. In: Proceedings of MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge, pp. 29–36.

Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C.S., Aizenstein, H.J., 2007. Optimum template selection for atlas-based segmentation. NeuroImage 34 (4), 1612–1618.