A novel atlas-selection approach for multi-atlas based segmentation using the correlation of inter-atlas similarities

P. Raudaschl[1], K. Fritscher[1, 2], P. Zaffino[3], G.C. Sharp[4], M.F. Spadea[3], R. Schubert[1]

[1] Department for Biomedical Image Analysis, UMIT, Austria
[2] Medphoton, Institute for Technology Development of Radiotherapy, University Clinic of Radiotherapy and Radio Oncology Salzburg, Austria
[3] Department of Experimental and Clinical Medicine, Magna Graecia University, Italy
[4] Massachusetts General Hospital, Harvard Medical School, USA

Abstract. Automated segmentation is a frequently applied task in the course of medical imaging. Furthermore, it is a substantial component of image-guided radiotherapy. Atlas based segmentation is one of the most frequently used approach for automated segmentation. Especially for multi-atlas based segmentation, segmentation quality and speed largely depends on the underlying registration and atlas selection strategy. In this work an atlas selection strategy that is based on the correlation of inter-atlas similarities within a set of atlas images is presented. Segmentation quality is analyzed by calculating dice coefficients and 95% Hausdorff distances for the left and right parotid with respect to different numbers of atlases. Results are compared to other state of the art atlas selection strategies. It can be shown that the developed atlas selection technique performs slightly better than NMI-based selection if a low number of atlases is used.

Keywords: Atlas-based segmentation, Atlas selection, Normalized mutual information, Deformable image registration

1 Introduction

Image segmentation is a frequently applied task in the course of medical imaging. There exist a variety of different segmentation approaches whereas the choice of the most suitable approach mainly depends on the question to be answered and on the underlying image data. Basically, segmentation techniques can be divided into manual, semi-automated and automated approaches. Especially in radiotherapy automated segmentation is widely used. Automated segmentation is a substantial component of image-guided radiotherapy (e.g. for segmentation of organs at risk). In order to accomplish a robust and reliable automated segmentation, a priori knowledge of structures that should be segmented is necessary. In case of atlas-based segmentation this

knowledge is available through already segmented atlas images. The segmentation of structures in new images is performed by registering these new images to an already segmented image (subject). A registration on several atlas images – also referred to as multi-atlas based registration - is also possible. After registration, voting schemes that mark single voxels as being inside or outside a segmented object are used. Due to high inter- and intra-subject variability of different structures, multi-atlas based segmentation approaches have shown to be more accurate than single-atlas based segmentation attempts [1]. Furthermore, non-systematic registration errors can be reduced therewith. Generally, the segmentation quality of atlas-based segmentation depends on the registration accuracy. In the case of multi-atlas based segmentation the selection of an appropriate voting strategy, but also the selection of atlases has an additional important influence on the segmentation accuracy. In addition, the selection of an appropriate subset of atlases can significantly improve the speed of the segmentation.
A number of different atlas selection strategies have been published. In [2] Rohlfing et al. compared segmentation approaches based on using a single individual atlas image, an average shape atlas image, the most similar atlas image and the application of all available atlas images with subsequent multi-classifier decision fusion. Their results have shown that the multi-classifier approach works best in most cases.
Aljabar et al. [3] analyzed the most common atlas selection strategies for the segmentation of brain structures in MRI images.

- **Atlas selection based on segmentation similarity:** Atlas selection according to a ranking that results of dice overlaps between manually segmented images and atlas images. In fact, this is an unrealistic approach, since dice overlap is unknown when starting the segmentation process, but it is often used as a reference for comparison with other atlas selection strategies.
- **Atlas selection based on image similarity:** The similarity between new images and atlas images is determined by using specific image similarity metrics. Based on the similarity values a ranking can be built that can be used for atlas selection. The most common similarity metrics are: sum of squared distances, cross correlation, mutual information and normalized mutual information.
- **Atlas selection based on demographics:** Atlas ranking according to non-image information (e.g. age or sex).

For a test dataset of 275 datasets, atlas selection based on normalized mutual information performed best.

In this work a new atlas selection strategy is presented that is based on image similarity. A correlation analysis is performed on inter-atlas similarities which are represented by normalized mutual information. On the basis of the results of this correlation analysis an atlas ranking is generated for multi-atlas segmentation.

2 Methods

2.1 Overview

The basis is an atlas database that is comprised of 18 rigidly registered CT datasets with manually segmented contours of the left and right parotid. In a first step image similarity based on mutual information between all pairs of atlases is calculated. Then the correlation between the resulting similarity values of each atlas to all remaining atlases of the test dataset is determined. Subsequently, an atlas ranking is built from the correlation values for each atlas which is used for atlas selection. After performing deformable B-Spline registration of all selected atlases to the new image, the label fusion approach proposed by Peroni [4], which is based on the method presented by Sabuncu [5] was applied to obtain a final registration result.

2.2 Image Similarity

The first step in the segmentation pipeline is the calculation of image similarities between all pairs of atlases. Normalized mutual information (NMI) is used as image similarity metric. MI is an entropy-based measure and it is based on the assumption that two images have statistical dependencies. MI expresses how much the uncertainty of one image decreases if the other image is known. MI is defined as [6]:

$$MI(A,B) = \mathrm{H(A)} + \mathrm{H(B)} - \mathrm{H(A.B)} \quad (1)$$

H(A) and H(B) is the Shannon entropy of image A and B, respectively. H(A,B) is the joint entropy of the two images. In order to achieve overlap invariance normalized mutual information is used. NMI is defined as:

$$NMI(A,B) = \frac{H(A) + H(B)}{H(A.B)} \quad (2)$$

2.3 Atlas Selection Strategy

Image similarities based on NMI also form the basis for further analysis which leads to the ranking of atlas images for a new image to be segmented.
Commonly, NMI-based selection approaches for multi-atlas based segmentation are based on similarities (expressed by NMI) between the new image to be segmented and all other atlas images. In the newly developed approach, these similarities also form the first column and row of a selection matrix S. In addition, the similarities among all atlas images based on NMI values are used and entered into matrix S (see also Figure 1). Based on the Pearson correlation (Figure 1: last row) of inter-subject NMI values (=column 1 of matrix S and columns 2..n (n=number of atlases)) a ranking for all atlases can be obtained (rows containing empty matrix entries were not taken into account for the calculation of the correlation). By this means, not the direct similarity between two images (in this case new image and atlas image), but the rela-

tionship (in terms of image similarities) of an image to an ensemble of other (atlas) images is used to find most appropriate atlases. Note, that in the presented approach NMI calculation was restricted to a three dimensional box like region around the parotid glands in the respective CT images. Furthermore, only voxels within a greyscale range of -200 to +200 were used for NMI calculation (the choice of these threshold values is based on statistics concerning voxel intensities within the parotid gland and surrounding soft tissue). Based on these facts and due to the underlying implementation of NMI calculation in the "Insight Segmentation and Registration Toolkit (ITK) [7]" the selection matrix S is not symmetric.

	NI	**Atlas 2**	**Atlas 3**	**Atlas 4**	...	**Atlas n**
NI		NMI NI-2	NMI NI-3	NMI NI-4	...	NMI NI-n
Atlas 2	NMI 2-NI		NMI 2-3	NMI 2-4	...	NMI 2-n
Atlas 3	NMI 3-NI	NMI 3-2		NMI 3-4	...	NMI 3-n
Atlas 4	NMI 4-NI	NMI 4-2	NMI 4-3		...	NMI 4-n
...	...	...	...	...	...	...
Atlas n	NMI n-NI	NMI n-2	NMI n-3	NMI n-4	...	
Correl		***Column NI-2***	***Column NI-3***	***Column NI-4***	...	***Column NI-n***

Fig. 1. Creation of the selection matrix S and calculation of Pearson correlation. (NI – New Image)

2.4 Evaluation Strategy

In order to evaluate the segmentation results of the developed approach the results were compared with the following atlas selection strategies:

- **Atlas selection based on direct image similarity:** According to NMI values between the datasets a ranking was generated. This selection approach is common and widely used.
- **Oracle selection:** The dice values after segmentation were used to build an atlas selection ranking. In reality this configuration is not possible, however, for testing purposes it is often used as a reference.

Besides the order of the atlas ranking, the number of the used atlases was also analyzed. Therefore segmentations with 3, 5, 7, 9, 11, 13 and 15 atlases were performed.
A leave-one-out strategy was used to evaluate and compare the different atlas selection approaches. For each subject, the remaining subjects were considered as potential atlases. The dice coefficient was used to quantitatively evaluate the segmentation accuracy. Hence, the overlap of segmented structures within the images can be analyzed. The dice coefficient is defined as [8]:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

A and B are labeled regions that are compared. Volumes are represented by the number of voxels. Furthermore, the 95% Hausdorff distance was also used for the analyses. Commonly Hausdorff distance is defined as [9]:

$$HD(A,B) = \max\big(h(A,B), h(B,A)\big) \quad (4)$$

where

$$h(A,B) = \max_{a \in A} \min_{b \in B} ||a - b|| \quad (5)$$

|| . || is a norm on the points of A and B. In contrast to the calculation of the maximum Hausdorff distance, 95% Hausdorff distance is based on the calculation of 95th percentile of the Hausdorff distance.
The average dice coefficients and the average 95% Hausdorff distances of the leave-one-out strategy were calculated with respect to a varying numbers of atlases.

2.5 Data and Tools

The atlas database consists of 18 subject's CTs of the head and neck. All images have a voxel size of 0.539 x 0.539 x 2.5 mm and a dimension of 512 x 512 x 89 slices. Left and right parotids were manually delineated by experts.
For atlas-based segmentation the software toolkit Plastimatch was used [10]. Plastimatch is an open source software package for image computation. It provides a sophisticated algorithm for multi-atlas based segmentation (MABS). All components of MABS can be configured separately (e.g. registration and optimization parameters). Hence, it is well-suited for the underlying question and the associated analyses.

3 Results

3.1 Average Dice Coefficient

In Figure 2 and Figure 3 the average dice coefficient of the left and right parotid of the leave-one-out analysis with respect to different numbers of atlases is visualized. It can be seen that the average dice coefficient distinctly rises with increasing numbers of atlases. Starting from 11 atlases and higher a constant plateau is reached. In general, (unrealistic) oracle selection based on dice overlap performed best for all test cases. For the left parotid, the new correlation-based atlas selection strategy performed better than the NMI-based approach for 3, 7, 9, 11 and 15 atlases. For the right parotid our newly introduced method outperformed NMI for 5, 7, 9, 11, 13 and 15 atlases. It is also obvious that the correlation-based method outperforms NMI if the number of atlases used is between 7 and 9. For higher numbers of atlases the NMI- and correlation-based approach are similarly with respect to average dice coefficient.

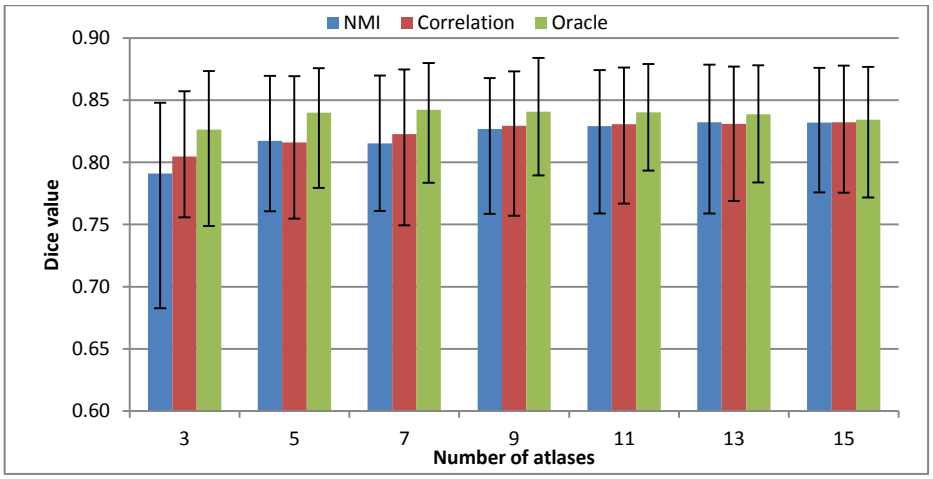

Fig. 2. Average dice coefficient of the left parotid for different numbers of atlases. Colored bars represent average dice coefficient (leave-one-out approach) of different atlas selection strategies. Whiskers indicate 0.95 and 0.05 percentiles.

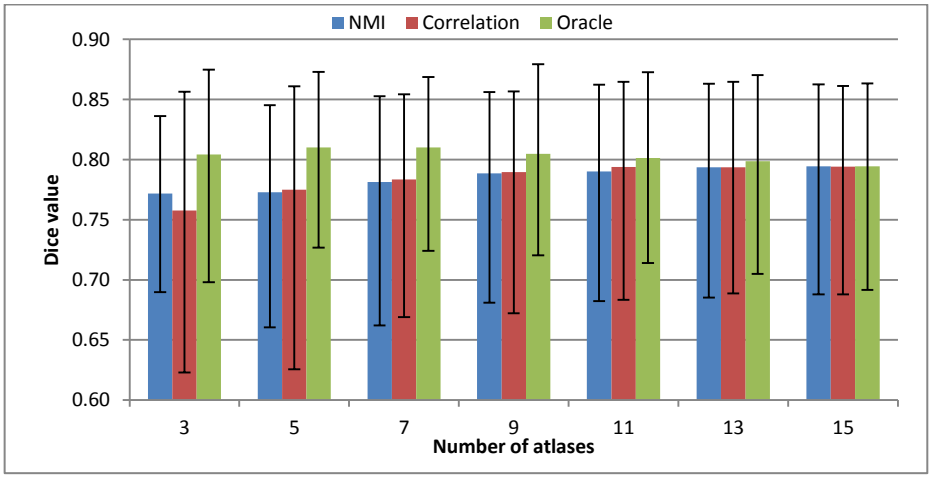

Fig. 3. Average dice coefficient of the right parotid for different numbers of atlases. Colored bars represent average dice coefficient (leave-one-out approach) of different atlas selection strategies. Whiskers indicate 0.95 and 0.05 percentiles.

3.2 Average 95% Hausdorff Distance

In Figure 4 and Figure 5 the average 95% Hausdorff distance of the left and right parotid with respect to different numbers of atlases is depicted. For the right parotid the new correlation-based selection strategy performs better than the NMI ranking for 3, 5, 7, 9, 11 and 15 atlases. For the left parotid the correlation-based method outper-

formed the NMI-based method for 3 and 11 atlases. From 13 atlases and more the changes of the average 95% Hausdorff distance are small.

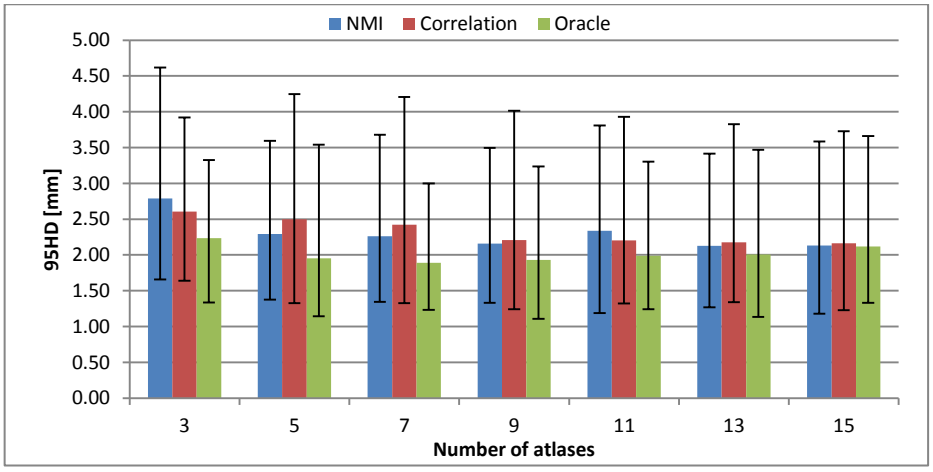

Fig. 4. Average 95% Hausdorff distance in mm of the left parotid for different numbers of atlases. Colored bars represent average Hausdorff distance (leave-one-out approach) of different atlas selection strategies. Whiskers indicate 0.95 and 0.05 percentiles.

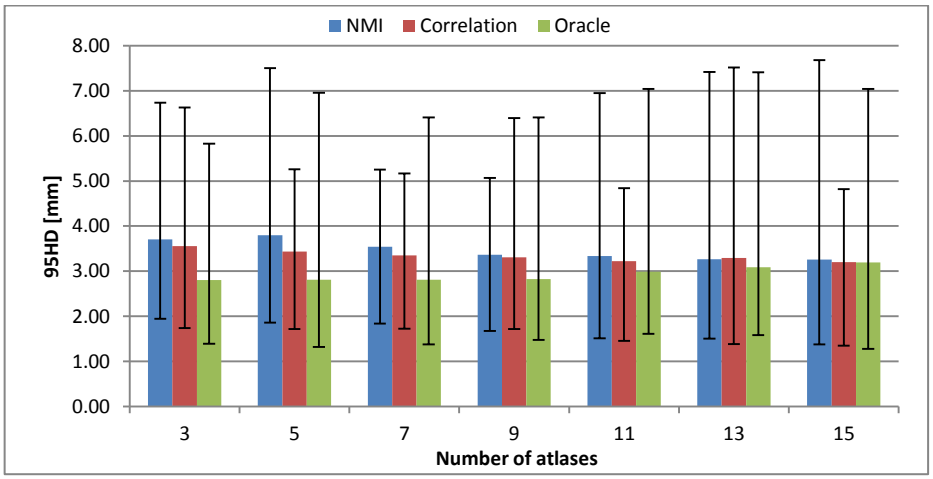

Fig. 5. Average 95% Hausdorff distance in mm of the right parotid for different numbers of atlases. Colored bars represent average Hausdorff distance (leave-one-out approach) of different atlas selection strategies. Whiskers indicate 0.95 and 0.05 percentiles.

4 Discussion

The application of image similarity expressed by NMI as atlas selection strategy is a common approach in the course of multi-atlas based segmentation. The developed

atlas selection strategy extends this approach by calculating the correlation of inter-atlas similarities. By this means, atlas ranking and selection can be performed.
The results have shown that the correlation-based selection strategy performs slightly better than common NMI-based selection concerning average dice overlapping. The average dice coefficient of the correlation-based method was higher in 11 of 14 cases. The advantages of the developed method are especially obvious if only a low number of atlases are used for segmentation/registration. This fact is of great importance if the speed of the segmentation is an issue. For a high number of atlases, NMI- and correlation-based atlas selections perform similarly. Regarding 95% Hausdorff distance, the advantages of the correlation-based method were smaller. With respect to the left parotid, NMI-based approach had lower average 95% Hausdorff distances. However, when looking at the percentiles, this relativizes. For the right parotid, the new correlation-based strategy had again a better performance, especially if a low number of atlases are used for segmentation.
Concluding it can be said, that the developed atlas selection technique performs slightly better or at least equal compared to NMI-based selection. In further analyses special characteristics of the developed atlas selection strategy also in combination with other atlas selection strategies will be investigated.

5 References

1. van Rikxoort, E.M., Isgum I., Arzhaeva, Y., Staring M., Klein S., Viergever, M.A., Pluim, J.P.W. and van Ginneken, B.: *Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus.* Medical Image Analysis, Vol. 14, No. 1, pp. 39-49, (2010).
2. Rohlfing, T., Brandt, R., Menzel, R. and Maurer, Jr., C.R.: *Evaluation of Atlas Selection Strategies for Atlas-Based Image Segmentation with Application to Confocal Microscopy Images of Bee Brains.* NeuroImage, Vol. 21, No. 4, pp. 1428-42, (2004).
3. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J.V., Rueckert, D.: *Classifier selection strategies for label fusion using large atlas databases.* Med Image Comput Assist Interv. Vol. 10, No. 1, pp. 523-531, (2007).
4. Peroni, M. *Methods and Algorithms For Image Guided Adaptive Radio- and Hardon-Therapy.* PhD thesis, Politecnico di Milano, (2011).
5. Sabuncu, M.R., Thomas Yeo, B.T., Leemput, K.V., Fischl, Br. and Golland, P.: *A generative model for image segmentation based on label fusion.* IEEE Trans Med Imaging, Vol. 29, No. 10, pp. 1714–1729, (Oct 2010).
6. Studholme, C., Hill, D.L.G. and Hawkes, D.J.: *An overlap invariant entropy measure of 3D medical image alignment.* Pattern Recognition. Vol. 32, No. 1, pp. 71-86, (1999).
7. Kitware. http://www.itk.org/. Last visited: 2014-06-17.
8. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V. and Rueckert, D.: *Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy.* NeuroImage. Vol. 46, No. 3, pp. 726-738, (2009).
9. Huttenlocher, D.P., Klanderman G.A. and Rucklidge, W.J.: *Comparing Images Using the Hausdorff Distance.* Pattern Analysis and Machine Intelligence, IEEE Transactions. Vol. 15, No. 9, pp. 850-863, (1993).
10. Plastimatch. http://www.plastimatch.org. Last visited: 2014-06-17.