Mini-project: Medical image segmentation

- Group 6 -

R.J.P. van Bergen (1234032), N.N.H.C. Bodenstaff (1223152), S.H.A. Verschuren (1245272)

Technical University Eindhoven, Dept. of Biomedical Engineering

October 14, 2019

Contents

1 Introduction

Image guided surgery is becoming a more commonly used technique worldwide [1]. Real-time digital images can play a role in deciding what the best options are for the surgeon. One of the challenges for this upcoming technology is the segmentation of images to distinguish different tissues. As digital images contain noise, it can be hard to know with certainty what kind of tissue each subarea represents. Hence, well generalized computer programs will need to be applied. Different features in images provide a basis for creating a method to segment tissues. However finding the best features to separate classes can be a demanding.

This paper investigates what methods are adequate for creating a segmentation method. This was accomplished by using multiple slices of brain images from different subjects, weighted with T1- and T2-sequences [2]. Different features extracted from the images are compared with each other and valued based on which combination of features provides the best segmentation of tissues. Generalization is the desired outcome, as it should be applicable to any brain image provided. Classification error and dice overlap are used for evaluation of the segmentation methods.

2 Methods

For segmenting images (in feature space) several methods can be used. Namely k-means clustering (KM) and k nearest neighbour (K-nn). KM segments the image based on locating k cluster centers in such a way that the distance between all points and their respective cluster center is minimised [3]. This method, however, is not used in this project due to the significantly higher computing times and classification errors found in the prototyping phase. K-nn compares an image (test set) to a already segmented image (train set). Based on this comparison the labels are computed though a majority vote of the k nearest neighbours [4].

The results of segmentation methods are rated based on the resemblance to the ground truth utilising the Sørensen–Dice coefficient (Dice) and classification error. The dice score is defined by formula 1 [5].

$$DSC = 2 * TP/(2 * TP + FP + FN) \qquad (1)$$

The classification error is defined by formula 2 [6].

$$E = (FP + FN)/(TP + TN + FP + FN) \qquad (2)$$

In formula 1 and 2, TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives respectively.

2.1 Dataset

The data sets for the images are made using provided code (segmentation_demo()). This code has as input the slice numbers and image locations and as output the features and the labels. The labels used differentiate white matter, gray matter, cerebrospinal fluid (CSF) and background.
The features being investigated are: the intensities of T1 and T2, Gaussian blurring (with sigma equal to 10) of T1 and T2, the absolute intensity of T1 minus T2, a median filter applied to both T1 and T2 and the distance from each point to the center.
The feature data sets are normalized for comparing purposes. Multiple features are used in the computation of the tissues segmentation.

2.2 Feature analysis

In order to compare and inspect different combinations of features, a scatter-matrix is created (see figure 4). The plots in this matrix contain the sampled slice 1-data from all subjects plotted in their respective feature space. The scatter matrix is now visually analysed. The 2D feature sets that were deemed to provide good enough cluster segmentation are now stored in an array for later use. These feature sets will be used in multiple combined K-nn algorithms later on (which is called "feature bagging"). They will not be used in one and the same segmentation algorithm, having in mind the "curse of dimensionality" [7].

2.3 Segmentation

The main segmentation is primarily done using k-NN classifiers, which segment the images in 4 classes. The main classifier will be a combination of k-NN classifiers that are trained on the different feature sets, as were described in section 2.2 ("feature bagging"). These sub-classifiers are trained based on a learning curve and on a combination of all training set images, comprised of 3 training-training images and 1 test-training image, in order to generate a generalisable model. Each sub-sub-classifier is trained on 3000 sample points, since down-sampling drastically improves computing times.

In the final step, a morphological binary opening is performed, which will "cancel out" some of the noisy background pixels, especially in the area around the skull.

The final segmentation results will be compared to both a combined atlas and a combined k-NN algorithm, to provide some context for its accuracy.

3   Results

3.1   Feature Analysis

As can be seen in figure 4, the 2D feature spaces providing the best separation of classes are: T1 with T2; T1 with T1 Gaussian; T1 with T2 Gaussian; T1 with T1 Median filtered; T1 with T2 median filtered; T1 with the distance to the center. Thus, these feature spaces are to be used in the feature analysis and segmentation. The other features show less or no separation of the classes. Furthermore, it is also visible that the absolute difference between T1 and T2 splits the classes in such a way they become scattered over the feature space.
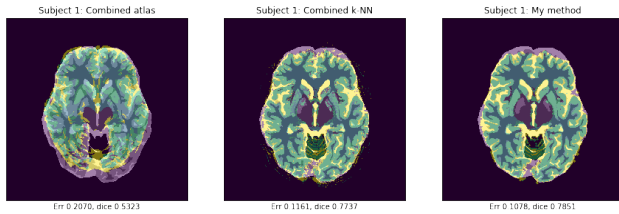
3.2   Segmentation

Figure 1: Tissue segmentation of the brain with combined atlases, combined k-NN and the method described in this paper for subject 1.

In figure 1, the outcomes of different segmentation methods are presented. These are all applied to subject 1. The other subjects yield equal results, as can be seen in appendix A. Visually, it is apparent that the combined atlas method yields the least accurate results, while combined k-NN and my_method yield similar results. It can however be spotted that the prediction by my_method seems less "noisy".

In figure 2 and 3, the classification errors and dice coefficients of all segmentations are presented respectively. From figure 2, it is apparent that the combining atlases algorithm yields the largest error. Furthermore, the method described in this report (my_method) has a slightly lower mean error than combining k-NN. The spread looks equal between the three methods.

In figure 3, the lowest dice score comes from the segmentation by combining atlases. The mean dice score created by my_method seems very slightly lower than that of combined k-NN.
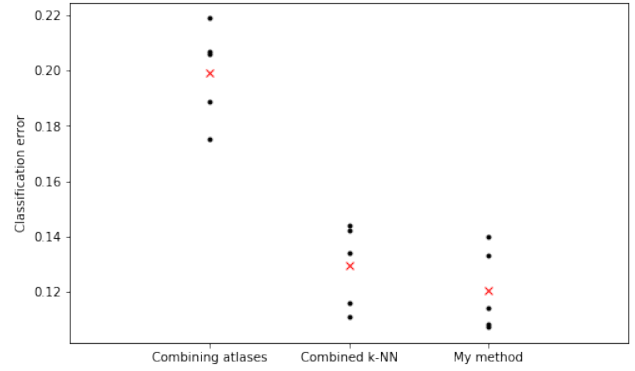

Figure 2: Strip chart containing the classification errors of the different subjects when segmented with combining atlases, combined k-NN and the method described in this paper
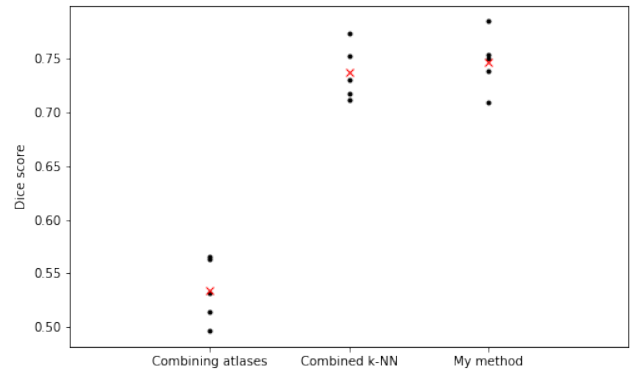

Figure 3: Strip chart containing the dice scores of the different subjects when segmented with combining atlases, combined k-NN and the method described in this paper.

4   Discussion

4.1   Feature Analysis

As seen in figure 4 T1 with T2, T1 with T1 Gaussian, T1 with T2 Gaussian, T1 with T1 median filtered, T1 with T2 median filtered and T1 with the distance to the center give the best class separation in feature space. Thus only these feature spaces were used in the segmentation. The other feature spaces suffered from severe overlapping of the classes. The only outlier in this is the feature representing the difference in T1 and T2. This feature splits up the different classes making it seem challenging to contain this feature in the segmentation. This hypothesis was confirmed during testing.

4.2   Segmentation

Of the three methods that were compared in this paper, my_method performed best on all images in this dataset. However, the segmentation greatly resembles the segmentation from the combined k-NN method. This makes sense, since my_method makes use of the same type of algorithms as the combined k-NN

method. What the created method excels at, is the reduction of noise in the subareas, especially at the boundaries. This can be partly explained by the use of morphological opening. Much like the combined k-NN method, however, it seems to struggle with seemingly "random", small and high-frequency variations in the scan intensity values, which mainly manifests itself in relatively high errors in sub-cortical areas. Combined atlas seems less viable to use than the other two methods, since while it describes the empty areas inside the brain better than the other methods, it fails on the outskirts and the subareas in the image. This can be explained by moderate variation in the train images that compute the atlas for segmentation of the desired image. It might also be more susceptible to small train sizes.

As can be seen in figures 3 and 2, my_method performs, on average, best on this specific data-set. It is, however, not possible to conclude a significant advantage of my_method relative to the other two methods, due to the relatively small sample size.

References

[1] S.K. Warfield. Real-Time Image Segmentation for Image-Guided Surgery, 1998.

[2] Adriënne M. Mendrik, Koen L. Vincken, Hugo J. Kuijf, Marcel Breeuwer, Willem H. Bouvy, Jeroen De Bresser, Amir Alansary, Marleen De Bruijne, Aaron Carass, Ayman El-Baz, Amod Jog, Ranveer Katyal, Ali R. Khan, Fedde Van Der Lijn, Qaiser Mahmood, Ryan Mukherjee, Annegreet Van Opbroek, Sahil Paneri, Sérgio Pereira, Mikael Persson, Martin Rajchl, Duygu Sarikaya, Örjan Smedby, Carlos A. Silva, Henri A. Vrooman, Saurabh Vyas, Chunliang Wang, Liang Zhao, Geert Jan Biessels, and Max A. Viergever. MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. *Computational Intelligence and Neuroscience*, 2015, 2015.

[3] David Arthur and Sergei Vassilvitskii. k-means++: The Advantages of Careful Seeding. Technical report.

[4] Petronella Anbeek, Koen L. Vincken, Glenda S. Van Bochove, Matthias J.P. Van Osch, and Jeroen Van Der Grond. Probabilistic segmentation of brain tissue in MR imaging. *NeuroImage*, 27(4):795–804, 10 2005.

[5] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, 15(1), 8 2015.

[6] Robert Nilsson. Classification Error Is environmental tobacco smoke a risk factor for lung cancer? Technical report.

[7] Eamonn Keogh and Abdullah Mueen. Curse of Dimensionality. In *Encyclopedia of Machine Learning and Data Mining*, pages 314–315. Springer US, Boston, MA, 2017.

Appendices

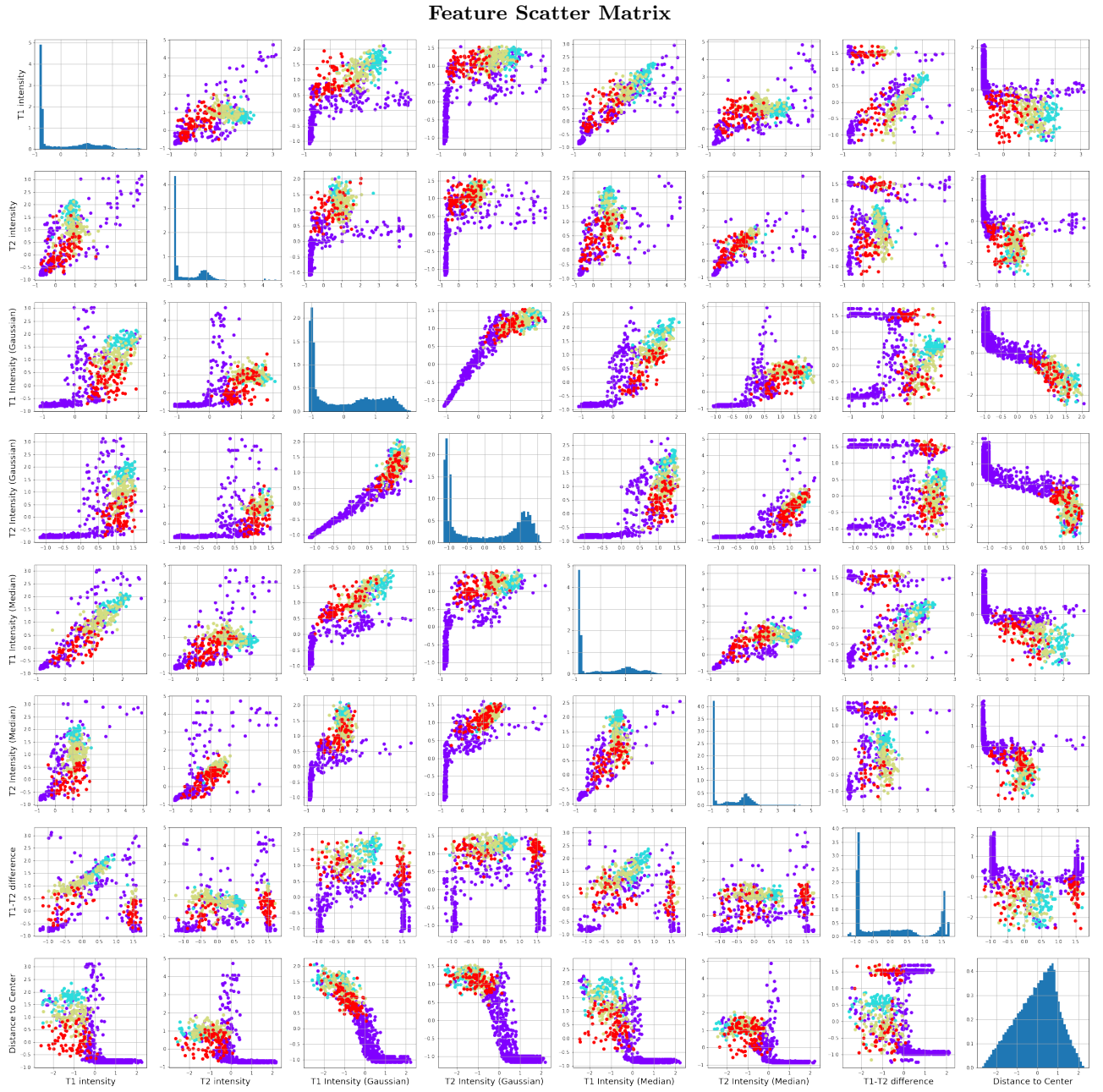## A Figures

Feature Scatter Matrix

Figure 4: Scatter matrix for all features, plotted over slice 1 of all subjects
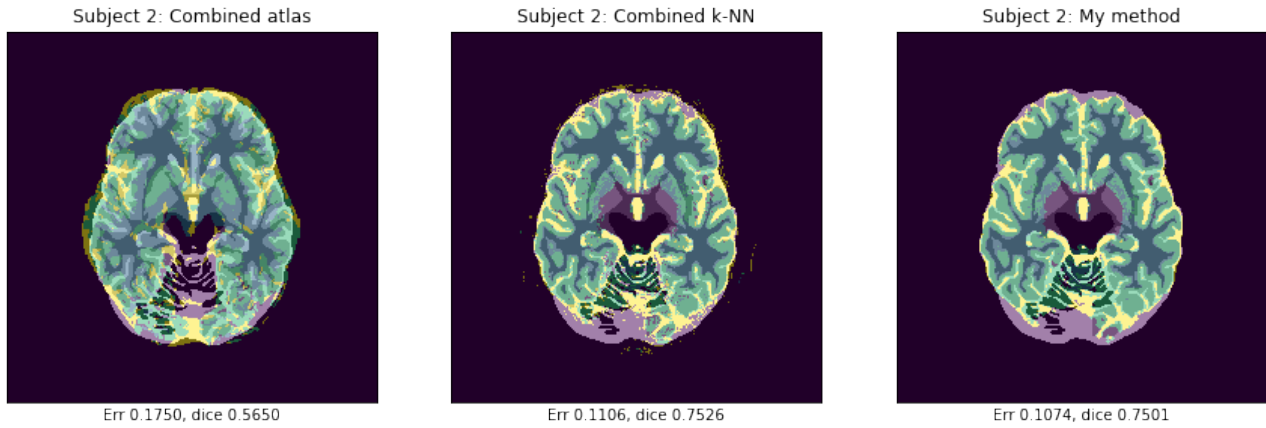
Segmentation Results - Subject 2

Figure 5: Tissue segmentation of the brain with combined atlasses, combined k-NN and the method described in this paper for subject 2.

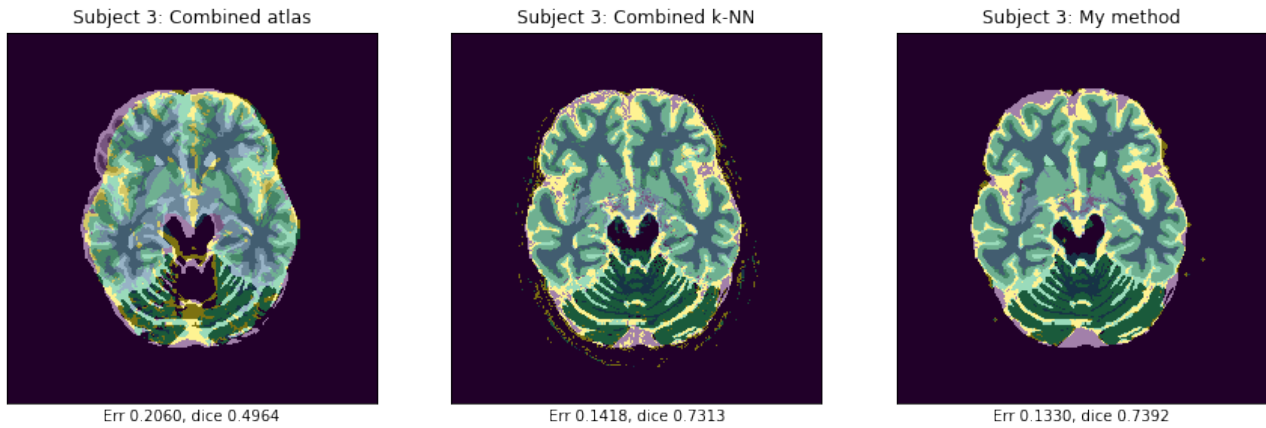**Segmentation Results - Subject 3**

Figure 6: Tissue segmentation of the brain with combined atlasses, combined k-NN and the method described in this paper for subject 3.
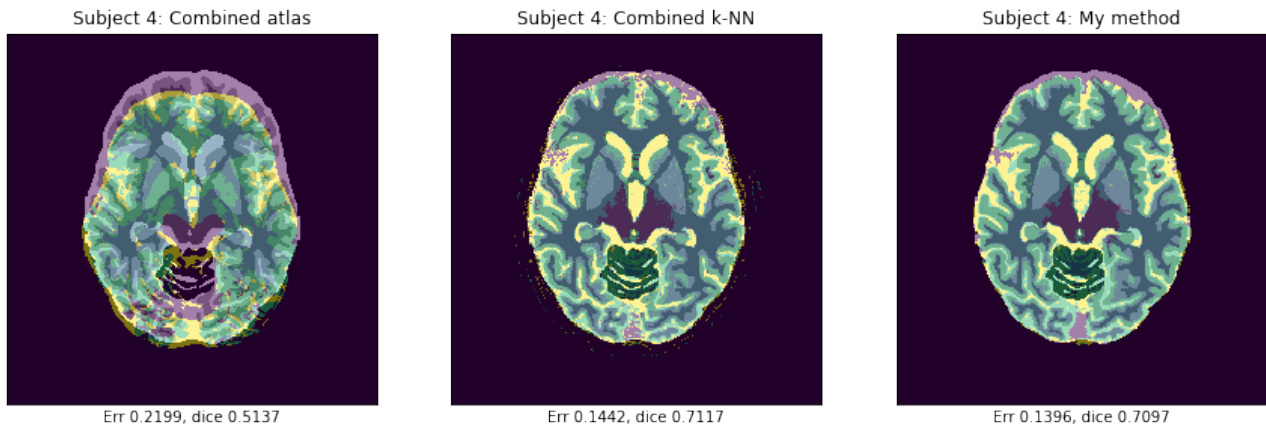
Segmentation Results - Subject 4

Figure 7: Tissue segmentation of the brain with combined atlasses, combined k-NN and the method described in this paper for subject 4.

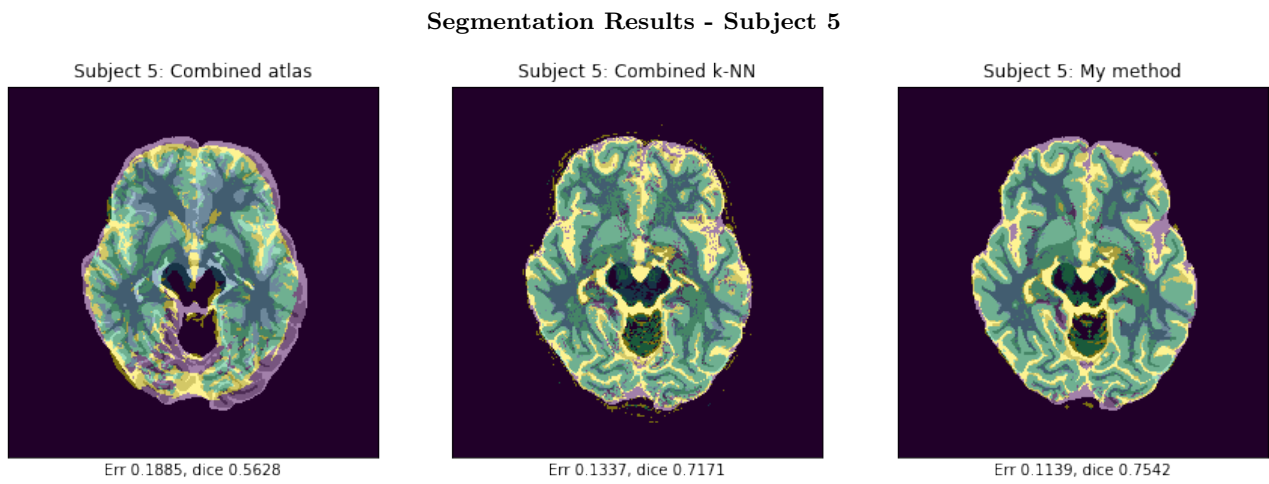**Segmentation Results - Subject 5**

Figure 8: Tissue segmentation of the brain with combined atlasses, combined k-NN and the method described in this paper for subject 5.