FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Comparison of Vessel Segmentation Techniques

BACHELORARBEIT

zur Erlangung des akademischen Grades

Bachelor of Science

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Alexander Gall
Matrikelnummer 1225540

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ-Prof. Dipl.-Ing. Dr.techn. Meister Eduard Gröller
Mitwirkung: Dipl.-Ing. Haichao Miao

Wien, 8. März 2017

Alexander Gall | Meister Eduard Gröller

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Comparison of Vessel Segmentation Techniques

Comparative Visualization and Evaluation of Image Segmentations

BACHELOR'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science

in

Medical Informatics

by

Alexander Gall
Registration Number 1225540

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ-Prof. Dipl.-Ing. Dr.techn. Meister Eduard Gröller
Assistance: Dipl.-Ing. Haichao Miao

Vienna, 8th March, 2017

Alexander Gall | Meister Eduard Gröller

Erklärung zur Verfassung der Arbeit

Alexander Gall
Engerthstrasse 259 10/6/19, 1020 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 8. März 2017

Alexander Gall

Abstract

Image segmentation is an important processing step in various applications and crucial in the medical field. When a new segmentation technique is introduced, validation and evaluation are essential for medical image analysis. But the automation of these processes is still not sufficient. Many algorithms have been published but there is still no satisfying way to assess whether an algorithm produces more accurate segmentations than another. More effort is spent on the development of algorithms than on their evaluation and therefore many researchers use the less complex subjective methods. For these techniques multiple experts are needed to visually compare several segmentation results, which is a very time-consuming process. Another way of comparing different results is the supervised evaluation method. Here we need experts, who manually segment reference images, which are used for comparison. As seen in recent researches there is a need for unsupervised methods due to many applications, in which user assistance is infeasible. The aim of this thesis is to provide an environment to visually and objectively evaluate segmentation results in the field of vessel segmentations. Our framework enables the comparison at voxel-level with various visualization techniques and objective measurements. These methods are meant to make the comparison more understandable for users. A subjective evaluation is realized through a comparative visualization by using a two- and three-dimensional comparison of voxels. Another general overview is provided by a maximum-intensity projection, which highlights the vessel structure. As purely objective evaluation technique, various metrics are used, to assure independence from experts or a ground truth. By using these techniques this paper presents an approach for evaluating differences in medical images, which does not rely on a permanent presence of an expert.

Kurzfassung

Die Segmentierung von Bildern ist ein wichtiger Bearbeitungsschritt in vielen Programmen und ein wesentlicher Bestandteil im medizinischen Arbeitsbereich. Eine Validierung und Evaluierung von neu entwickelten Segmentationstechniken ist unerlässlich im Gebiet der medizinischen Bildanalyse. Die Automatisierung dieser Prozesse ist aber immer noch nicht zufriedenstellend. Obwohl viele Algorithmen und Evaluierungsarten publiziert wurden, gibt es bis heute immer noch keine befriedigende Möglichkeit zu überprüfen, ob ein Algorithmus besser ist als ein anderer. Die meisten Forscher richten ihre Anstrengungen mehr auf die Entwicklung von Algorithmen, als auf deren Evaluierung und nützen daher die weniger komplexe, subjektive Evaluierung. Für diese Technik werden aber viele Experten benötigt, deren zeitaufwendige Aufgabe es ist, unzählige Segmentierungen optisch zu vergleichen. Eine andere Möglichkeit zum Vergleichen von Segmentierungsresultaten ist die überwachte Evaluierung. Hier werden Experten benötigt, welche ein Referenzbild manuell segmentieren, das zum Vergleich verwendet werden kann. In den jüngsten Forschungen zeigt sich aber, dass unkontrollierte Evaluierungen benötigt werden, da bei vielen Anwendungen keine Unterstützung durch Experten gegeben ist. Das Ziel der vorliegenden Bachelorarbeit ist es eine Anwendung zu schaffen, in der es möglich ist Resultate von Gefäßsegmentierungen sowohl visuell, als auch objektiv zu vergleichen. Das Framework ermöglicht eine Evaluierung auf Voxel-Ebene mit verschiedenen Visualisierungstechniken und objektiven Messwerten. Diese Methoden sollen die Ergebnisse des Vergleiches für den Benutzer verständlicher machen. Die subjektive Evaluierung wird über zwei-dimensionale und drei-dimensionale, komparative Darstellungen von Voxeln erreicht. Durch eine Maximum-Intensitäts-Projektion, bei welcher die Gefäßstruktur hervorgehoben wird, entsteht ein grundsätzlicher Überblick der vaskulären Umgebung. Für eine rein objektive Evaluierung wurden verschiedene Metriken verwendet, welche eine Unabhängigkeit von Experten oder einer "Ground Truth" ermöglichen. Durch die Anwendung der genannten Techniken, präsentiert diese Arbeit ein Konzept für die Evaluierung von Unterschieden in medizinischen Bildern, ohne auf das permanente Vorhandensein von Experten angewiesen zu sein.

Contents

CHAPTER 1

Introduction

1.1 Motivation

Image Segmentation and its visualization are an important field in today's medicine. Most diagnostic procedures need medical imaging support for localizing damages or to search for abnormalities. In the complex area of the human brain even more caution is needed. Brain damage is caused by various circumstances but many are due to occlusion of vessels, like a stroke. To find these disorders fast and effectively, radiologist need applications, which give a good overview over the vascular system. Therefore, many segmentation techniques were developed over the years. Image segmentation allows, for instance, the localization and highlighting of occluded vessels. Because of the great amount of segmentation techniques, it is hard to choose which one fits the needed situation best. Although there are so many methods, there is no satisfactory performance measure, which would help to compare different algorithms. As Price [1] wrote about 30 years ago "Computer vision suffers from an overload of written information but a dearth of good evaluations and comparisons." Even now not much has changed and it is hard to compare different segmentation results or parametrizations of a method. The most common way to compare segmentation results is to let an expert (i.e. radiologist) visually compare the different outcomes, as described by Zhang et al. [2]. For this tedious task not only a skilled person is needed, but it is also a time-consuming job and time is rare in medical fields. Another drawback of this approach is that the results still depend on human intuition, as Zhang noted [3]. To reduce the work and improve time-management, we can use the supervised evaluation methods, in which the expert is only needed to manually segment a reference image. The segmentation, which is done by hand, is often called ground truth. The problem is that we mostly have a large variety of images with unknown content and unacquainted parametrizations, from which we want to create a comparison. Because of this circumstance unsupervised evaluation methods are necessary. As state by Zhang et al [2] the main advantage of unsupervised segmentation methods is, that the comparison of segmentation results against a manually-segmented reference image ceases. Not only enables this benefit a objective comparison, furthermore it is possible for segmentation algorithms to perform self-tuning of their parameters. This work proposes methods and capabilities to combine a subjective comparison of segmentation results as well as an objective evaluation. This objective comparison is essential in cases where a fast and precise medical imaging is needed. With this segmentation and objective evaluation, applications can support diagnostic processes in real-time and enable important decision aid and visual assistance for the medical staff.

1.2 Medical Background

The brain is the most important organ of the human body. Its complex structure with billions of nerves and strong blood supply makes diagnoses as well as operations difficult. Cipolla [4] defined it like this: "As an organ, the brain comprises only about 2% of body weight yet it receives 15–20% of total cardiac output, making the brain one of the most highly perfused organs in the body." This high perfusion is mandatory to ensure the proper working of the brain. Even a little reduction in the blood supply leads to permanent damage in the affected region. If the brain has an undersupply of blood and oxygen for too long, a stroke can emerge. But also vascular changes like thickness and stiffness effects the brain. Grinberg et al. [5] stated that the most predominant disorders of the vascular structure in the brain of elderly are Cerebral atherosclerosis, Small vessel disease and others. These injuries can be detected by medical imaging in combination with segmentation algorithms.

1.2.1 Vascular Structure of the Human Brain

The primary blood supply is provided by two paired arteries, the Aa. carotides internae (supplies the anterior part) and the Aa. vertebrales (supplies the posterior part), which is described by Anderhuber et al. [6]. The anterior and the posterior cerebral circulation interconnect and create the circulus arteriosus cerebri or also called Circle of Willis. As shown in Figure 1.1 the Aa. vertebrales combine and form the A. basilaris (AB), which splits into the branches of the A. cerebri posterior (ACP). These arteries branch into the A. communicans posterior (ACoP), which then end into the A. cerebri anterior(ACA). This artery can be seen well in the schematic representation in 1.1. Finally the circle (anastomosis) closes with the branches of the A. cerebri anterior merging over the A. communicans anterior (ACoA). In figure 1.2 a three-dimensional representation of this vascular structure is shown. This image is generated in our framework and modified to give a good overview over the important arteries, which are discussed in this thesis.

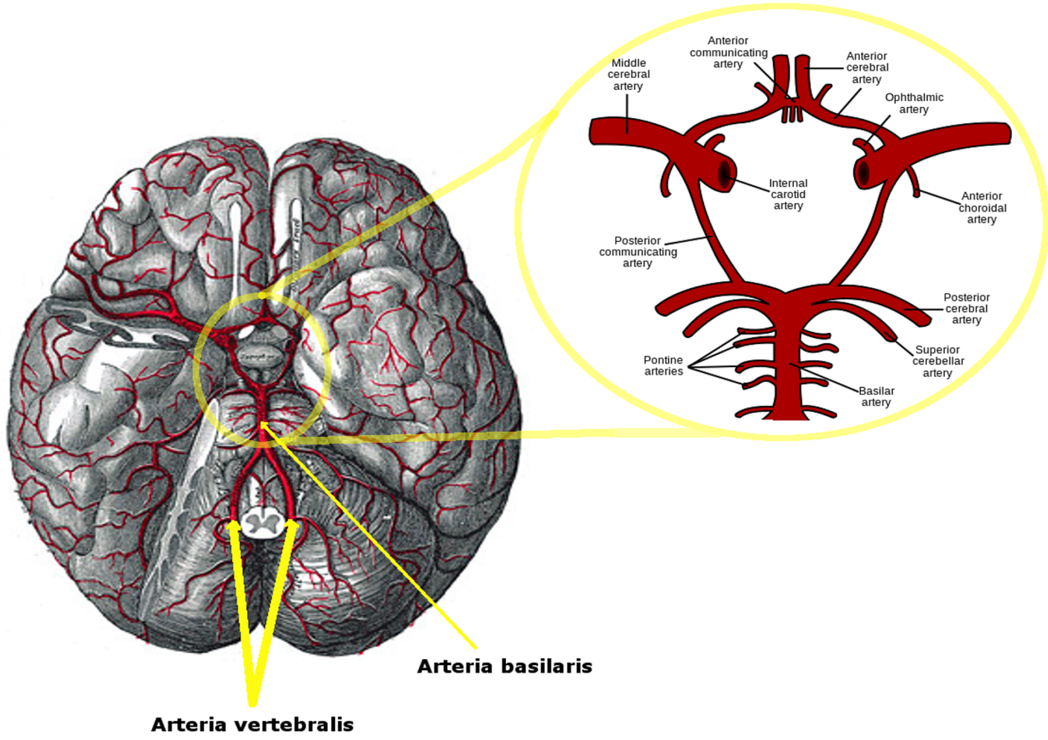


Figure 1.1: View of the vascular system of the head. The illustration shows the most prominent vascular structures of the human brain. The Circle of Willis is zoomed out and represented as schematic representation. The Figure of the arteries of the head is modified and taken from [7]. The schematic representation is modified and taken from [8].

1.2.2 Acquisition of Medical Image Data

There are many techniques and reasons for the acquisition of medical image data. Typically medical images are used for diagnosis, therapy planing, intraoperative navigation, research and other tasks, as Preim et al. [9] described. To fulfil these various tasks the data has to be of high-quality, in example having a high resolution, and should be showing the sought anatomy completely. Therefore it is important to choose an appropriate imaging modality to achieve a useable result. In the region of the brain it is crucial to have an accurate representation of the vascular system to enable medical actions. Some techniques use invasive methods, which yield better image quality, but have less patients tolerance and others use non-invasive methods. Some modalities to acquire medical data of the human brain, which can be used, to detect traumas and abnormalities, are Computer Tomography, Magnetic Resonance Imaging and techniques like Angiography.

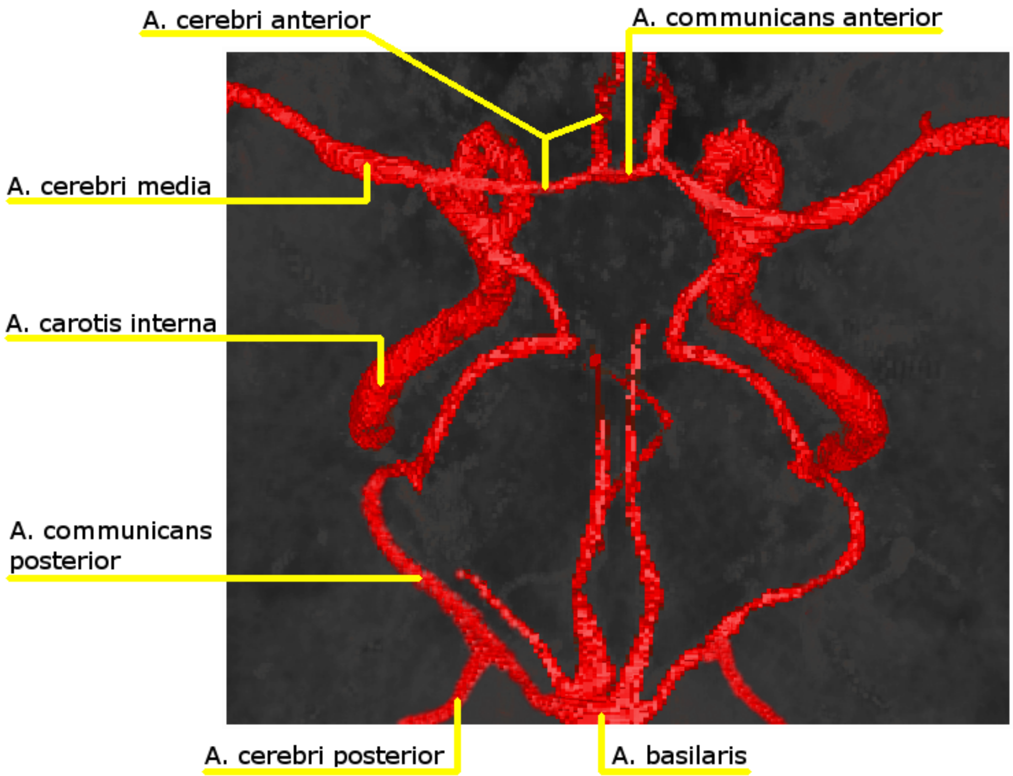


Figure 1.2: This image shows the Circle of Willis in a three-dimensional representation. The depiction is taken from our framework and modified to delineate and enhance the different arteries.

Computer Tomography

X-ray Computer Tomography (CT) expands the older 2D X-ray imaging with an volumetric representation of the objects. The more complex CT data, is composed out of a series of individual X-ray images, which create one volume when combined. The in-plane resolution is lower in comparison with normal X-ray imaging and some artefacts have to be considered. In return, CT is widely available and has a fast image acquisition with the ability to localize pathologies. In combination with cerebral angiography it is applicable to display blood vessels inside the brain.

Magnetic Resonance Imaging

In Magnetic Resonance Imaging (MRI) a patient is placed in a strong magnetic field. Due to the different properties of human tissue the Hydrogen nuclei act differently under magnetic fields. Therefore MRI is appropriate to distinguish soft tissues and enables a high contrast where CT is less effective. Since MRI does not need X-rays, it also can not harm the patients and the complex data can be used for various diagnostic tasks. The disadvantages of MRIs are their rarity due to the high acquisition costs and the need of good segmentation algorithms as well as skilled radiologists.

Angiography

To even more improve the representation of blood-filled structures Angiography is used. This technique can be used in addition to previous introduced imaging modalities. Angiography is a invasive method, in which a contrast agent is injected into the blood vessel. This yields more visible vascular structures and helps improving the segmentation process.

1.3 Aim of the Work

The purpose of this thesis is to provide an environment to visually and objectively evaluate different segmentation results of vascular structures of the brain. The insights of this work should help radiologist in their work, as well as other researcher to compare their segmentation results more efficiently. This however should make it possible to improve the diagnostic process on the one hand by more precise segmentation algorithms through evaluation and on the other hand by supporting real-time diagnosis for the medical staff. The goal is to resign manual interaction and enable automatic segmentation and evaluation, which in the future can yield to self-tuning and parametrization in the field of medical image analysis.

CHAPTER 2

Related Work

The field of medical image analysis and visualization has expanded significantly over the years. The need to extract clinically relevant information from medical data to generate high-quality visualizations for radiologists increased with better imaging modalities. These various segmentation techniques, which have been developed and presented in the literature, are generally not suitable for all images or applications. To compare these algorithms and methods researcher have began to draw their attention on evaluation techniques. The goal is to develop segmentation algorithms, which extract the object of interest as good as possible, and validate them against other implementations. With this comparison a sufficient visualization can be produced. To make this process efficient and easy to use for doctors, it has to be automated, which is a complex task and still not solved. Preim et al. [9] described the problem like this: "Medical image data, anatomical relations, pathological processes, image modalities, and biological variability exhibit such a large variety that automatic solutions for the detection and delineation of certain structures cannot cope with all such cases."

2.1 Segmentation Methods

Every medical image analysis follows a simple pipeline, which starts with preprocessing and filtering of the data. After that the important part is the extraction of the anatomical structure for the required task. The goal of segmentation is the recognition of relevant objects, as well as their accurate delineation. Unfortunately there is no single segmentation algorithm, which can extract all useful objects from every medical image. Therefore some requirements as noted by Preim et al. [9] as well as segmentation techniques described by Kirbas and Quek [10] are required to characterize the algorithms.

2.1.1 Requirements

- Robustness
  This means the algorithm should work for a large variety of cases. Only robust methods can be used effectively for the comparison of different images and real-time diagnosis.

- Accuracy
  Segmentation results should be validated for their accuracy. That means the results should be compared against their ground truth, or respectively against the correct reference image.

- Reproducibility
  Reproducibility or Precision, as described by Chang et al. [11], means that independently of the user, the algorithm performs very similar result every time its applied to the same image.

- Speed
  This requirement determines that the execution time of an segmentation should be fast enough to enable an interactive use of the application. Sometimes this characteristic is also called Efficiency [11], which combines the processing time with the computational complexity.

2.1.2 Segmentation Techniques

Because there is no single segmentation method for all objects of interests and image modalities, Kirbas and Quek [10] categorized the techniques into six groups, which can be seen in figure 2.1.

- Pattern recognition techniques
  Pattern recognition imitates human vision and tries to extract objects through intensities, homogeneous regions and similar cases. A fundamental pattern recognition technique would be thresholding in all its variations.

- Model based approaches
  The model based approach assumes that every object of interest has an certain geometry, which can be described as a model. Therefore a prior knowledge about the structure (shape, size, etc.) of the object has to exist. In a two-step process an initial contour is determined and then adopted to local features [9].

- Tracking-based approaches
  This technique follows a track, which should represent a vessel. On the track, which is mostly chosen by a user, local operators are applied. So the optimal path of the vessel can be tracked with help of centerlines and boundaries.

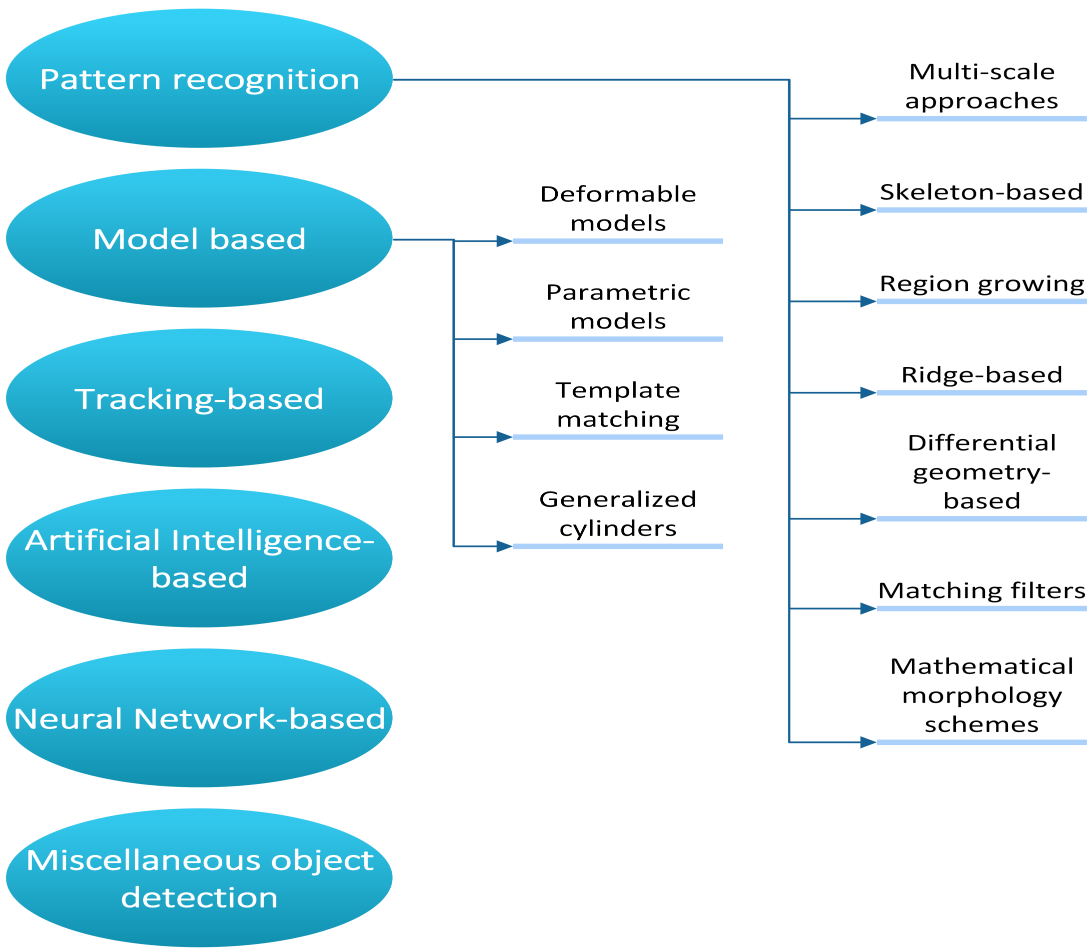


Figure 2.1: Categorisation of segmentation techniques according to [10]. The illustration shows the basic categorisation of segmentation methods, which can be applied to extract sought structures.

- Artificial Intelligence-based approaches
  Artificial Intelligence-based approaches use prior knowledge for the segmentation and the description of the vascular structure. This means with information like shape, intensity patterns or even a whole model, rules for the characterisation of the object can be described and used for guidance of low-level image processing algorithms.

- Neural Network-based approaches
  These systems are made for self-learning with a great amount of training sets. The network consists of several nodes, which perform elementary algorithms through their input and parameters and allow together complex computations. To do so, the network has to be trained every time a new feature is added, but allow segmentation

of objects, from which no prior knowledge is existing.

- Miscellaneous tube-like object detection approaches
  Although not developed for vessel extraction, this method can be adapted, due to its tube-like structure. This approach combines a lot of the other techniques like the generalized cylinder approach or the matching filters approach to find any form of tube-shaped objects like roads or pipelines.

2.2 Evaluation Methods

It is crucial in medical image analysis to be able to measure the quality of a segmentation results. This is not only important to measure the improvements of a single algorithm but also the possibility to compare multiple algorithms. Recent research aims to solve the problem of evaluating different segmentation techniques and their parametrization and to characterize the different evaluation methods. As can be seen in Figure 2.2 Zhang et al. [2] grouped the processes of evaluating a segmentation method in a hierarchy.

2.2.1 Segmentation evaluation hierarchy

The task of evaluation can be divided into two major categories, the subjective evaluation and the objective evaluation. This distinction determines if an expert (i.e. radiologist) visually evaluates the results or if the evaluation is based on objective parameters.

Subjective Evaluation

As mentioned above subjective evaluation needs a person, who visually compares two segmentation results. Although it is not always possible to have one or more experts and the results are not consistent due to its inherently subjective nature, subjective methods are still the most widely type of evaluation. As Zhang stated [2], this methods are time-consuming, intrinsic and therefore can not be successfully used in real-time system to help picking an adequate segmentation algorithm or even change its parametrization. Thus the aim for state-of-the-art applications in the field of medical image segmentation is to not rely on manual intervention and abolish its need, as Crum et al. mentioned [12]. But still there are some "goodness" parameters, which an objective method cannot reflect properly compared to humans, like Wang explained [13]. This means that some desirable properties of the segmentation are often created because of human intuition. A good explanation for this is provided by Zhang et al. [2], where the authors say "[...] manually-segmented images are based on humans' semantic understanding of the real-world objects in the image".

Objective Evaluation

To overcome the need for visual comparison from experts and differentiate between methods and their various parametrizations, an objective evaluation is needed. Therefore

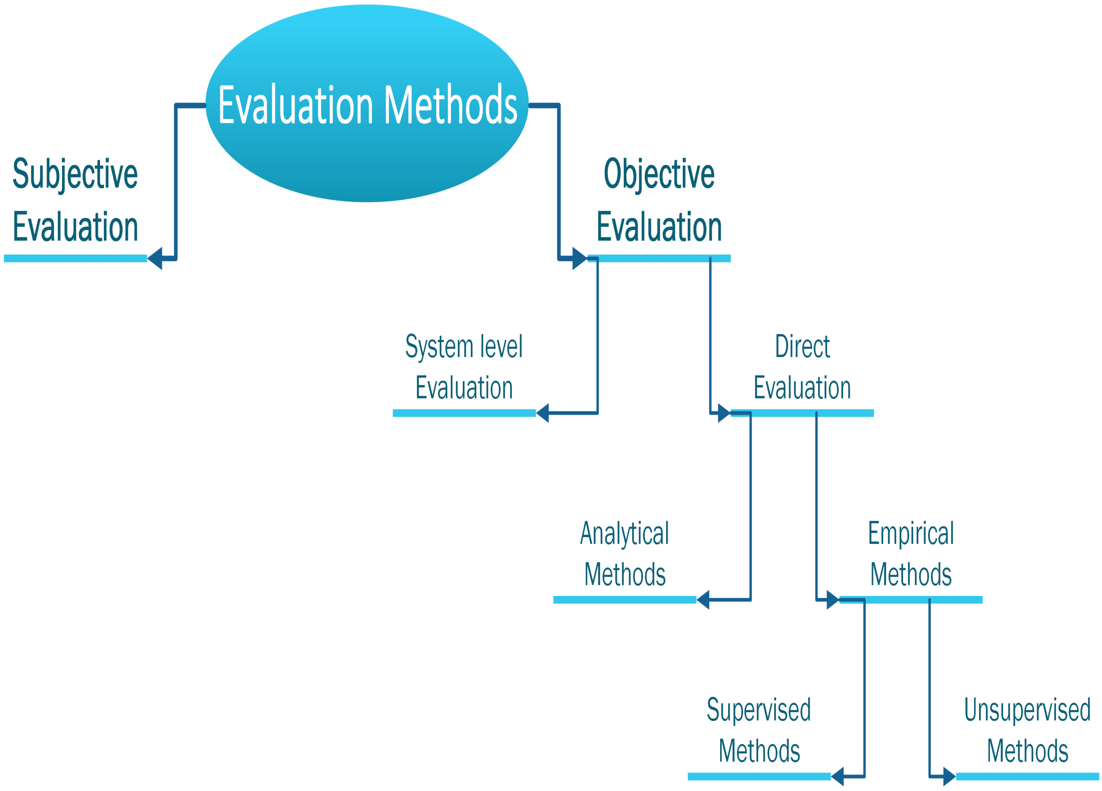


Figure 2.2: Hierarchy of segmentation evaluation methods. The basic division of evaluation tasks can be made into subjective methods and objective methods. Objective methods can further be divided, where the subjective evaluations relays purely on experts. Figure is modified and taken from [2].

we must distinguish System level evaluation, which uses the overall performance of an end system, from direct evaluation techniques. The direct approaches can further be evaluated analytically or empirically, based on whether the method itself, or the result is evaluated. The focus however lies on the empirical methods, which can be differentiated in supervised methods and unsupervised methods. The biggest impact on recent research have the unsupervised methods, which enable objective evaluations without any experts. The problem with the supervised methods lies in the need of a ground truth or gold standard, which can only be obtained by manually segmented reference images from experts. The next section will describe the individual methods in more detail.

- **System level Evaluation**
  This evaluation type examines the impact of different segmentation methods on the overall system [2]. Not the single segmentation technique is evaluated but rather all images will be segmented with a set of techniques and the best overall performance for the result is taken. This indirect evaluation does not select the best

technique for a single image and therefore only indicates that the current result is more favourable for that system.

- **Direct Evaluation**
Direct methods have the advantage that various segmentation techniques can be applied on the different images and the best technique can be individually selected. As consequence an overall better evaluation can be obtained than for a system level evaluation.

- **Analytical Methods**
The analytical methods asses the segmentation method directly with its properties and principles. As Polak et al. describes [14] the methods emphasise on analyzing the characteristics of the algorithms, like processing strategy, complexity, and efficiency. So it is possible to evaluate the algorithms after their processing strategy, where parallel algorithms are more suitable for faster implementation [3]. This however limits the use of this methods to only evaluate algorithmic or implementation properties, for which a certain prior knowledge is needed.

- **Empirical Methods**
These important methods try to improve the evaluation process by judging the segmentations after their results. The way in which the original image and the segmented result are used, can be differentiated into empirical goodness methods or empirical discrepancy methods. With the empirical goodness methods the segmentation result is judged by "goodness" measures [3]. These measures are derived from human intuition and can be best described with statistical measures like the uniformity within segmented regions, inter-region contrast, and region-shape [14]. Empirical discrepancy methods, in contrast emphasize more on the difference between the segmentation result and the ideal segmentation. Therefore some metrics are needed, which calculate different errors between a reference image and the actual result. For this object-by-object comparison a reference image, often called ground truth or gold standard, is needed, which is mostly done by an expert.

- **Supervised Methods**
The supervised methods are derived from the empirical discrepancy methods, which means they compare the segmentation result against a manual segmented image. The basic time-consuming procedure is to let an expert segment an image, which serves as gold standard, but is still not accurate or reliable. Then this image is compared with an automatically segmented image. The higher the similarity of the machine-segmented image compared to the manual result of the human becomes, the better is the quality of the segmented image [2]. This should yield a finer resolution, but because of the reference image the comparison is still not fully objective. Another disadvantage is described by Kohlberger et al. [15] where the authors stated that supervised methods are applicable for the development of algorithms, but afterwards there is no automated method for the evaluation of

segmentation techniques available. This is because there is no way to get a ground truth for this task after the deployment.

- **Unsupervised Methods**
  This methods are derived from the empirical goodness methods and deal with the problem of evaluation without a ground truth. These methods "[...] do not require a reference image, but instead evaluate a segmented image based on how well it matches a broad set of characteristics of segmented images as desired by humans [2]." This characteristic of unsupervised methods makes it possible to compare segmentation results where no gold standard is available. Furthermore this allows self-tuning algorithms based on the evaluation results. Since no user assistance is needed, unsupervised methods qualify also for real-time diagnosis when sufficient segmentation criteria or metrics are met.

2.3 Evaluation Measures

There are many metrics and measures possible to create an objective evaluation measure. The most metrics are based on the intra-region uniformity, inter-region disparity, barycenter distance, shape, volume differences and combination of the previously mentioned metrics. In Table 2.1 some of the measurements, used for comparison of segmentation results, are specified.

Metrics	
Jaccard	$JC(A,B) = \frac{\lvert A \cap B\rvert}{\lvert A \cup B\rvert}$
Dice	$D(A,B) = \frac{2\lvert A \cap B\rvert}{\lvert A\rvert + \lvert B\rvert}$
Volumetric Overlap Error	$VO(A,B) = 100 * (1 - \frac{\lvert A \cap B\rvert}{\lvert A \cup B\rvert})$
Relative Volume Difference	$RV(A,B) = 100 * (\frac{\lvert A\rvert - \lvert B\rvert}{\lvert B\rvert})$
Volume Similarity	$VS(A,B) = 100 * (\frac{\lvert\lvert A\rvert - \lvert B\rvert\rvert}{\lvert A\rvert + \lvert B\rvert})$
Martin error	$P_{ji}(A,B) = (1 - \frac{\lvert A_j \cap B_i\rvert}{\lvert A_j\rvert} * \lvert A_j \cap B_i\rvert)$

Table 2.1: Possible Metrics for evaluation, where A is a a set of voxels from the first Image and B stands for a set of the second image.

- **Jaccard coefficient**
  The jaccard coefficient is a region-based spatial overlap measure. It measures the ratio of the intersection area of two sets (A,B) divided by the area of their union [11]. The coefficient gets zero when A and B are disjoint and one when the sets are equal. This coefficient is often used synonymously with the Tanimoto coefficient.

- **Dice coefficient**
  The Dice coefficient computes the ratio of the intersection area divided by the mean sum of each individual area. This coefficient is, like the jaccard coeffcient, a spatial overlap measure, which is derived from a reliability measure [11].

- **Volumetric Overlap Error**
  This error measure is derived from the jaccard coefficient and given in percentage. As stated by Heimann et al. [16] it is the ratio between intersection and union and returns zero for an exact segmentation and 100 if the segmentation and the reference do not overlap at all.

- **Relative Volume Difference**
  This measure is used to detect over- or undersegmentations, but due to its asymmetry it is not considered a metric. If both volumes are identical the result of the measure is zero, which does not imply that A and B are identical or overlap with each other [16]. The results are given in percentage.

- **Volume Similarity**
  This measure checks the similarity of two sets, irrespective of the position of the points of the sets. If the number of elements in both sets are equal, the result of the measure is one percent and zero percent, when one of the sets is empty, as noted by Fernandez in [17].

- **Martin error**
  The martin error measure produces a real-value output in the range of [0,1], where one is an error and zero means a correct segmentation. It is a better measure for supervised evaluation, since Polak et al. [14] stated "The measure is shown to be effective for qualitative similarity comparison between segmentations by humans, who often produce results with varying degrees of perceived details [...]". Here A_j is the jth fragment in the reference image and B_i is the ith fragment in the segmentation result.

2.4 Comparative Visualization

Not only the segmentation techniques and the evaluation methods are important to compare segmentation results. To understand the results of these processes, a good information visualization is needed. It is a common task to compare various images against each other and describe their differences and similarities. In the medical field a precise delineation of the anatomical structure is required to inspect abnormalities. Schmidt et al. [18] proposed some features, which should be implemented to allow a useful comparative visualization:

- **Scalability**
  The system should allow to compare the differences of large sets of images.

- **Focus+Context**
  A general overview over the matches and mismatches should be provided. Further the system is supposed to enable interactions to change or adjust the context.
- **Flexibility**
  The framework may aim for different types of images and apply various metrics.

Apart from these features the system should also enable a quick identification of the differences and provide information about the underlying data. Furthermore the encoding of the differences with abstract parameters, like color, improve the insights of the results. To enhance the analysis of this comparisons, an appropriate visualization is needed. A taxonomy for such comparative designs was described by Gleicher et al. [19]. The authors divided the possibilities, to present data structures, in three main categories:

- **Juxtaposition (separation)**
  Here the objects are presented in side-by-side views. This separate arrangement relies on the viewer's memory, but enables to see patterns better. It is possible to spot differences easier but the layout needs more space.
- **Superposition (overlay)**
  This method displays multiple objects in the same space. The blending or overlay of semi-transparent images makes the depiction easier to understand and reduces the required screen space. However, this type needs some sort of symmetrical data and is difficult to display when the data is too dense.
- **Explicit Encodings (aggregation)**
  With this technique relationships between objects are visually encoded. To highlight this connections the link between them has to be known. One difficulty of this methods is, to find the most significant relationship to compute. When a connection is displayed, other links are not clearly visible and lead to a loss of information.

To ensure the most possible information context for the viewer, different visualizations are needed. Two-dimensional, as well as three-dimensional depictions help to maintain the balance between precise details and a general overview. These different methods of displaying information and encoding the data, help to illustrate the differences and similarities between the segmentation results and enable an effective evaluation.

CHAPTER 3

Methodology

The current state of the art does not provide a optimal algorithm for the segmentation of vascular structure of the brain, therefore we propose techniques for the comparison and evaluation of these segmentation results. This thesis demonstrate methods and visualizations, which enable a subjective as well as an objective evaluation of segmentation results. The first step is to load the medical data. The images used in this thesis are generated with an MRA (Magnetic resonance angiography), which yields high-contrast images of the vascular structure of the brain. After the loading of the images the next step is to apply several segmentation algorithms with different parametrizations. It is also possible to load existing segmentations results, to make a comparison. After processing the images a proper visualization of the segmentation is needed. Therefore the results are displayed as a 2D slice view and 3D volume with an MIP (Maximum intensity projection) from the original image. If more than one input image is given statistics are shown, in which an objective evaluation based on metrics, is displayed.

3.1 Introduction

Our framework consists of three main stages, which can be seen in figure 3.1. The first stage focuses on the generation of a segmentation result or the converting of pre-segmented images for subsequent tasks. The second stage deals with the visualization of these results and the third stage takes care of the objective comparison of these images. In our approach we access data values at every voxel (e.g. intensity values) and distinguish between the algorithm, which segmented the voxel, and display the amount of algorithms, which successfully segmented it. Another important feature we implemented is the highlighting of the segmented voxel in different colors, which clearly identifies the algorithm, which defined a concrete position as vessel, or the segmentation distribution of the assumed vessel. With the combination of 2D views and direct volume rendering, which is used by our function for highlighting the segmentation, we greatly improve the vessel visualization for the comparison. To define the relationship between these results we defined some metrics, which help evaluating the segmentations in an objective way.

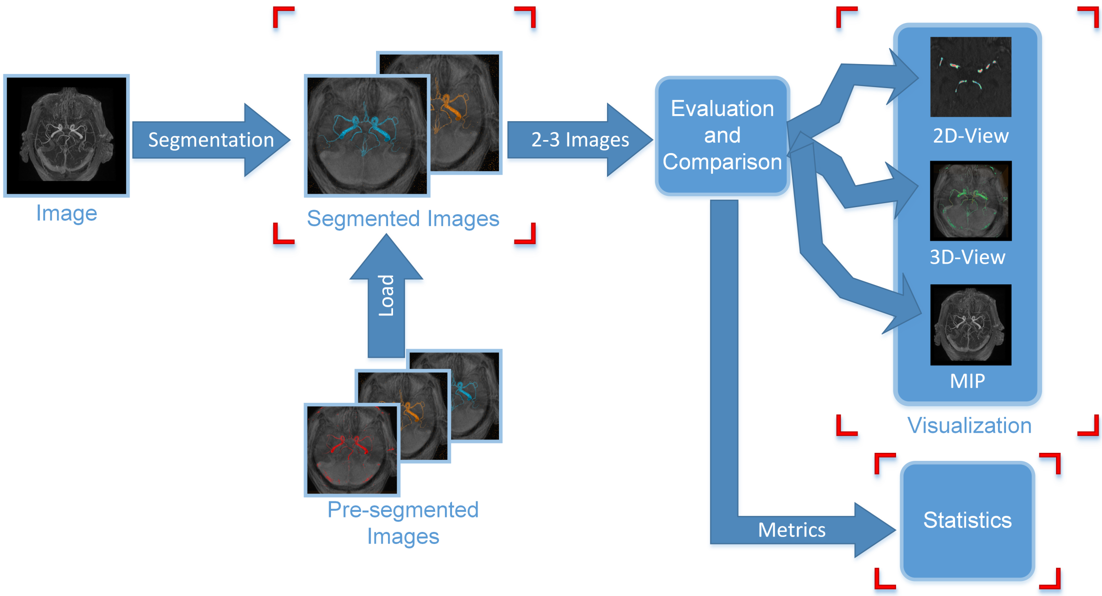


Figure 3.1: Overview of the concept behind our framework. The input for the evaluation and comparison stage consists of segmentation results from pre-segmented images or results from segmentations of the framework. The output provides the visualization in three different versions and the statistics computed through metrics. The three main stages of the framework are marked red.

3.2 Segmentation

To enable a segmentation of unprocessed images a set of segmentation algorithms is implemented. These methods should help to extract the vascular structure of the given medical data. The set contains a threshold algorithm and a watershed algorithm. Because these are global segmentation algorithms, it is difficult to identify parameters that are applicable to a large number of images. The parameters for the threshold and watershed can be chosen from a predefined set. The methods used, in this thesis, offer a good basic segmentation and will be explained in the next sections.

3.2.1 Threshold-Based Segmentation

A threshold defines a global upper and lower limit for intensity values. With this interval, which can be extended to use multiple intensity intervals, a binary image is generated. This pattern recognition technique is qualified for generating a basic segmentation mask. Its possible to detect disturbing structures like bones, which can be removed to highlight the vascular structure. Since thresholds are mostly straightforward, they are often used as the first step in a segmentation workflow. The important part of this method is the threshold selection. Only if a meaningful interval is chosen, the region of interest is sufficiently segmented. Mostly this section is supported by the histogram of the image, where a local minima distinguish two tissue types. The second and here used way, to select the thresholds, is to have knowledge about the data. If the image modality and the body region of interest is known, estimations about the intensities of the tissues are possible.

The advantages of the intensity-based threshold segmentation and why we used it, is because the method is mathematically less complex, easy to implement, robust and fast to calculate even on big images. The disadvantages of simple threshold-based segmentation however are that we only consider the intensity values, which ignore any relationships between the voxels. Therefore voxels at edges, or due to noise, can be missed. Noisy images can also generate false background images, which can lead to problems with following algorithms and missing boundaries. And last even with histograms and knowledge of the image, it is difficult to adjust the threshold interval [9].

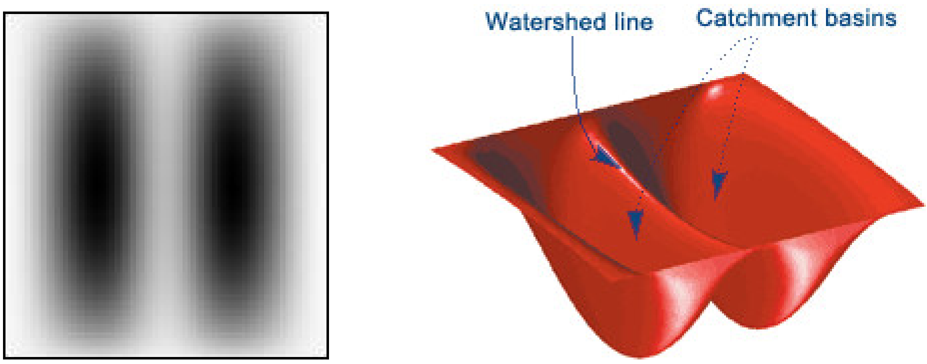
3.2.2 Watershed Segmentation

Another pattern recognition method is the watershed segmentation. We use the originally mathematical morphology approach, because it has been successfully applied to many segmentation task, like the extraction of the human brain. The watershed segmentation uses a type of region growing method in combination with an gradient image. The image with its intensity values get interpreted as topographic landscape, with hills and valleys like in figure 3.2. Catchment basins are the local minima from which the surface gets flooded, and partition the image into basins and watershed lines. By flooding the surface the catchment basis merge at their watershed location sooner or later as illustrated in 3.3. These merged region define the segmented areas. Our framework can then characterize these areas by their flood level and the initial global threshold, the so called pre-flooding.

We used the watershed segmentation, because the boundaries form closed and connected regions due to the gradient image, which can not be achieved with the threshold method. This yields more accurate results and is still robust and reproducible, which is in section 2.1 described. Also because of the basins and watersheds the locality is increased, which makes parallel implementation possible. This method can also be applied in different situation due to the numerous implementation types, like the Marker-based watershed or the rain-falling watershed. The main disadvantages of the watershed segmentation are that if too much noise or local irregularities are located in the data, many small region arise which lead to over-segmentation.

3.2.3 Segmentation Masks

In our approach any arbitrary segmentation mask can be used for the comparison. Because of the huge amount of segmentation algorithms and ways to implement them,



(a) Initial Synthetic Image. (b) 3D presentation of the Surface.

Figure 3.2: The two images show the principle of the watershed transformation. From the synthetic image (a) with two dark blobs, bright areas are taken as high values and dark areas as low values. Therefore we can see basins in the dark areas and otherwise watersheds (or hills) in the brighter areas (b). Pictures are taken from [21].

(a) Initial Image. (b) Topographic Surface. (c) After serveral floodings.

Figure 3.3: The illustration shows the watershed process applied to the initial grey image (a). From the initial image a gradient image is generate (b), in which high intensities form hills and low intensities form valleys. Then the surface gets repeatedly flooded which combines regions (c). The images are taken from [22].

we used the application MeVisLab to generate segmentation masks for the evaluation. This was necessary to test the comparison and visualization with different segmentation results. In the next section we explain the generated segmentations. In order to get other results than with our own algorithms, we used a region growing technique as well as an hessian based technique and their combination. The results can be seen in figure 3.4.

Region Growing Technique

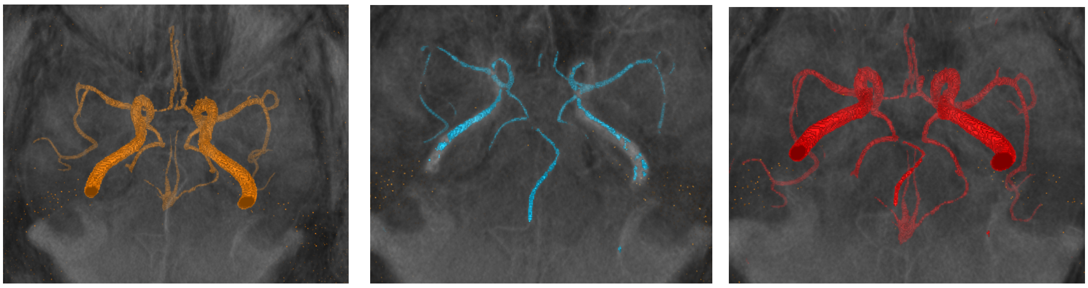
In this approach we preprocessed an image with an threshold and then segmented it with an region-growing method. Several seed points in the arteries were chosen to get a precise segmentation. The results is shown in figure 3.4a.

Hessian-based Technique

This technique computes the Hessian matrix of an image by using derivatives of a Gaussian. With these filters it is possible to compute the vesselness and detect bright, tubular structures in the image. Mostly very small vessels are segmented by this method, which can be seen in the image 3.4b.

Combination of Region Growing Technique and Hessian-based Technique

The combination of this methods is achieved with the summation of the two segmentation results. In our approach a very accurate segmentation of the Circle of Willis is possible, which is depicted in 3.4c.



(a) Region Growing technique. (b) Hessian-based Technique. (c) Combination of (a) and (b).

Figure 3.4: The three figures show the output of the MeVisLab segmentations in our framework. Every result is achieved through a different segmentation technique. As can be seen in figure (c) the combination of the region growing method and the hessian based approach yield a good result.

3.3 Visualization

After the segmentation the next step is the evaluation and comparison. Here the segmented images get analysed and rendered. This stage deals with the process of visually illustrating differences and similarities in the results, which is often called comparative visualization. Our framework takes one to three images and checks for every voxel, which algorithm has indicated it as vessel. After every check the voxel gets an parameter with an color, which belongs to one of the chosen algorithms, to mark it. After marking every found vessel on a new image, it gets forwarded to the visualization stage, in which the rendering process is started. There are three specific widgets for the visualization: A 2D slice view, a 3D view and a maximum intensity projection. These three widgets have an unique transfer functions, which are mapped to the colors of the algorithms or a distribution map to utilize the available information in the most efficient way.

3.3.1 Evaluation and Comparison

Before the segmented images can be evaluated, they have to be compared against each other. The matches as well as the mismatches of the segmentations have to be detected in every slice and made visible. Therefore the images get analysed by the framework and every voxel gets assigned to the algorithm, which has detected it as vessel. After that the segmentation masks are compared and the data is delineated. Then a new evaluation mask is generated out of this information. This mask serves as atlas to identify every voxel and its belongings to an algorithm. In the final step the data is forwarded, rendered and displayed.

3.3.2 2D View

The first method to display the information of the evaluation, is a two-dimensional view. The three-dimensional medical data is depicted as several slices. In the framework this representation is utilized because the majority of radiologist are skilled in handling slices of medical imaging data, as Wiebel et al. noted [23]. With this visualization a more precise and distinct representation of the data is possible, which enables to distinguish fine details. To even further enhance this precise examination we highlight the vessels and blend a grey-value reference image onto it. The blended image emphasizes vessels, which are not detected and contributes to the identification and localisation of the surrounding area. The framework displays the current slice number and enables the scrolling trough the slices with the mouse wheel, as well as to zoom in and out with the right mouse button. It is also possible to move the image with the middle mouse button to view a certain scope.

3.3.3 3D View

Although the two-dimensional view enables a detailed representation of the anatomy, it is unsuitable to provide a general overview. Therefore this visualization serves as comprehensive impression of the vascular system. To display a three-dimensional structure the direct volume rendering technique (DVR) is used. This informative and cutting-edge visualization facilitate the display of volumetric data by considering the data as a semi-transparent light-emitting medium, which enables a three-dimensional representation [23]. The framework analyses the evaluation mask and maps every voxel of the data, with a transfer function, to an opacity and color value. The function is used without interpolation to ensure a accurate representation of the voxels. Like in the 2D view a grey-value reference image is used to blend in missing details of the anatomy. To make the segmented vessels highly visible in the 3D representation, they are colorized like in the 2D view. An advantage of this method is the interactive use of the volume, which allows rotations in all directions with the right mouse button, as well as moving the image with the middle mouse button. Further it is possible to zoom in and out, with the mouse wheel, in different angles to analyse the segmentation closer.

3.3.4 Coloring

To improve the depiction of vessels and enhance the information content, the segmentation in both 2D and 3D representation is colorized. The process, in which the vessels are highlighted, depends on the amount of chosen algorithms. When up to two segmentations techniques are selected, the voxels are marked in the color of the particular technique, which found them. The color, in which the voxels can be marked, is illustrated as the color of the technique's name in the legend. When both algorithm specify a voxel as belonging to a vessel, it is marked green. If only the first algorithm finds a vessel the voxel is marked orange, if the second method detects one it will be made blue. Through this interpretation a distinct differentiation of the various segmentation techniques can be made. Therefore it will be visually highlighted which algorithm found a certain voxel, as illustrated in figure 3.5. When more then two algorithms are selected, it is not possible to assign a unique color for every possibility. In the case that two algorithms found the same vessel, it is not possible to visually label, which two segmentations found the voxel. To still be able to visually quantify the amount of techniques, which segmented the voxel, a new method for their representation has to be applied. Therefore our framework uses a distribution map for displaying the results.

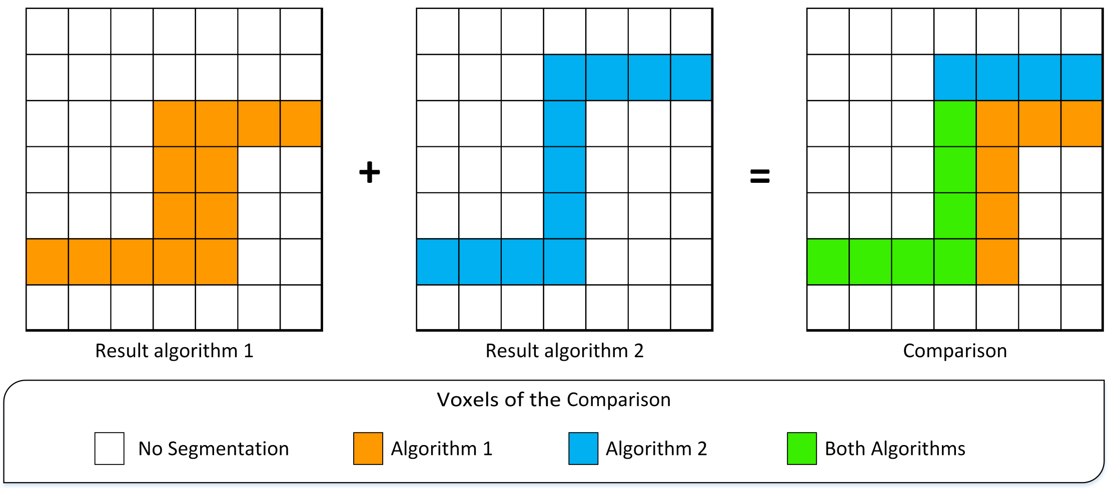


Figure 3.5: The illustration shows how two segmentation results are compared and generate an evaluation mask. This mask is labelled with unique colors to identify which algorithm found a certain voxel. The first algorithm is marked orange and the second one is blue. If a voxel is determined as vessel from both techniques, it is labelled green.

Distribution Map

To enable an intuitive way of presenting more than two algorithms at once, we use a distribution map. With this method it is possible to encode the differences and similarities through a mapping, based on colors. Thereby the amount of algorithms, which segmented a voxel, is represented with an unique color. This can be seen in figure 3.6. If one algorithm finds a vessel, it is getting marked blue, if two techniques find it, it will be labelled green and if three found it, a red color is used. Due to this depiction the user is able to identify which voxel got segmented frequently by various algorithms. The distribution map thus grant a good overview of the evaluation.

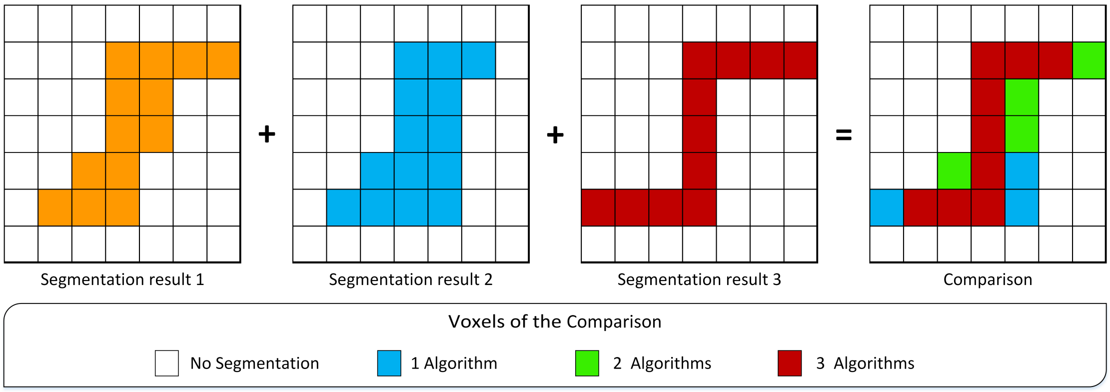


Figure 3.6: For the 3D visualization a distribution map is used. In this illustration three algorithms are compared against each other. The differences and similarities of the three segmentation masks together create an evaluation mask. In this mask it is visually not possible to identify a certain algorithm but rather the amount of techniques, which identified a voxel as belonging to a vessel.

Color Guidelines for Visualization

In order to be able to differentiate between more then two algorithms, the choice of colors is important. A suitable color for the display of information is crucial if the interrelationships and design within data should be easy to observe, as explained by MacEachren et al. [24]. The aim was to find a diverging scheme for the distribution map, which emphasizes on low and high values. For two algorithms it was important to find colors, which have a good dissimilarity and highlight the label. The label marks that both techniques found the same voxel. With the guidelines of Cynthia A. Brewer written in [24] and the online tool "ColorBrewer" [25], appropriate colors for the visualization were chosen.

3.3.5 Maximum Intensity Projection

Another Method to display the vascular structure is the maximum intensity projection (MIP). This visualization technique was chosen because it enables the user a good three-dimensional view, which assists the viewer's perception. Therefore the depiction helps to orientate oneself in the volume and allows to asses contrast-enhanced vascular structures. This method uses ray-tracing to display the voxels with the highest intensity for each pixel in the view plane [9]. It is applied on the unsegmented images and no specific transfer function is required. The voxels are usually mapped linearly to the brightness of the grey values of a reference image. This is done in order to not impair the impression, a coloring was renounced. The advantages of MIPs is the fast calculation and the good overview of the vascular structures. On the other hand the technique poorly reflects depth relations and small vessels can disappear. Due to that fact it only serves as additional support for experts. The framework enables the rotations of the MIP in all directions with the right mouse button, as well as to zoom in and out with the mouse wheel. Further the image can be moved with the middle mouse button to rearrange the scope.

3.3.6 Arrangement of the Views

Due to the limitations in screen space, a suitable arrangement of the provided views has to be found. The design of the layout is important to deliver the results in an intuitive and informative way. To achieve this goal, approaches like juxtaposition, superposition and special encodings, which are described in section 2.4, are used. We combine in our framework all three techniques to obtain the best possible visualization. The design of the result window can be seen in 3.7. To get a good overview of the segmentation result, a 3D view is placed on the lower left side. When the viewer needs a general view of the surrounding structure a glance at the lower right side shows him the MIP. If the user has found the region of interest with the MIP and the 3D view, he then can investigate the area in more detail by looking at the 2D view in the upper left corner. To not only have the visualization in focus, the calculated statistics and the colors for the algorithms are shown in the upper right corner. On the left side of both the 2D and 3D widget the user can switch the current displayed result of the respectively technique with radio buttons.

3.4 Statistics

To enable an objective and automated evaluation, our framework consist of different metrics to measure the differences and similarities. After the comparison of the segmentation results the statistics, out of four parameters are shown. For computing, more than two techniques have to be selected. The metrics are compared against the first chosen algorithm or against the pre-segmented image. This evaluation is composed out of:

- **Runtime**
  When an intern segmentation algorithm is used, the execution time is captured. This time, measured in seconds, can be compared against other techniques. If a pre-segmented image is used, the runtime is displayed as zero.
- **Volumetric Overlap Error**
  This overlap error is a modified jaccard coefficient. It is one of the most popular

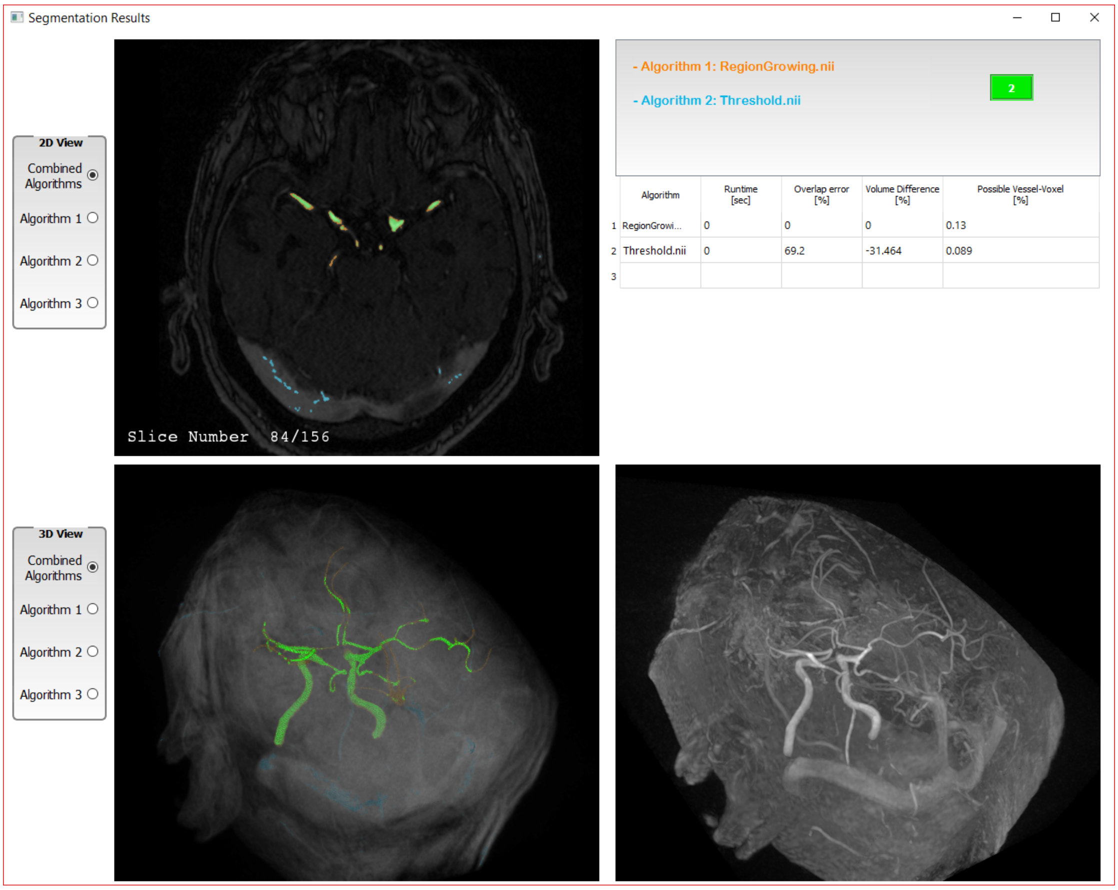


Figure 3.7: The image shows the design of the result window in the framework, where two algorithms were selected. The arrangement of the layout is so conceived that each visualization is easily perceived. In the lower left corner the 3D segmentation is placed, which gives a good view over the found vessels. The lower right corner contains the MIP, which is used to get a quick overview over the surrounding. In the upper left corner a 2D representation is placed to inspect the region in more detail. At last statistics, calculated with metrics, are displayed on the upper right side. Because only two algorithms are selected, the legends show the colors of the respective technique, with which the voxels are marked

metrics to measure the segmentation accuracy. The result is a value between zero and hundred percent. We used this metric to get an insight about the overlapping voxels.

- **Relative Volume Difference**
  The volume difference between two sets of voxel is given in percentage. With this measure the framework detects over- and under-segmentation.

- **Vessel to Voxel Ratio**
  Another metric to identify how many percent of vessels are found is the "Vessel to Voxel ratio". It is similar to the Dice coefficient and measures the percentage of voxels, which were identified by the first technique and the other segmentation methods. Because there should be no need for a ground truth, the voxels which the segmentations have found, is divided by the amount of grey values in the unsegmented image. The grey values represent the voxels, which can be possible vessels.

3.5 Implementation and Tools

Our framework is composed out of three different libraries and is implemented in C++. For the segmentation of the medical data the open-source library "Insight Segmentation and Registration Toolkit" (ITK) version 4.10 was used. To enable a qualitative image processing and visualization the open-source "Visualization Toolkit" (VTK) library in version 7.0 was utilized. An intuitive graphical user interface was developed with the cross-platform library Qt, in the version 5.7. The framework was implemented on Visual Studio 2013 in 32Bit. The tool MeVisLab was used to segment some images and to speed up the prototyping process.

MeVisLab

MeVisLab is a powerful framework, which enables developers to prototype segmentation processes and interact with their results. As the developers of MeVisLab [20] state "MeVisLab represents a powerful, modular framework for image processing research and development with a special focus on medical imaging." In this framework it is possible to choose from different modules of the ITK (Insight Segmentation and Registration Toolkit) and VTK (Visualization Toolkit) library and build your own processing pipeline, from segmentation to visualization, which is illustrated in 3.8.

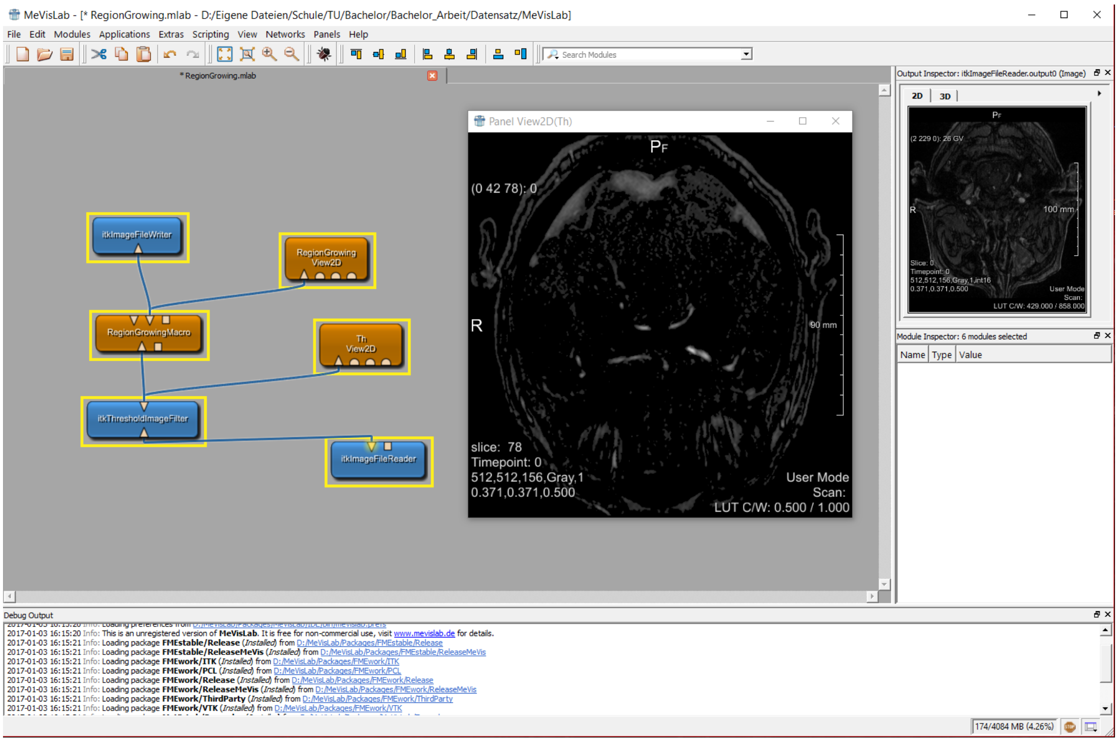


Figure 3.8: Screenshot of the MeVisLab Interface. In this picture various modules are used to build a segmentation pipeline, where at the end results can be displayed in 2D views.

CHAPTER 4

Results and Discussion

To demonstrate the proposed methodology from chapter 3 we analysed different MRA images from the human brain. In this chapter, we will evaluate the different segmentations and metrics, as well as the various comparative image visualizations. We will be highlighting different regions and vessels from the medical datasets to illustrate the applicability of our approach and show the effectiveness of the framework. To give an comprehensive insight of the utilization, the chapter is divided in three main section: An overview of the interface and structure of the framework, and a detailed workflow for both cases of two, or more selected algorithms, is described. This should give a good overview over the usage and show the possible fields of applications.

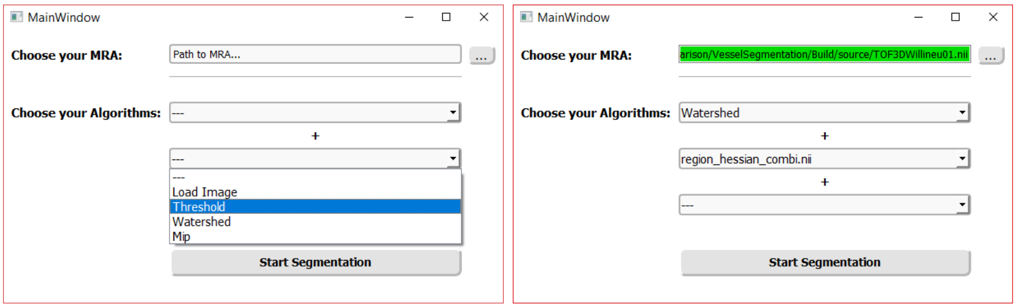
4.1 Structure of the Framework

Main Window

The framework is divided in two windows. At the start the user has to choose the necessary inputs in the menus of the main window, which is shown in figure 4.1. First a grey-value MRA image has to be chosen and loaded. This image serves as unsegmented reference. When the loading was successful, the user needs to select up to three algorithms. He can choose from the already implemented techniques or load pre-segmented images. After the inputs have been provided, the segmentation is started with the "Start Segmentation" Button.

Result Window

The next window, which appears, displays the different visualizations, as well as the statistics. Here the results of the evaluation are shown. The basic structure of the Result window is is identical when two or more algorithms are used, apart from the legend in the left corner, which changes with the amount of chosen techniques. The legend display in



(a) Freshly opened main menu. (b) Main menu after choosing an MRA.

Figure 4.1: The two figures show the main menu after starting the framework. At the start the menu has no MRA and no algorithms selected, which is illustrated in (a). After choosing a MRA and successfully loading it, the text-line displays the path to the image in green. Furthermore it is visible in (b) that techniques have been chosen and are ready to be applied by pressing "Start Segmentation".

both cases the current applied segmentations and current loaded pre-segmented images. When two algorithms are selected the legend shows the technique's name in the color of the voxel, which it has segmented. This is shown in figure 4.2. If more than two methods are chosen, then the range of the color scheme is displayed, which stands for the amount of techniques, that segmented the voxel. This different legend is illustrated in figure 4.5.

The Result window is designed in such a way, that the greatest possible amount of information is perceived by the viewer. Therefore the framework divided the evaluation results in three specific visualizations. The separation or juxtaposition method, which is described in section 2.4, is used to assist the process of comparing the differences and similarities. Because the framework allows to swap the images in both 2D and 3D views with radio buttons, the segmented results can be compared simultaneous in the different widgets. When the window shows up, the two separated depictions will be displayed. These show the comparison of the selected techniques. With the radio buttons it is possible to show the colored segmentation results of the first algorithm in the two-dimension widget and also the result of the first algorithm in the three-dimensional widgets. To get a better overview of the surrounding areas in the representation the framework uses superposition to blend in the grey-value image. This allows to add more information without distracting with an additional separated image. To even further enhance the visualization and highlight the detected vessels an explicit encoding is used. Therefore the framework emphasizes the relationship between the techniques and their found voxels with an unique color. Altogether our system offers the features, which are described in section 2.4. The system can compare sets of images and gives a good interactive overview, over the evaluation results. Further the framework is quite flexible because it compares the result also in an objective way with metrics.

4.2 Comparison of two Algorithms

In this section we focus on the comparison of two algorithms. This enables the detailed analyses of the segmentation techniques, which segmented a voxel. The evaluation of two algorithms instead of more, helps to emphasize the precise differences and similarities of the detected structures. The first test case aims for the general comparison of the segmented vessels and the second case focuses more on the objective comparison after an editing of the results.

4.2.1 Subjective Evaluation

The first test case, which will be demonstrated by our framework, is the segmentation and evaluation of a MRA with two selected algorithms. The medical image dataset contains all essential arteries of the Circle of Willis according to the gold standard. For the comparison the image will be segmented with an implemented watershed algorithm and the pre-segmented result of the region growing - hessian approach, which is described in section 3.2. The second scenario has a different data set, in which the ACoP dexter and a part of the ACP sinister are missing. This information was given by experts. In this test case we use a pre-segmented image, which was segmented with a threshold.

First Scenario

Due to the informative visualization of our framework the compared results can be analysed precisely. In both representation it can be unambiguously seen which technique segmented which voxel. As can be viewed in the result window in figure 4.2 the watershed algorithm is orange and the region-hessian approach is blue. At the first look we can interpret that the watershed method finds small vessels but due to the noise in the image many false positive results appear. This conclusion can be seen in the 2D view, where a lot of the cerebral cortex and other structures are segmented. Another insight is provided by the 3D widget, in which we discover that most of the bigger vessel are green, which means that both algorithm found similar voxels. In the 3D view very few blue voxels are displayed, which means that the region growing - hessian approach has not found any additional voxels compared to the result of the watershed. To get a closer look onto the blue colored vessels, we have to view the more detailed 2D widget.

To view an example, in which the framework shows a big difference in the segmentation, we investigate the region shown in different views in figure 4.3. In the image 4.3a an axial representation of the brain is shown. In this picture we zoomed the Circle of Willis, where we marked a vessel yellow. This was done in the 3D view as well as in the MIP, where the vessel was not found by the two techniques. The AB was segmented successfully by both algorithms, which is shown green in the segmentation. The ACoP sinister, which is marked with a yellow arrow, however was only segmented by the watershed method. This is displayed in figure 4.3b. Then we zoomed in a little bit further and rotated the view to better see the segmented region in the 3D widget as well in the MIP. The MIP gives a good overview of the surrounding, but has wrong depth relations, which is noticeable at the A. carotis interna. In figure 4.3c we zoomed in even more to emphasize the correct segmentation of the ACoP sinister by the watershed technique in orange. In the first image (figure 4.3a) we can detect that not only the watershed, but as well the region growing - hessian approach have successfully segmented the ACoP dexter.

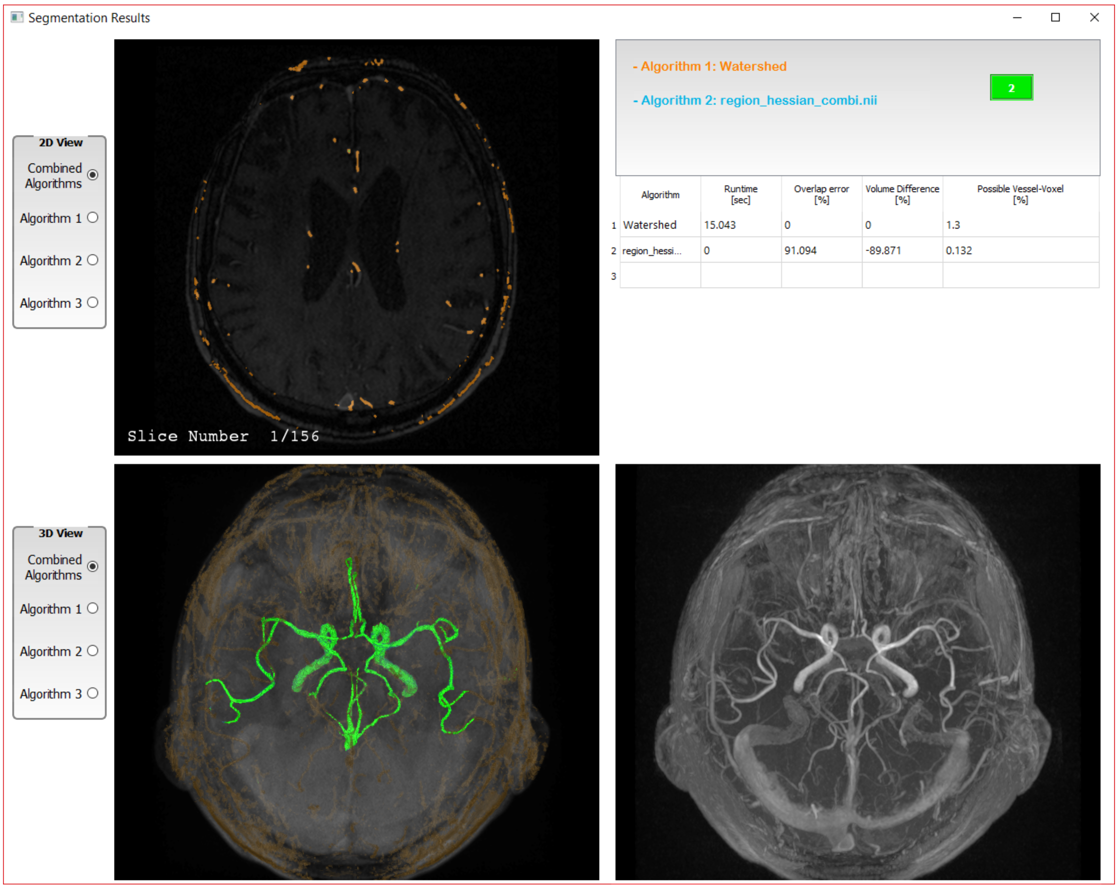
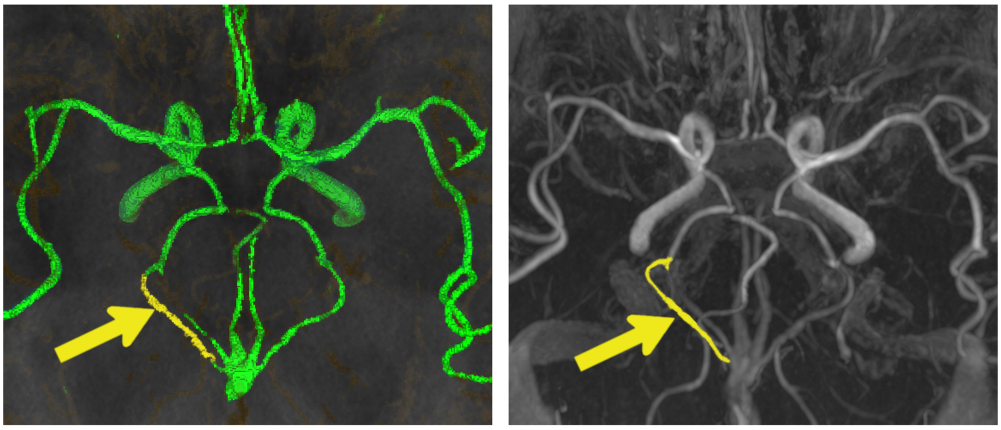
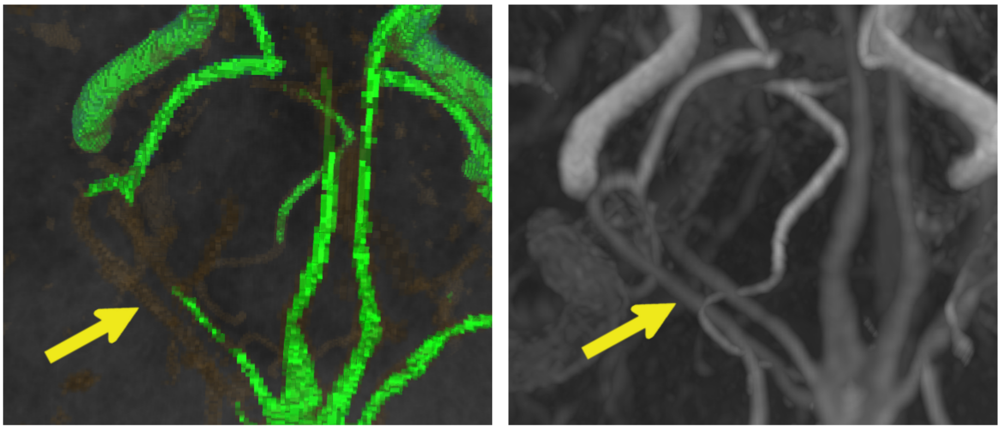


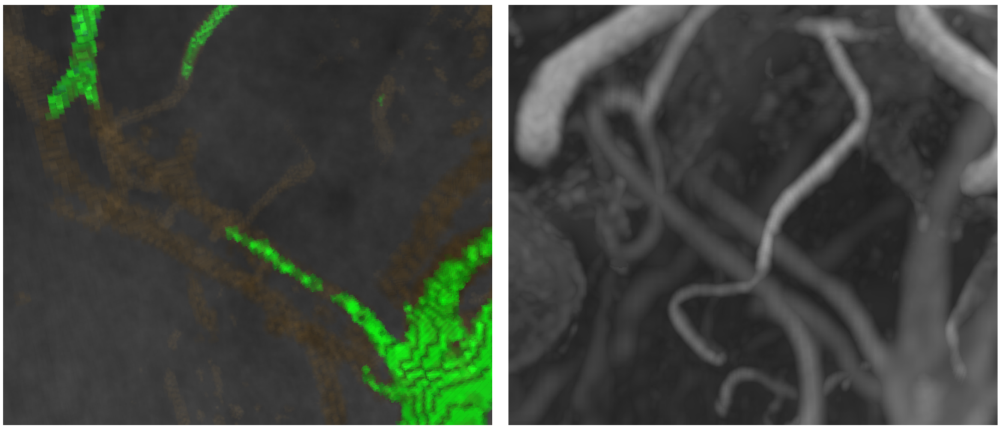
Figure 4.2: The image shows the Result Window of the comparision with two algorithms. In this case as first algorithm a watershed technique was used. The second method, which will be compared against the watershed algorithm, is the pre-segmented region growing - hessian result. The window shows in the right upper corner the legend with the colored names of the techniques as well as the statistics for the current evaluation. Below, hence in the lower right corner, lies the MIP which gives a general overview. In the upper left corner is the 2D view for detailed examinations and below, in the lower left corner, is the 3D view for the environmental representation over the found vessels.



(a) Axial view with yellow marked ACoP sinister.



(b) Rotated axial view of the ACoP sinister.



(c) Zoomed in rotated axial view of the ACoP sinister.

Figure 4.3: These images show the 3D view and the MIP of the segmented Circle of Willis. From picture (a) to (c) they are getting rotated and zoomed in, to get a good view of the ACoP sinister. In the images we can see the green voxels, which both algorithms (watershed technique and region growing - hessian approach) have segmented. The yellow marked vessel is the ACoP, which was only segmented by the watershed algorithm in orange.

Second Scenario

In the second scenario we analyse a pre-segmented image, which was segmented by a threshold. To highlight the comparison process and provide a better insight in the objective evaluation, we compare the threshold filtered image with the same result, which was edited outside the framework. As can be seen in figure 4.4 we have a slightly rotated axial view of the Circle of Willis. The vessel, which we investigate closer, is the ACoA, which closes the anastomosis. In the MIP we can see that the connection between the left side of the ACoA and the ACA is present. In the original segmented image this link was missing. Therefore the image was edited outside the framework to correct the segmentation. In the 2D as well as in the 3D view, we can see the comparison of the unedited image which is orange and the edited result in blue. The orange voxels are missing, because both results have the same voxels segmented except the gap, which was corrected and is shown blue. In the 2D view we can see the difference in more detail.

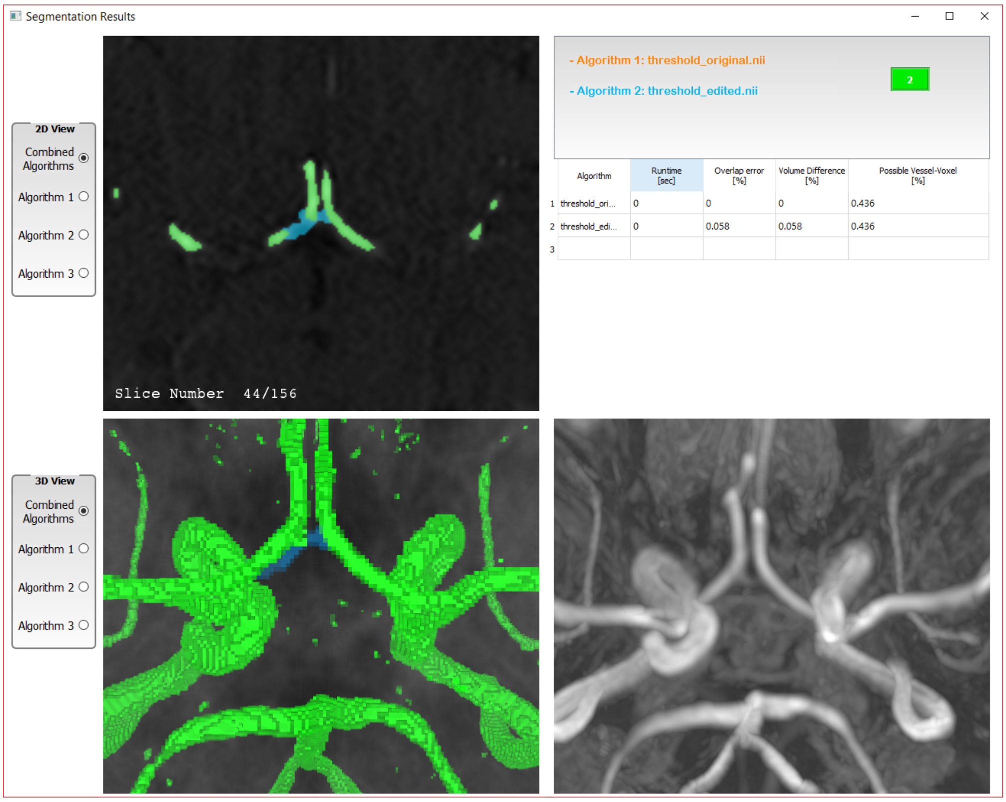


Figure 4.4: This illustration shows the comparison between a threshold based segmentation and its edited and corrected version. As can be seen in the 2D and 3D view the blue voxels are the edited part, which connects the left side of the ACA and the right side by creating the ACoA. Because the images do not have any other differences, the remaining voxels are all green.

4.2.2 Objective Evaluation

First Scenario

Not only visually the differences and similarities can be detected. As can be seen in the left corner of figure 4.2 an objective evaluation is presented. Here is stated that the execution time of the watershed algorithm only needed approximately 15 seconds. Further the fact, that noise leads to segmentation errors in the watershed method, is verified due to the fact that the algorithm has 1.3% of the grey values detected as vessels. The region growing - hessian approach has only detected 0.13% of the image as possible vessel. Also this technique overlaps for 9% (91% overlap error) and has a volume difference around -90%, which indicates an undersegmentation. This evaluation increases in significance with optimization of the technique's parameters. With every change the segmentation can be adopted in hindsight of the new evaluation values. Therefore this results give a hint that a reduction of the noise will increase the segmentation accuracy and adjust the overlap error towards zero.

Second Scenario

More significant statistics can be observed in scenario two, which is illustrated in 4.4. Here we can notice the influence of two metrics: the overlap error and the volume difference. The framework indicates with a overlap error of 0.06% that the difference between the images is marginal. This also can be observed, with the volume difference, which shows a difference of again 0.06%. That means that 99.94% of the voxels are equally segmented and that in the edited image additional voxels were found. The minor change is so small, that the Vessel to Voxel Ratio is around 0.44% and shows no change between the two images. With the other two metrics we can clearly detect the modification of the image, which can be verified visually. This test case illustrates that the framework shows even small changes in the widgets, as well as through the metrics.

4.3 Comparison of three Algorithms

In this section we will demonstrate two test cases to show the comparative visualization with more then two algorithms. To emphasize the comparison between the different segmentation approaches, we chose the same data set, as in the first test case from the last section. This should help to make the comparison and evaluation process more visible without distraction, due to a changed underlying MRA image. Thus, all essential arteries of the Circle of Willis are once more completely contained according to the gold standard. For the comparison three pre-segmented images were chosen. The first segmented image was created with a threshold, the second with a hessian approach and the last one with a region growing technique.

4.3.1 Subjective Evaluation

The two separated scenarios are intended to illustrate the different insights, which can be obtained through the evaluation of the framework. Both test cases have the same segmentation techniques but show different regions of the brain. The overview of the evaluation can be seen in the result window in figure 4.5. Here we can detect, in the upper left corner, the changed legend. The distribution map is displayed in a color scheme from black over blue and green to red, which shows the amount of techniques that segmented a voxel. It is also visible that the voxels, found by the segmentation algorithms vary highly due to the different colors in the widgets. The vessels, which are detected by all three techniques, are marked red. As example the A. cerebri media dexter and sinister are successfully segmented in this result.

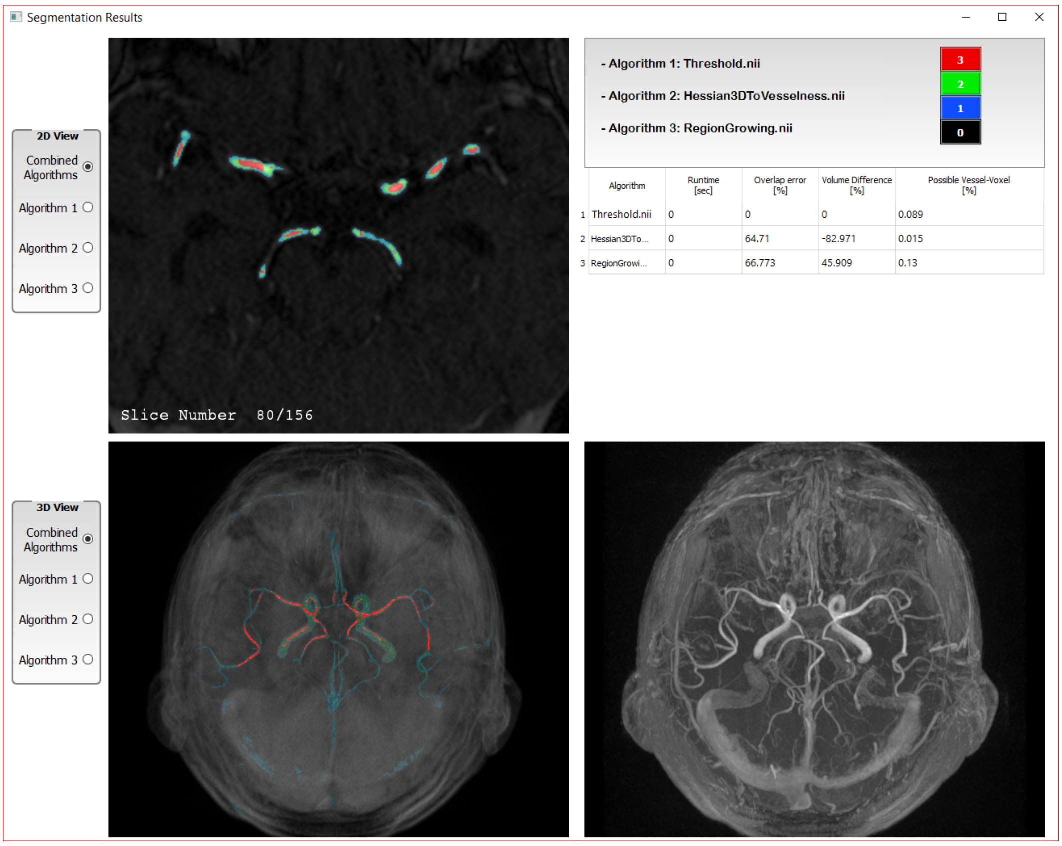


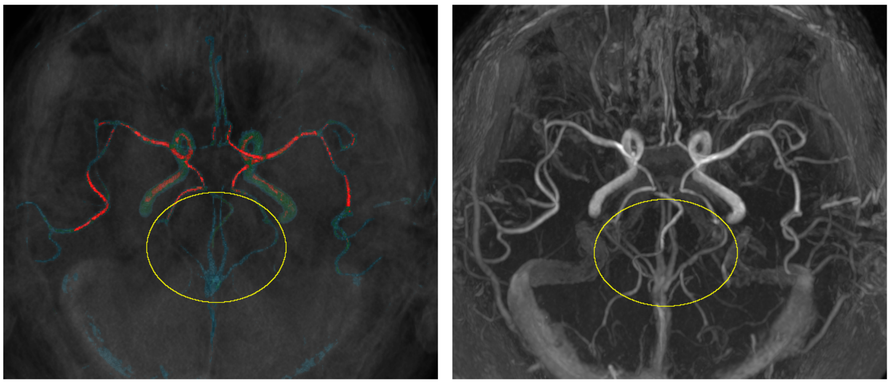
Figure 4.5: The image shows the Result Window of the comparison with three algorithms. As techniques a threshold, a hessian and a region growing method were used. The general structure of the Result Window is identical to the case with two algorithms. One modification happened in the legend. Here we have not colored the individual technique, but a color scheme, which labels the amount of methods, that segmented a voxel.

First Scenario

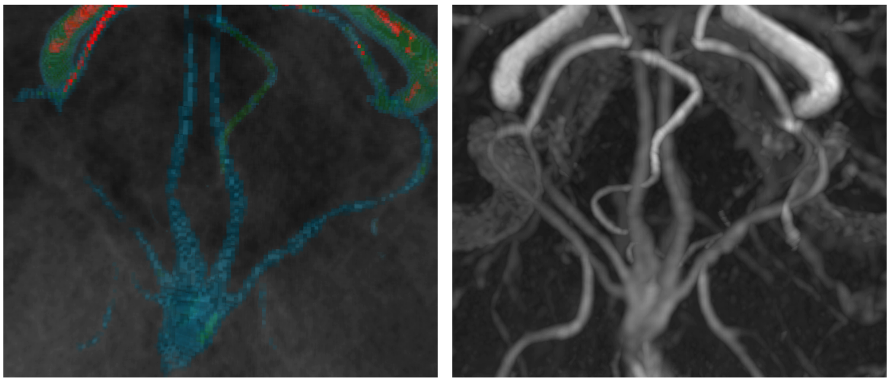
The first scenario can be seen in figure 4.6. In the illustration 4.6a an axial overview of the Circle of Willis is given. The region, which we analysed closer, is yellow encircled. This area contains the AB as well as the ACoP dexter and sinister. In the next image 4.6b the region is zoomed in, to examine its voxels more precisely. As can be seen by the blue voxels, the AB and its surrounding has nearly only been segmented by one algorithm. Some green voxels indicate that a second algorithm detected some structures of the vessel. The ACoP dexter and sinister are also almost completely detected by one technique. To distinguish now which method found the currently viewed area, the framework enables to switch between the selected algorithms in both 2D and 3D widgets. The radio buttons, which facilitate this effect, are displayed in the result window in figure 4.5. As in the illustration 4.6c can be seen, we switched from the 3D comparision view to the region growing depiction. Here it is possible to identify that this is the technique, which segmented the AB as well as the ACoP dexter. To analyse another result, we picked the threshold, which is shown in figure 4.6d. The picture shows the reason for some green voxels in the area of the AB in the comparison view.

Second Scenario

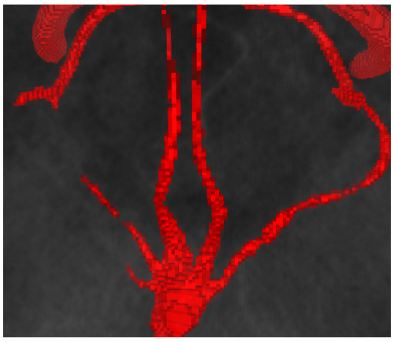
In the second test cases we examine the upper part of the Circle of Willis. The region of interest is yellow encircled in figure 4.7a. Here we look at the ACoA, which is zoomed in the image 4.7b. As can be seen in this case the vessel was not detect by all three algorithms and only the branch points were found by more then one technique. Further it can be noticed that the connection between the ACoA and the left part of the ACA is missing. In the MIP a link between them is assumable. To further enhance this guess, the framework enables a more detailed 2D view. This precise representation is shown in figure 4.7c. Here the visible gap is marked with an yellow arrow. In this representation we can clearly examine the individual voxels. Further it is, due to the blended grey value image, visible that a connection between the ACoA and the left part of the ACA should exist. The points on where the vessels would merge is marked green, which indicates that more then one algorithm has segmented the branch points. In this view as well as in 4.3b it is also visible that the cranial fragments of the ACA were segmented partly by all three methods.



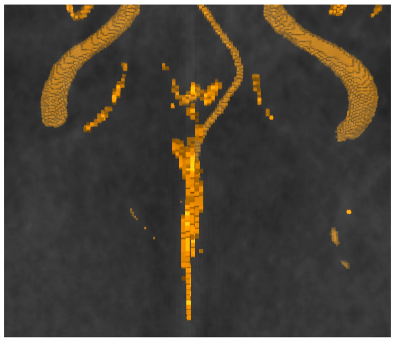
(a) Axial view of the AB marked with a yellow circle.



(b) Zoomed in axial view of the AB with surroundings.

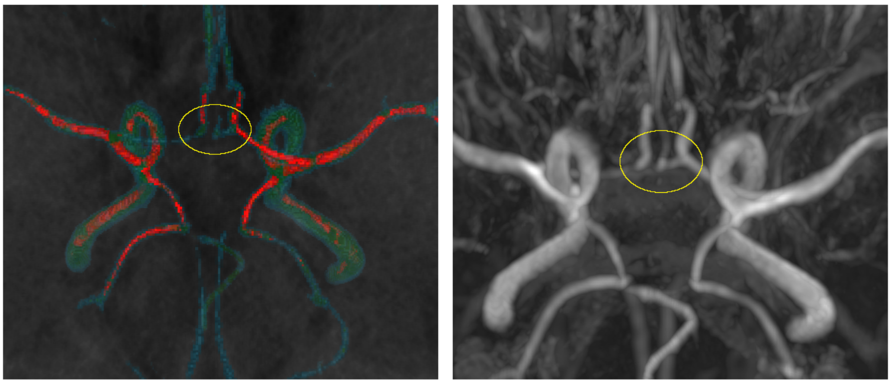


(c) Area of the AB segmented with Region Growing.

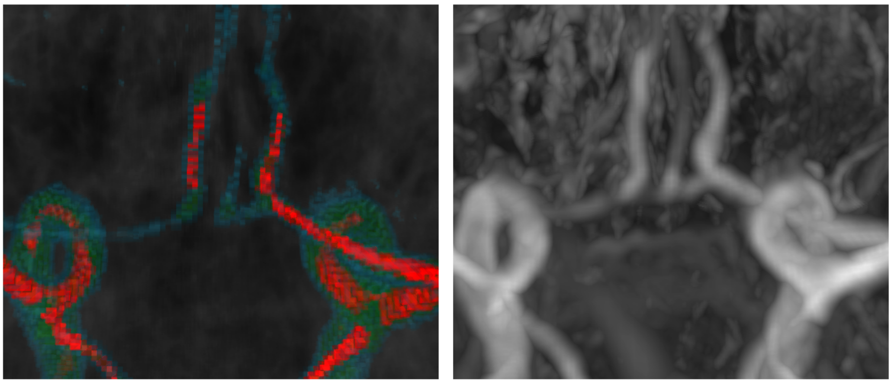


(d) Area of the AB segmented with Threshold.

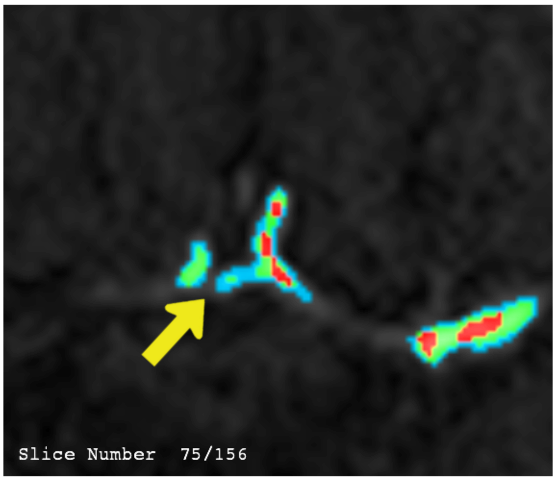
Figure 4.6: In this illustration the 3D view and the MIP of the segmented Circle of Willis is visible. The figures (a) to (b) are zoomed in to better visualize the AB and their surrounding vessels. The interesting region is yellow encircled and shows that only one of the three algorithms found the AB (blue color). Further it can be seen that no algorithm detected the ACoP sinister. Figure (c) displays the region growing technique, which is the only method that could segment the area of the AB. In figure (d) the result of the Threshold algorithm is shown, which detected a few voxels of the AB and is the reason for the green voxels in the area of the comparison.



(a) Axial view of the ACoA marked with a yellow circle.



(b) Zoomed in axial view of the ACoA.



(c) Slice view of the ACoA with a marked gap.

Figure 4.7: This figure shows an axial view of the Circle of Willis in the 3D widget and in the MIP. In both cases here the ACoA was encircled in yellow. In image (b) we zoomed in the yellow marked area from figure (a) to display the vessel of interest in more detail. As can be seen here the ACoA is detected by neither technique. Further the vessel is not closed on the left side with the ACA. However some cranial parts of the ACA were detected by all methods and are therefore marked red. To better analyse these segmentations a 2D view is displayed in figure 4.7c with a gap marked with an yellow arrow. Here it is possible to detect the gap of the ACoA and examine the detailed visualization of the algorithms, which segmented the voxels in its neighbourhood

4.3.2 Objective Evaluation

For not relying exclusively on visual quantification, the objective statistics provide calculations for the comparison. These results of the metrics can be seen in figure 4.5 in the upper left corner. Because all three results originate from pre-segmented images the runtime was zero. Further it is visible in the table that the most voxels were segmented by the region growing technique, and not by the threshold algorithm. The least amount was found by the hessian approach. This fact is supported by the statistics, which indicate that the hessian approach yielded an undersegmentation due to the volume difference of -82.97%. The region growing technique however had a volume difference of 45.91%, which implies an oversegmentation. But it is still closer to zero and indicates a more identical segmentation. Another insight, that is provided by the objective evaluation, is the overlap error, which shows that 35.29% of the voxels in the first image overlap with the hessian approach (error of 64.71%). In addition the region growing overlaps with 33,23% for the first image (error of 66.77%). Again this statistics are useful for comparison with edited results or different parametrizations of the methods.

4.4 Future Work and Limitations

Although many visualizations, metrics and features are implemented in our framework, it still has some limitations. These limitations can be solved by future work, to make the approach even more efficient and useful. In the current state it is possible to compare many different segmentation results, however only three of them at the same time. It would improve the visualization and the calculation of metrics, when more than three techniques could be evaluated in the same execution.
Another limitation of the framework is the field of application, in which the metrics work. Currently the calculation only covers the volume of the anatomical structures. An improvement of the statistics would be to include the shape and appearance of the vessels. As result the framework would have the possibility to compare even the radii of the vascular structures.
Furthermore our test cases show that images with noise have a huge impact on the metrics. The additional voxels of the noise distort the computations of the volumes. The preprocessing step, in which noise can be reduced, depends strongly on the image acquisition and the prior knowledge of these parameters and is therefore not the focus of our work. But when built in, it would enhance the calculations, done for evaluation even more.

CHAPTER 5

Conclusion

In this thesis a framework for comparative visualization and objective evaluation was proposed. The comparison of medical data sets and their evaluation without the need of an expert are important. For this reason we presented methods to quickly analyse these segmentation results and detect differences and similarities at voxel level. The aim was to support doctors as well as researchers in the field of medical image analysis. We used various approaches to enable a subjective evaluation with different views as well as purely objective methods to not rely on experts or a ground truth. The use of special state-of-the-art visualization techniques combined with different views, like a three-dimensional-, two-dimensional and MIP-widget, enables an accurate comparison. Furthermore a particular data encoding with colors and a distribution map was developed to represent the information in the most efficient way. The framework supports up to three segmentation results, which can be separately viewed in 3D and 2D. Therefore unsegmented images may be used, whose vessels can be extracted with built in segmentation techniques, or pre-segmented results can be loaded. The comparison can then be subjectively or objectively evaluated based on the calculated metrics.

The feasibility of the described methods is demonstrated in the Result chapter. Here our work shows that the comparison is intuitive and at the same time precise enough to investigate which voxel belongs to which technique. It also makes the benefit of statistics visible. The metrics enable a detailed acquisition of information concerning changes. Even though the framework has promising results and can be used versatile, there are still features, which can be implemented. With an extension for the framework to compare even more segmentations at the same time, more information could be provided for the user. Also adding different metrics, which include the shape of vessels, can improve the accuracy for future tasks.

Acronyms

BA Arteria basilaris

ACP Arteria cerebri posterior

ACA Arteria cerebri anterior

ACoP Arteria communicans posterior

ACoA Arteria communicans anterior

MIP Maximum intensity projection

Bibliography

[1] K. Price, “Anything you can do, i can do better (no you can't),” *Comput. Vision Graph. Image Process.*, vol. 36, pp. 387–391, Nov. 1986.

[2] H. Zhang, J. E. Fritts, and S. A. Goldman, “Image segmentation evaluation: A survey of unsupervised methods,” *Computer Vision and Image Understanding*, vol. 110, pp. 260–280, May 2008.

[3] Y. Zhang, “A survey on evaluation methods for image segmentation,” *Pattern Recognition*, vol. 29, no. 8, pp. 1335 – 1346, 1996.

[4] M. J. Cipolla, “The cerebral circulation.” `https://www.ncbi.nlm.nih.gov/books/NBK53081/`, 2009. [Online; Accessed 18. Dezember 2016].

[5] L. T. Grinberg and D. R. Thal, “Vascular pathology in the aged human brain,” *Acta Neuropathologica*, vol. 119, no. 3, pp. 277–290, 2010.

[6] F. Anderhuber, F. Pera, J. Streicher, and A. Waldeyer, *Waldeyer - Anatomie des Menschen: Lehrbuch und Atlas in einem Band.* De Gruyter Studium, De Gruyter, 2012.

[7] “Blutversorgung des gehirns.” `https://de.wikipedia.org/wiki/Blutversorgung_des_Gehirns`. [Online; Accessed 4. January 2017].

[8] “Circle of willis illustration.” `https://en.wikipedia.org/wiki/Circle_of_Willis`. [Online; Accessed 27. December 2016].

[9] B. Preim and C. Botha, *Visual Computing for Medicine: Theory, Algorithms, and Applications.* The Morgan Kaufmann Series in Computer Graphics, Elsevier Science, 2013.

[10] C. Kirbas and F. Quek, “A review of vessel extraction techniques and algorithms,” *ACM Computing Surveys (CSUR)*, vol. 36, no. 2, pp. 81–121, 2004.

[11] H.-H. Chang, A. H. Zhuang, D. J. Valentino, and W.-C. Chu, “Performance measure characterization for evaluating neuroimage segmentation algorithms,” *NeuroImage*, vol. 47, no. 1, pp. 122 – 135, 2009.

[12] W. Crum, O. Camara, and D. Hill, “Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis,” *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1451–1461, Nov. 2006.

[13] Q. Wang and Z. Wang, *A Subjective Method for Image Segmentation Evaluation*, pp. 53–64. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[14] M. Polak, H. Zhang, and M. Pi, “An evaluation metric for image segmentation of multiple objects,” *Image and Vision Computing*, vol. 27, pp. 1223–1227, July 2009.

[15] T. Kohlberger, V. Singh, C. Alvino, C. Bahlmann, and L. Grady, “Evaluating segmentation error without ground truth,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 528–536, Springer, 2012.

[16] T. Heimann, B. van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. Cashman, Ying Chi, A. Cordova, B. Dawant, M. Fidrich, J. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmuller, R. Kitney, H. Kobatake, H. Lamecker, T. Lange, Jeongjin Lee, B. Lennon, Rui Li, Senhu Li, H.-P. Meinzer, G. Nemeth, D. Raicu, A.-M. Rau, E. van Rikxoort, M. Rousson, L. Rusko, K. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J. Waite, A. Wimmer, and I. Wolf, “Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets,” *IEEE Transactions on Medical Imaging*, vol. 28, pp. 1251–1265, Aug. 2009.

[17] M. Martin-Fernandez, S. Bouix, L. Ungar, R. W. McCarley, and M. E. Shenton, “Two methods for validating brain tissue classifiers,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 515–522, Springer, 2005.

[18] J. Schmidt, M. E. Gr"oller, and S. Bruckner, “Vaico: Visual analysis for image comparison,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, pp. 2090–2099, Dec. 2013. Demo: http://www.cg.tuwien.ac.at/ jschmidt/vaico/.

[19] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, “Visual comparison for information visualization,” *Information Visualization*, vol. 10, no. 4, pp. 289–309, 2011.

[20] “Mevislab.” `http://www.mevislab.de/`. [Online; Accessed 28. December 2016].

[21] M. Steve Eddins, “The watershed transform: Strategies for image segmentation.” `https://de.mathworks.com/company/newsletters/articles/the-watershed-transform-strategies-for-image-segmentation.html`. [Online; Accessed 3. January 2017].

[22] S. Beucher, “Image segmentation and mathematical morphology.” `http://cmm.ensmp.fr/~beucher/wtshed.html`. [Online; Accessed 3. January 2017].

[23] A. Wiebel, F. M. Vos, D. Foerster, and H. C. Hege, “Wysiwyp: What you see is what you pick,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, pp. 2236–2244, Dec 2012.

[24] A. MacEachren and D. Taylor, *Visualization in Modern Cartography.* Modern Cartography Series, Elsevier Science, 2013.

[25] “Colorbrewer.” `http://colorbrewer2.org`. [Online; Accessed 16. January 2017].