Aus dem OncoRay – Nationales Zentrum für Strahlenforschung in der Onkologie
Direktor: Herr Prof. Dr. med. Michael Baumann

Optimisation and Validation of a Swarm Intelligence based Segmentation Algorithm for low Contrast Positron Emission Tomography

D i s s e r t a t i o n s s c h r i f t
zur Erlangung des akademischen Grades
Doktor der Medizinischen Biometrie und Bioinformatik
Doctor rerum medicinalium (Dr. rer. medic.)
vorgelegt
der Medizinischen Fakultät Carl Gustav Carus
der Technischen Universität Dresden

von

Dipl.-Inf. (FH) Robert Haase

aus Karl-Marx-Stadt jetzt Chemnitz

Dresden 2014

1. Gutachter:

2. Gutachter:

Tag der mündlichen Prüfung:

gez.: _____

Vorsitzender der Promotionskommission

‘The purpose of computing is insight, not numbers.’

Richard W. Hamming (Hamming, 1962)

Contents

List of Figures

List of Tables

List of Abbreviations

ACO	Ant colony optimisation
A_{ROC}	Area under the ROC-curve
Bq	Bequerel, unit of radioactivity
C	Concentration, or rather tracer concentration usually given in Bq$/$ml
CNR	Contrast-to-noise-ratio
CT	Computed tomography
CTV	Clinical target volume
D	Dice index
$\bar{d}_c$	Mean contour distance
$\bar{d}_{\text{bil,c}}$	Mean bilateral contour distance
$d_{\text{e,min}}(a, B)$	Euclidean distance between a voxel a and a set of voxels B
d_H	Hamming distance
d_h	Hausdorff distance
$\delta\bar{V}$	Relative volume error
ΔV	Absolute volume error
$\bar{d}_{\text{wei,bil,c}}$	Mean weighted bilateral contour distance
ϵ_{Class}	Classification error
GTV	Gross tumour volume
FAZA	$[^{18}\text{F}]$-fluoroazomycin-arabinoside, a PET tracer for measuring hypoxia
FDG	$[^{18}\text{F}]$-fluorodeoxyglucose, a PET tracer for measuring glucose-metabolism
FMISO	$[^{18}\text{F}]$-fluoromisonidazole, a PET tracer for measuring hypoxia
FN	A set of false-negative elements
FP	A set of false-positive elements
FWHM	Full-width-at-half-maximum
J	Jaccard index
$K(A)$	Contour of A
MRI	Magnetic resonance imaging
$n_{\text{seed,rel}}$	Number of seeded ants, relative to the number of inspected voxels
PET	Positron emission tomography
PSO	Particle swarm optimisation
PTV	Planning target volume
p_{ij}	Probability of an ant of transiting from node i to node j, used in ACO algorithms
p_r	Reproduction probability of an ant, used in ACO algorithms
p_s	Survival probability, used in ACO algorithms
p_{TN}	Specificity
p_{TP}	Sensivity
ROC	Receiver operating characteristic

SUV	Standard uptake value
τ	Pheromone intensity
T	Threshold
T_{abs}	Absolue threshold
T_{rel}	Relative threshold
$T_{\text{opt,method}}$	Optimal threshold according to a given method
$T_{1/2}$	Half-life
TBR	Target-to-backgroud-ratio
TMR	Target-to-muscle-ratio
TN	A set of true-negative elements
TP	A set of true-positive elements
VOI	Volume of interest
vwROC	Voxel-wise receiver operating characteristic
Y	Youden index

1. Introduction

In modern radiooncology, therapy individualisation is one of the hot topics in current research. An important issue in this field ist biologically-adapted radiotherapy treatment planning based on findings from tomographic imaging. Computed tomography (CT) is used as the standard imaging technique for target volume delineation and radiotherapy treatment planning. In many tumour entities Magnetic Resonance Imaging (MRI) is applied to retrieve more detailed information about the tumour and its surrounding tissues. Positron Emission Tomography (PET) is increasingly applied in diagnostics for oncology, because the combination of molecular or metabolic imaging using PET with morphological imaging using CT provides for advantages radiotherapy treatment planning, response monitoring (Zöphel et al., 2008) and prediction of regions of local failure (Abramyuk et al., 2009). Combined PET and MRI scanners are furthermore available and increasingly clinically applied (Platzek et al., 2013) for some years incorporating the versatile imaging sequences available in MRI scanners with the possibility of molecular and biological imaging using PET. In clinical routine, standard clinical PET tracers such as $\left[^{18}\mathrm{F}\right]$-fluorodeoxyglucose (FDG) are used for supporting radiooncologists in defining an appropriate metabolic tumour volume for irradiation. Researchers are one step ahead: other tracers such as $\left[^{18}\mathrm{F}\right]$-fluoromisonidazole (FMISO) for hypoxia imaging are investigated for their potential for biologically-adapted radiotherapy treatment planning (Thorwarth et al., 2007; Lee et al., 2008; Thorwarth and Alber, 2008, 2010). No matter how biologically-adapted radiotherapy treatment planning based on PET imaging using different tracers may be implemented in the future, it will not be possible without a reliable technique for target volume definition. While for automatic target volume definition based on FDG PET a number of reliable and broadly observer-independent algorithms have been published (Daisne et al., 2004; Nestle et al., 2005; Schaefer et al., 2008; Hatt et al., 2009; Nehmeh et al., 2009; Hatt et al., 2011; Schaefer et al., 2012; Hofheinz et al., 2012c), this is not the case for FMISO PET. One reason is the limited contrast between target volume and background observed in clinical FMISO PET images (Abolmaali et al., 2012). Furthermore, reproducibility of FMISO PET measurements is still part of controversy (Nehmeh et al., 2008; Okamoto et al., 2013).

Aside from therapy planning based on biological imaging, these imaging techniques may be used for statistical analysis of patient cohorts. Researchers working in this field concentrate on therapy monitoring and therapy response prediction. It is known for longer that measurements of intratumoural partial pressure of oxygen is a potential prognostic tool for example for patient survival (Höckel et al., 1993). Non-invasive tumour hypoxia evaluation using FMISO PET may also be used for estimation of patient prognosis (Rajendran et al., 2006; Eschmann et al., 2007; Zips et al., 2012). Thus, FMISO PET may be

the basis of a decision support tool for therapists. But again, to determine FMISO PET imaging derived tumour properties such as mean signal intensity or the amount of FMISO-positive volume for further statistical processing, reliable volume definitions are essential.

The present thesis was developed in the environment of researchers investigating optimal FMISO PET imaging and analysis protocols (Abolmaali et al., 2012; Zips et al., 2012). As soon as researchers validated FMISO PET imaging to use it for therapy response prediction or biologically-adapted radiotherapy, a reliable tool for target volume definition in FMISO PET should be available. Thus, the aim of this work is optimisation and validation of a new segmentation algorithm purpose-made for FMISO PET image analysis presented earlier (Haase, 2010). Algorithm development must not end with a proof-of-principle, as in numerous examples before. New segmentation algorithms for medical imaging need to be optimised for sustaining the challenges of clinical routine. Furthermore, a detailed validation of new algorithms is strongly needed; not just to increase clinicians trust in modern information technology. But also because of the possible serious consequences of automation in medicine. Furthermore, abilities and limitations of new algorithms must be researched, documented and emphasised. This thesis shows how the proposed algorithm was optimised and to what degree resulting target volume definitions may be trusted. Finally, recommendations for its application in radiooncological research and possibly in clinical routine will be given.

2. Fundamentals

2.1. Image analysis

The presented work involves various working fields of image analysis research. To introduce fundamentals of this field, the image analysis pipeline is introduced in this section. Afterwards, the different processing steps of the image analysis pipeline will be explained in detail in the context of PET/CT imaging. Another focus of this chapter are methods for contour comparison. Several methods will be presented as well as a systematisation of these. Finally, an introduction to the research field of swarm intelligence will be given, because the proposed algorithm originates in this field.

2.1.1. Image analysis pipeline

In the field of computer graphics, operations are divided into two subspecialties. On the one side, the transfer of objects acquired by imaging techniques are analysed to derive abstract object information. This subspecialty is called image analysis. On the other side, the field of image synthesis describes the transfer of abstract objects into visualisations. Thus, for every step of image analysis, there exists a counterpart in the field of image synthesis. For example a digital camera may be the counterpart of a computer screen. The circle of information processing from image acquisition through pattern recognition to visualisation is usually divided into two so called pipelines. The first is the image analysis pipeline, which transforms real world objects to virtual information and is introduced in this section in detail. The second is the image synthesis pipeline which transforms virtual information into visualised objects.

The image analysis pipeline describes the different shapes of an image and the needed transformation processing steps from its origin to the representation in electronic form. Figure 2.1 shows the processing steps of the pipeline. The origin of the image is related to physical effects, for example attenuation of radiation in tissue. Those effects are utilised in technical devices to acquire an image. The acquired image does not essentially has to be a visually interpretable form like two-dimensional and coloured. Such a form is usually created in a further step, called image reconstruction. A reconstructed image may then be further manipulated, for example for the purpose of contrast enhancement, error correction and noise reduction. Alternatively, parametric images or maps can be generated to describe physical aspects measured in the imaged object. Afterwards image processing and pattern recognition techniques are applied to extract abstract information from the image such as contours, meshes, landmarks, vectors

or textures. From these data objects and the corresponding image, statistical image analysis methods deliver information, such as average values, maxima and area or volume measurements.

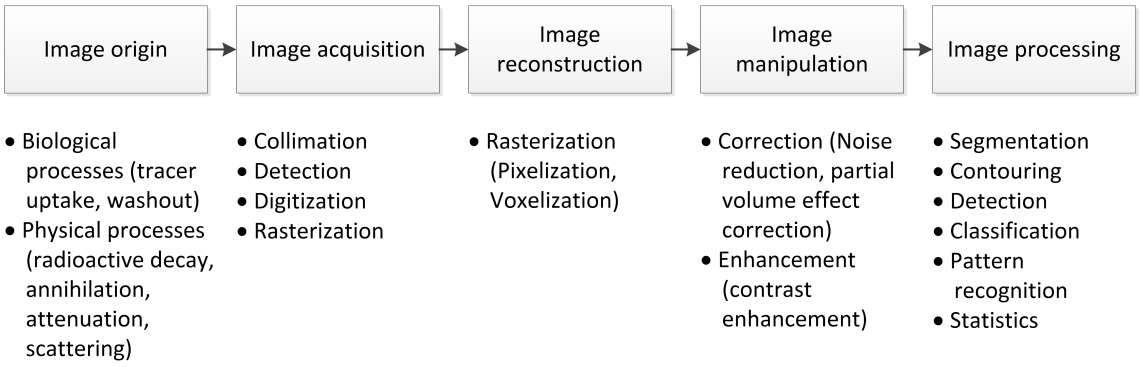


Figure 2.1.: The image processing pipeline connects a number of processes from acquisition of the image to the final derival of information from it.

2.1.2. Image origin

Before the actual process of taking an image can start, physical processes of the real world are needed to create some kind of measurable effect that later forms the image. In classical photography, light beams illuminate the object of interest. Light that is reflected from the object to the camera is used to expose a film. Afterwards the film is further processed to create an image of the photographed object.

In medicine, there are many physical processes utilised for imaging of morphology and functionality of biological subjects. Ultrasound is used to visualise tissues and organs which are easy to reach from the skin and transfer sound waves into and from depth. X-rays are used to generate body projections in medical radiography or furthermore three dimensional images in computed tomography. Magnetic properties of nuclei are used in Magnetic Resonance Imaging (MRI), for example if differentiation of soft tissues is in question. To keep it simple, only the processes behind PET and CT are introduced. For PET imaging, a radioactive, positron-emitting isotope is bound to a molecule that plays an active role in the metabolism of the patient. The solution of this molecule is called a PET tracer. Radiopharmacists develop new PET tracers and are responsible for the production of established ones. The most commonly used PET tracer in clinical routine is $[^{18}\mathrm{F}]$-fluorodeoxyglucose (FDG). This molecule is a glucose derivative marked with a positron-emitting $[^{18}\mathrm{F}]$flourine atom. Similarly to glucose, FDG is processed by cells for energy production. Thus, if a specific organ has a high energy consumption, more glucose and alternatively FDG is taken into the cells of this organ, where only FDG is trapped. The high FDG-uptake into the organ can be measured using a PET scanner. As a result of higher energy consumption the brain shows higher signal intensity in PET images of the head than for example muscle in the neck. However, before a signal in the scanner can be measured, physical effects are in charge. In PET imaging, radioactive decay, or more precisely beta decay plays an important role. Decay is the stochastic process transforming a single unstable atom nucleus into another nucleus. In the case of beta decay, the former nucleus emits an electron or a positron and thus, changes the charge of the nucleus. This process can be described using the physical property half-life. Assuming there are a number of unstable atoms of the

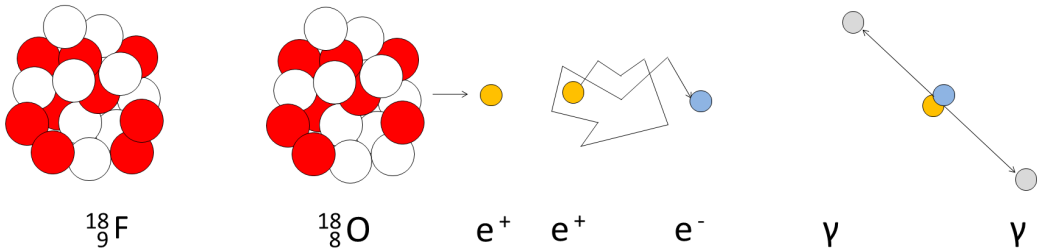
same nuclide, the half-life of the nuclide is the time passing by until half of the nuclei decayed. In PET imaging, as the name suggests, positron-emitting nuclides are utilised for imaging. The emitted positron migrates in the surrounding tissue and causes several further effects. For example, it gets scattered when it collides with other particles. The distance the positron may travel into the surrounding volume is limited by the positron range which furthermore is related to the energy with which the positron is emitted. Table 2.1 gives an overview on different positron ranges for a selection of radionuclides. Finally, the positron and an approaching electron annihilate by emitting two photons in approximately opposite directions. These two photons, which may again be scattered, ideally induce the measurement in the detector ring of the PET scanner. On their way through the body, the photons may furthermore be stopped. The behind lying process is attenuation. The attenuation coefficient of tissue describes the intensity of the photon beam that can be measured after the photons have passed the tissue. The intensity after transmitting the beam through the volume I depends on the emitted intensity I_0, the attenuation coefficient α and the length of the way through the volume x. The association between these parameters is described by the Beer-Lambert law:

$$I = I_0 e^{-\alpha x} \tag{2.1}$$

The whole process from emission of the positron through decay, its annihilation and the emission of the photon pair until the photons leave the body is shown in Figure 2.2. For CT imaging attenuation and scattering play a role as well. The principal differences between PET and CT will be explained in more detail in the next section.

Table 2.1.: Positron ranges in water of selected nuclides taken from (Valk et al., 2004)

Nuclide	half-life $t_{1/2}$/min	**mean range / mm**	**maximum range / mm**
^{11}C	20.4	1.1	4.1
^{15}O	2.0	2.5	7.3
^{18}F	109.8	0.6	2.4
^{68}Ga	68.3	2.9	8.2



(a) radionuclide (b) decay (c) scattering (d) annihilation

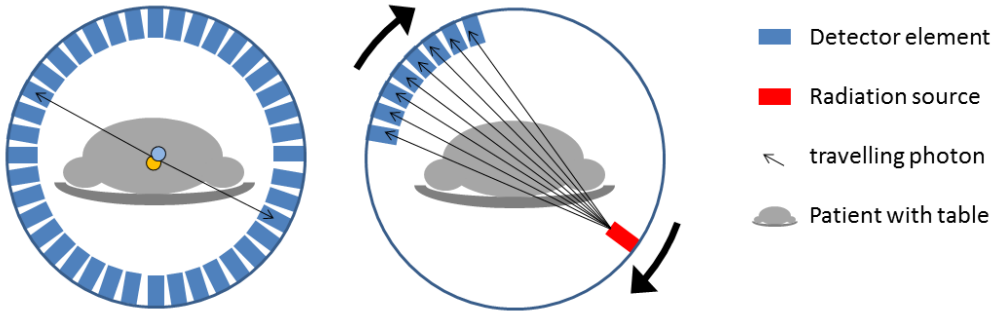
Figure 2.2.: In principle, PET imaging utilizes positrons which are emitted during radioactive decay of so called positron-emitters (a). These positrons travel into the surrounding tissue by some millimetres (b) until they get destroyed during an annihilation process with an electron (c). Two photons are emitted from the location of the annihilation with opposite directions. These pairs of photons may later be measured and the respective signal will be used for image reconstruction.

2.1.3. Image acquisition

In CT imaging, photons are emitted from outside the body and sent through it. The radiation source is mounted within the ring of the CT scanner. On the opposite of the ring, the detector is mounted. Thus, emitted photons which passed the body are measured on the other side of the patient. The resulting data, so called sinograms, can be compared to x-ray images which are created by measuring photons which had been sent through the body. However, radiation source and detector in a CT are mounted on a rotating gantry. That allows acquisition of sinograms from different directions. Thus, collection of sinograms can later be used to reconstruct a slice image of the scanned body region and furthermore three dimensional images. To reduce the impact of scattered photons, a collimator is mounted in front of the detector. The collimator is a physical raster that only allows photons to pass which were sent from the direction of the source. In PET imaging, the detector is most commonly a full ring. Collimators are not incorporated in PET scanners because the position of the source is unknown. The principal design scheme of PET and CT scanners is shown in Figure 2.3. In practice, both scanners are mounted together in one device to eliminate misalignment issues when combined analysis of resulting images shall be performed later. Furthermore, the data format acquired in PET scanners is different from CT scanners. The electronics processing signals from the detector elements of the ring store the time of an event and its position, the detector-number, in a so called list-mode file. List-mode data can later be used to reconstruct a three-dimensional PET image. While image correction is usually performed as a further step in the image analysis pipeline, some errors can be eliminated before saving the list-mode data. For example the electronics of the scanner may drop so called single events. If one detector element measures a photon and no other detector measures another one within a defined time frame, the measured, so called single event, may be dropped, because it would be useless for reconstruction. However, it may also happen that randomly coincidently detector elements measure a signal even though both photons did not appear by an annihilation but by two annihilations at different places, just within a short time frame. These events, so called randoms, are acquired and stored in the list-mode data and may later be corrected.

2.1.4. Image reconstruction

After PET and CT image data have been acquired, reconstructing the data to three dimensional image stacks is the next step. Reconstruction in this case means the rasterisation of measured data to a predefined image matrix. Another word for this process in three-dimensional tomographic imaging is voxelisation, because measured data is mapped to a three dimensional voxel matrix. Choosing the right size of the image matrix is crucial to allow accurate image interpretation after reconstruction (Busk et al., 2008). Therefore the width and height of the image matrix may be calculated from the size of the field of view and the spatial resolution of the scanner. The field of view is predefined by the technician before performing the scan. The spatial resolution r is limited by the used scanner technique and can be measured using phantom experiments (Brambilla et al., 2005). The value of r describes the minimum size of an object that could be detected using the scanner. The Nyquist–Shannon–Kotelnikov sampling theorem (Shannon, 1998) can be adapted to the issue of choosing the right voxel size: It is necessary to

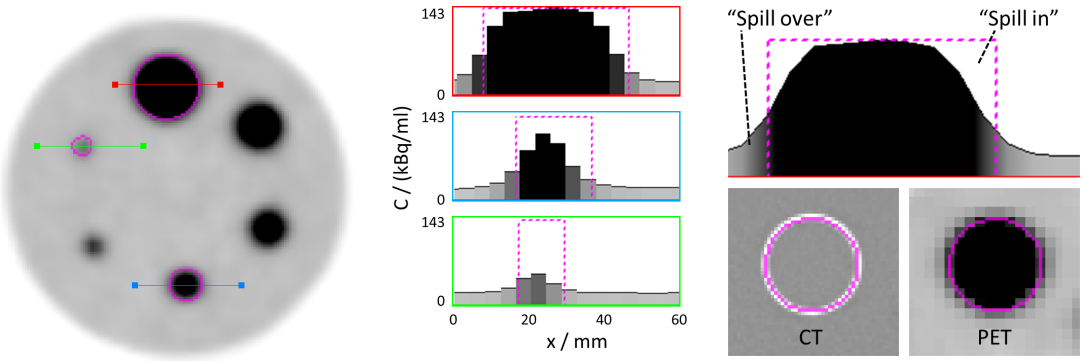


(a) PET scanner (b) CT scanner

Figure 2.3.: PET and CT scanners are both photon-measuring devices for the purpose of imaging. But the layout of both scanners differs: The PET scanner (a) consists of a complete ring of detector elements which measure photons coming from the space inside the ring. The CT scanner (b) consists of a detector mounted in opposite position to a radiation source. Both, detector and source are mounted on a gantry which rotates around the patient. Photons are emitted from the one side to the other side to image radiation density of the patient from different directions. The fact that the direction of the photons is known allows to mount a collimator in front of the detector. This collimator improves image quality by blocking scattered photons.

reconstruct the image stack using a voxel size that is half of the spatial resolution of the scanner or even lower. Otherwise, an object may be not detectable, especially if the contrast is low, because of the partial volume effect (PVE). The PVE is a result of averaging signal intensities within a scanned volume. If there are two objects with differing signal intensities, their signals are measured together. Thus, the determined signal intensity is a weighted average from both subvolumes. Because of the typical spatial resolution of clinical PET scanners, which is about $5\,\mathrm{mm}$, objects smaller than $10\,\mathrm{mm}$ may not be detectable. An example for such objects and a visualisation of the partial volume effect is given in Figure 2.4. Even if such objects are visible in a PET image, the measured signal intensity within the volume may be decreased compared to larger objects where the actual activity concentration is equal. The signal distorting effects observed in PET images on the boundary of an object are sometimes also designated to PVE. However, if activity concentration within an object with theoretically homogeneous activity distribution is different from the activity distant from its boundary, signal from outside the object has 'spilled in'. The terms 'spill-over' and 'spill-out' are used, if activity outside the volume is changed, because of the signal intensity within the volume (Hofheinz et al., 2012b). In fact spill-in and spill-out describe the same effect but from different perspectives. Even though these effects may lead to decreased signal within a target volume, structures may still be visible. For example, if a small lesion takes up a huge amount of tracer, it may still be visible in the PET image, even though it could be smaller than the resolution of the PET scanner. In such a case, the measured uptake is decreased to the actual uptake. These lesions should not be considered for statistical analysis of signal values unless the uptake is PVE-corrected. Such a correction is still possible if volumes are very small, for example $< 5\,\mathrm{ml}$ (Hofheinz et al., 2012b).

For both modalities, CT and PET, several approaches for reconstruction are available and developing more accurate algorithms is still content of current research. For example, introduction of time-of-flight-PET, a technique which allows using the time difference of signal events in detector elements for a more



(a) PET Phantom measurement (b) Partial volume effect (c) Spill in and spill over effects

Figure 2.4.: Partial volume effect and spill in/spill over effects can be shown using a simple phantom experiment (a): Glass spheres (pink) are filled with a solution containing a predefined activity concentration. These spheres are put in a 'background' solution with a different activity concentration. Spheres and background are scanned in a PET scanner. In the resulting imaging data, it is obvious that a sphere must be larger than a defined size so that the maximum activity within the volume is actually measured: All spheres contain a solution with the same activity concentration. But if the spheres are too small (green profile), the partial volume effect leads to decreased measurement of maximum activity (b). The spill-in effect lead to imaging data showing decreased activity concentration measured within the sphere near its glass wall, because decreased activity concentration outside the sphere 'spills into' the sphere. Outside the glass wall, increased activity concentration is measured, even though it is not actually present there. Activity concentration is measured outside, because it is 'spilled over' of the sphere.

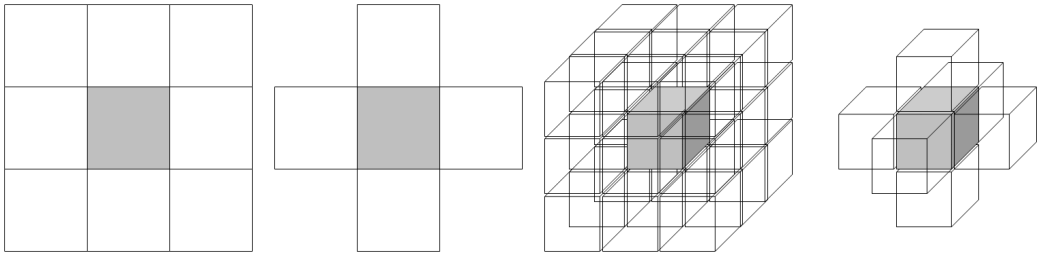
precise location estimation of the annihilation, is still a quite new concept used in clinical routine (Yasuda et al., 2012).

PET reconstruction further allows so called dynamic reconstruction of list-mode data. If the patient was scanned for a sufficient amount of time at one position, a time resolved and thus four-dimensional reconstruction can be applied. By splitting the list-mode data into several frames with a shorter frame-duration time, it is possible to reconstruct several three dimensional image stacks out of one scan. These reconstructed images may then further be processed and allow interpretation of tissue dynamics (Thorwarth et al., 2005). However, in clinical practice reconstruction is usually performed in three dimensions only, because of the higher effort and complexity of four-dimensional reconstruction and image interpretation. Furthermore, dynamic PET measurements deserve more time for image acquisition and may thus be not practical for clinical routine.

2.1.5. Image manipulation

Reconstructed images or image stacks may be manipulated to change image properties such as contrast or brightness. More sophisticated algorithms are available for partial volume effect correction (Hofheinz et al., 2012b) or background subtraction (Nehmeh et al., 2009). In principle, all image manipulation techniques have one property in common: input and output are images. In general, image manipulation always implies that one image is transformed to another image.

For many of these techniques, information from neighbouring voxels is combined in one voxel. Thus, voxel neighbourhoods are introduced. In two-dimensional images, a square drawn around a pixel including its neighbouring pixels is known as the Moore-neighbourhood. It contains all 8 straight or diagonal connected pixels. The so called von-Neumann-neighbourhood contains only the four pixels that have a joint edge with the pixel in question. Both neighbourhoods are visualised for the two- and three-dimensional case in Figure 2.5



(a) 2d Moore-neighbourhood ($n = 8$) (b) 2d von-Neumann-neighbourhood ($n = 4$) (c) 3d Moore-neighbourhood ($n = 26$) (d) 3d von-Neumann-neighbourhood ($n = 6$)

Figure 2.5.: Visualisation of pixel neighbourhoods in two-dimensional (2d) images (a,b) and voxel neighbourhoods in three dimensional (3d) images (c,d). The number of corresponding neighbouring voxels is given as n.

2.1.6. Image processing

The term 'image processing' is used synonymously for two fields of electronic data processing: a) the whole image analysis pipeline with all underlying processes, and b) for the process of transferring rasterised image data to non-image data, such as vectors, contours or other abstract objects. In the following, the term will be used for the latter purpose. Furthermore, some words used in daily practice of image analysis are introduced:

Definition 1 (Segmentation)**.** *The process of dividing a set of image elements I into two disjunct subsets $I_p \subseteq I$ and $I_n \subseteq I$ with $I_p \cup I_n = \emptyset$ and $I_p \cap I_n = I$ is defined as segmentation. The subsets I_p and I_n are defined as positive segment and negative segment, respectively.*

In image analysis practice the first subset I_p represents all voxels which fulfil one or more conditions, the second subset $I_n = I \setminus I_p$ represents the remaining voxels which do not. One of the most commonly used condition is a threshold on voxel properties such as grey values.

Definition 2 (Thresholding)**.** *The process of segmenting a set I of voxels $v \in I$ into two subsets by allocating all voxels with grey value g_v above or equal to a threshold $g_{\min}$ to I_p and the remaining voxels to I_n is defined as thresholding.*

$$\forall v \in I_p : g_v \geq g_{\min} \tag{2.2}$$

Definition 3 (Delineation)**.** *The process of extracting the spatial boundary between segments is called delineation.*

Synonymously, the term contouring may be used. The result of delineation processes are contours or meshes.

Definition 4 (Contour)**.** *A contour is defined for two-dimensional images as a polygone enclosing a segment of pixels.*

Meshes, three dimensional representations of object boundaries, are uncommon in medical imaging due to the higher computational effort needed to process them. In practice, slice-by-slice delineation of three-dimensional segments is performed resulting in a collection of contours describing the segment. If the delineation was performed on thresholded segments, the result can be called iso-contour. This term is derived from topographic contours, which express that the enclosed area lies above the foreclosed area.

After an image or a set of voxels has been segmented, properties of the resulting subsets can be used to classify the image.

Definition 5 (Classification)**.** *The process of allocating a set of voxels to a class by interpreting properties of the set is called classification.*

A class can be seen as a category of sets which fulfil certain conditions. The distinction from image segments is the fact that classes usually contain segments from different images. In applied medical image processing for oncology, conditions for classification can be for example a minimum volume or size of a segment to build up 'large'- and 'small'-tumour-classes.

2.2. Image quality estimation

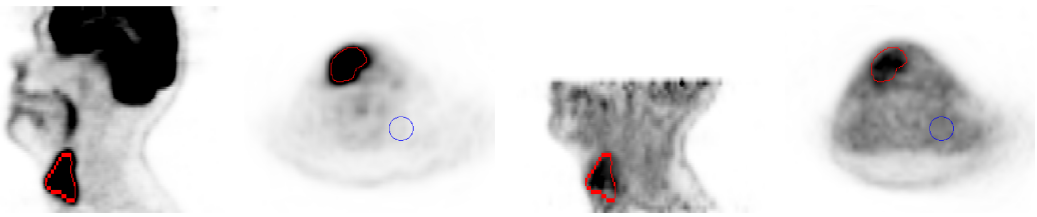
To understand why image segmentation for FMISO PET is more complicated than for example for FDG PET, some measurements of image quality are introduced. Delineation of a target object deserves images where target and background can be distinguished from each other; the contrast between both must be high. For the purpose of contrast estimation, two definitions of contrast are introduced which may be used in PET image analysis (Abolmaali et al., 2012) : The target-to-background ratio TBR and the contrast-to-noise ratio CNR. The TBR can be calculated by dividing activity concentration C within the target by the concentration in a defined background volume:

$$TBR = \frac{C_{\text{target}}}{C_{\text{background}}} \tag{2.3}$$

If the background volume is defined in a specific tissue, such as neck muscle, the abbreviation TMR for target-to-muscle ratio may be used. It's definition is identical with equation (2.3). While TBR measures the ratio between signal intensities of two volumes, CNR allows to estimate how far the activity concentration within target object is above the noise σ in the background volume:

$$CNR = \frac{C_{\text{target}} - C_{\text{background}}}{\sigma_{\text{background}}} \tag{2.4}$$

In the literature, various definitions of contrast are published, which often differ in definition of target and background. In FMISO PET analysis for example, two different definitions of background are commonly used: either activity concentration measured in blood samples is assumed as background (Thorwarth et al., 2005; Nehmeh et al., 2008; Lee et al., 2008) or activity concentration measured in defined volumes within the PET data set (Abolmaali et al., 2012; Zips et al., 2012; Bittner et al., 2013; Chang et al., 2013; Okamoto et al., 2013). In FDG PET image analysis, the standardised uptake value (SUV) is most commonly used as a uptake measurement allowing inter-patient comparison (Thie, 2004). In FMISO PET, TBR estimation is used commonly to numerate tracer uptake in comparison to background activity. Figure 2.6 shows contrast differences between FDG PET and FMISO PET. Furthermore, some measurements use mean activity inside the target for contrast definition, others use the maximum. On the other hand, measurement of maximum activity concentration as observed in a single voxel suffers from noise. Thus, calculating the so called peak activity concentration may be preferred. This is the mean activity concentration of voxels in the neighbourhood the voxel with maximum concentration. However, SUV_{peak} is again dependent on the volume definition being analysed (Vanderhoek et al., 2012).



(a) (b) (c) (d)

Figure 2.6.: For the estimation of contrast between target volume and background, a volume of interest (red) and a background volume (blue) drawn on an FDG PET image (a, b) and an FMISO PET image (c, d) are given. TMR, TMR_{max}, CNR and CNR_{max} in this example were 15.2, 24.2, 78.2 and 127.7 in the FDG PET data set and 1.6, 2.4, 9.3 and 21.2 in the FMISO PET data set, respectively. Thus, in this case independent from the used contrast measurement, between FDG and corresponding FMISO PET data set the contrast is different by about an order of magnitude. Signal intensity measured using the SUV was also higher in FDG PET by about an order of magnitude: SUV and SUV_{max} were 10.4 and 16.6 for FDG and 2.0 and 3.0 for FMISO PET, respectively.

2.3. Segment and contour comparison

When contours are available after segmentation, their comparison is the next step to determine the quality of the contours. The differences between contours from segmentation algorithms and contours representing a given reference standard are investigated to determine the quality of a given algorithm. This section gives an overview of various measures of contour similarity. These so called figures of merit allow measuring the quality of segmentation algorithms as well as studying inter- and intra-observer variabilities. The overview was collected and summarised erlier (Weichert, 2011) and was extended with relationships between the figures of merit.

2.3.1. Definitions

As introduced in definition 1 in section 2.1.6 a segment of a medical image stack is defined as a set A of voxels. The image stack itself can be defined as a set I of voxels as well. The theory of sets may then be applied for the purpose of contour comparison. This allows defining $|A|$ as the number of voxels in segment A. Furthermore, $A \cap B$ is defined as the intersection of two sets A and B. Analogously, the operators $\cup$ for union and $\setminus$ for the subtraction of two sets are defined. The fact that a medical image stack is always bounded, it has a defined size, induces that the negation $\bar{A}$ of set A is also bounded. It contains all voxels $a \in I$ of the image stack that are not member of A:

$$\forall a \in I(a \in \bar{A} \wedge a \notin A) \vee (a \in A \wedge a \notin \bar{A}) \tag{2.5}$$

To apply statistical methods when comparing two sets A and B, four more sets are introduced: Assuming set B is the reference segment and set A is the segment resulting from a given method, the sets TP of truly-positive segmented voxels is defined. Analogously, the set of false-positives FP, true-negatives TN and false-positives FP are defined. By application of the introduced operators and assuming that the whole image I is investigated, these four sets are defined:

$$TP = A \cap B \tag{2.6}$$

$$TN = I \setminus A \setminus B \tag{2.7}$$

$$FP = A \setminus B \tag{2.8}$$

$$FN = B \setminus A \tag{2.9}$$

Figure 2.7 shows how the sets A and B are connected with the sets TP, FP, TN and FN.

2.3.2. Voxel-wise sensitivity and specificity

Assessing the quality of a method used for classification can be done by counting the number of cases where the method classified the samples correctly positive and incorrectly negative. These measurements allow determining the true-positive rate, also called sensitivity, of the classifier. As introduced above, corresponding measurements exist for the number of true positive and false negative voxels for segmentation algorithms. But analysing those does not value a classifier but the performance of a segmentation algorithm. To highlight this difference, the true-positive rate of a segmentation algorithm will

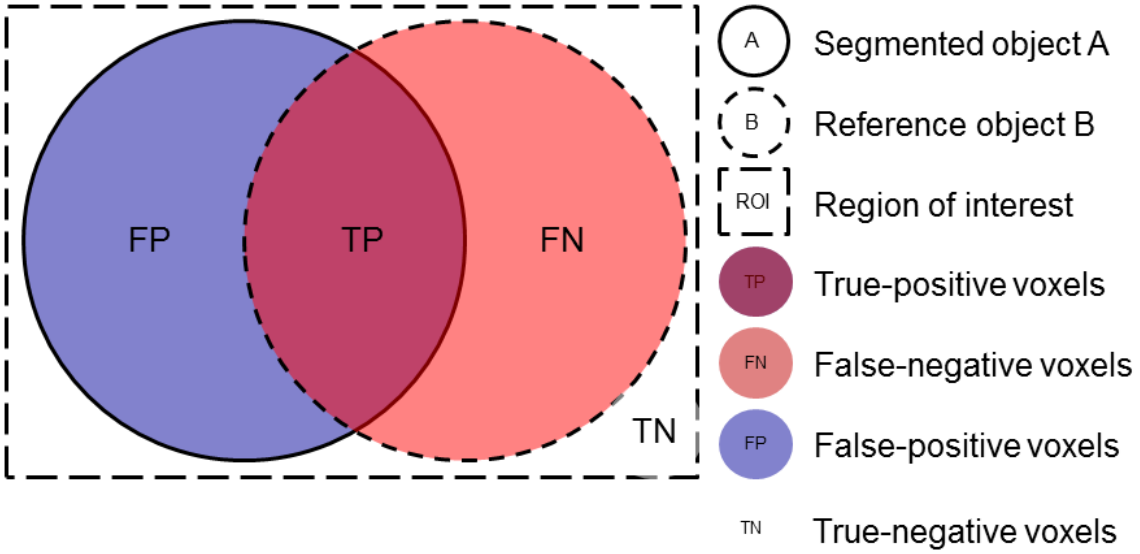


Figure 2.7.: To apply the theory of sets to contour comparison, the voxels enclosed by a contour are defined as a set of voxels. Furthermore, the intersection of a set A and a reference set B can be seen as the set of true-positive voxels TP. Analogously, false-positive FP, false-negative FN and true-negative TN sets can be visualised.

be denoted as voxel-wise sensitivity p_{TP}:

$$p_{TP} = \frac{|TP|}{|TP| + |FN|} \tag{2.10}$$

Analogously the voxel-wise specificity p_{TN}, or false-negative rate, is defined:

$$p_{TN} = \frac{|TN|}{|TN| + |FP|} \tag{2.11}$$

Both rates alone do not measure the quality of a segmentation result meaningfully. Imagine a segmentation algorithm which segments all voxels in I together to the positive set $I_p = I$ and no voxels in the negative set $I_n = \emptyset$. It obviously has a sensitivity of $p_{TP} = 100\,\%$. But its sensitivity is $p_{TN} = 0\,\%$. To value both measurements together, a number of possible combinations were published, such as the Youden index (Youden, 1950):

$$Y = p_{TP} + p_{TN} - 1 \tag{2.12}$$

Alternatively, also the voxel-wise true-rate p_T, also called accuracy, is introduced:

$$p_T = \frac{|TP| + |TN|}{|TP| + |FP| + |TN| + |FN|} \tag{2.13}$$

Accuracy also uses sensitivity and specificity in a combined way to describe the performance of a classificator or a segmentation algorithm. But while the Youden index weights sensitivity and specificity equally, accuracy takes the positive $|TP| + |FN|$ and negative $|TN| + |FP|$ volumes into account. If the target

volume is distinctly smaller than the background, the Youden index is recommended to prevent wrong conclusions taken from a specificity-dominated accuracy measurement.

2.3.3. Receiver operating characteristic

For rating classifiers using sensitivity and specificity measurements in a more visual way, the receiver operating characteristic (ROC) is useful. So called ROC-curves allow visualising the performance of a classifier in a diagram. A free parameter $x_{cut-off}$ used by the classifier serves as variable cut-off value to distinguish between positive samples $x_j \geq x_{\text{cut-off}}$ and negative samples $x_j < x_{cut-off}$. To draw an ROC-curve, a range of cut-off values $x_{\text{cut-off}} \in [\min_i(x), \max_i(x)]$ serve to split the samples into two subgroups. To distribute the data points resulting from applying cut-off-values uniformly to the range, $x_{cut-off,i}$ can be defined as:

$$x_{cut-off,i} = \min_j(x) + \frac{i}{n_i}(\max_j(x) - \min_j(x)) \tag{2.14}$$

$$\text{with } i = 0...n_i \tag{2.15}$$

The number of calculated data points on the curve n_i is chosen freely and should be less than the number of available samples n_j. For every data point, sensitivity p_{TP} and specificity p_{TN} can be calculated as explained in the previous section. The pairs of sensitivity and specificity values are then plotted in a squarish diagram in the form $(p_{TP}, 1 - p_{TN})$. Thus, the x-axis represents the false-positive-rate of the classifier and the y-axis represents the true-positive-rate. Additionally, a straight line from the data point $(0, 0)$ to the data point $(1, 1)$ is drawn to simulate a randomly guessing classifier. If the plot of the visualised classifier lies below this straight line, the classifier performs worse than making random decisions on classification of the samples.

Two classifiers for recurrent tumour detection are shown as example in Figure 2.8 to explain ROC interpretation. In this example, the maximum target-to-background-ratio TBR_{max} of FMISO PET scans of 25 patients is used to classify the patients into two groups: patients with and without recurrent tumours. The underlying data on tumour recurrence is taken from an exploratory study which was published earlier (Zips et al., 2012). It is known, which patient was local-recurrence free and which was not within a certain time. Taking this information and TBR_{max} as the cut-off parameter, allows drawing those curves. Both curves visualise that the classifiers perform better than a random guess. But the second classifier outperforms the first, because

- its area A_{ROC} under the curve is higher,
- there is one data point on the curve with the shorter Euclidean distance to the upper left corner of the plot,
- sensitivity and specificity in this optimal data point is a better compromise than any point on the first curve and
- the corresponding Youden index is maximised in this data point.

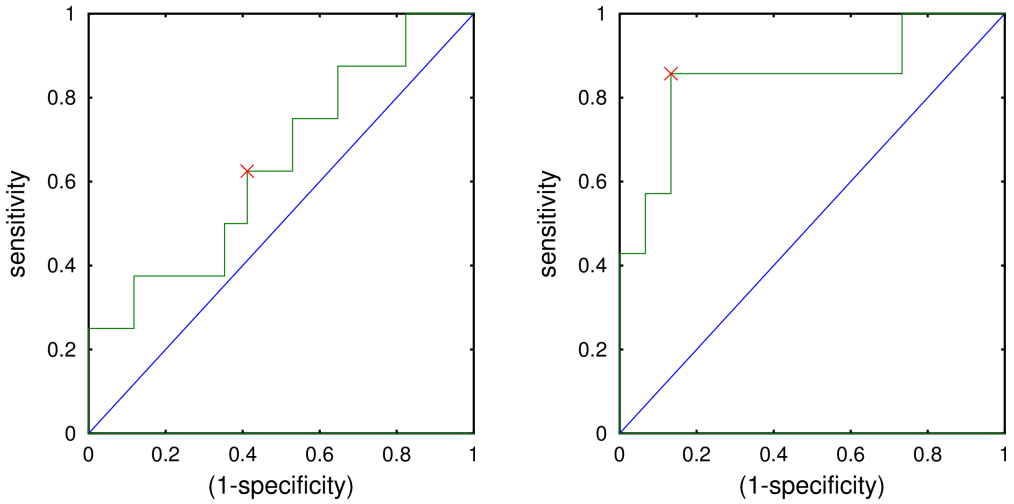


Figure 2.8.: Two example ROC-curves give a visual impression of how well recurrent-tumours can be detected using two classifiers. In a) the TBR_{max} in the proximity of the tumour in an FMISO PET image before therapy start and in b) after two weeks of combined radiochemotherapy. The ROC curve in b) suggests that the latter classifier performs better for the detection of recurrent tumours. This is consistent with findings from multivariate Cox analysis performed in a study on the same data set of patients (Zips et al., 2012). The red cross represents the optimal cut-off value which is defined as the data point with the shortest distance to the point $(0, 1)$ where sensitivity and specificity are optimal.

2.3.4. Voxel-wise receiver operating characteristic

If ROC-curves are applied in the field of image segmentation, they may be used to determine the ability to threshold an image correctly (Haase et al., 2011b; Shepherd et al., 2012). The term voxel-wise receiver operating characteristic vwROC is then used. In such a curve, a number of voxel-wise sensitivity measurements are plotted against corresponding specificity measurements. The free parameter needed to draw the curve, is the threshold which is applied to the image to segment it in two segments. Figure 2.9 shows example vwROC curves of two PET images segmented and compared to a volume definition on CT as reference standard. The vwROC curve gives a visual impression of any threshold to segment an image correctly. Thus, in this context vwROC is an image property, not a property of a classifier.

2.3.5. Jaccard and Dice index

For measuring the degree of overlap of two segments A and B, the Jaccard index J and Dice index D (Dice, 1945) have been established in medical image processing (Han et al., 2011; Dewalle-Vignion et al., 2012; Xia et al., 2012). Both figures of merit deliver values which are always in the range of $[0, 1]$. Zero expresses that both segments do not overlap and 1 expresses that both segments are identical:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2.16}$$

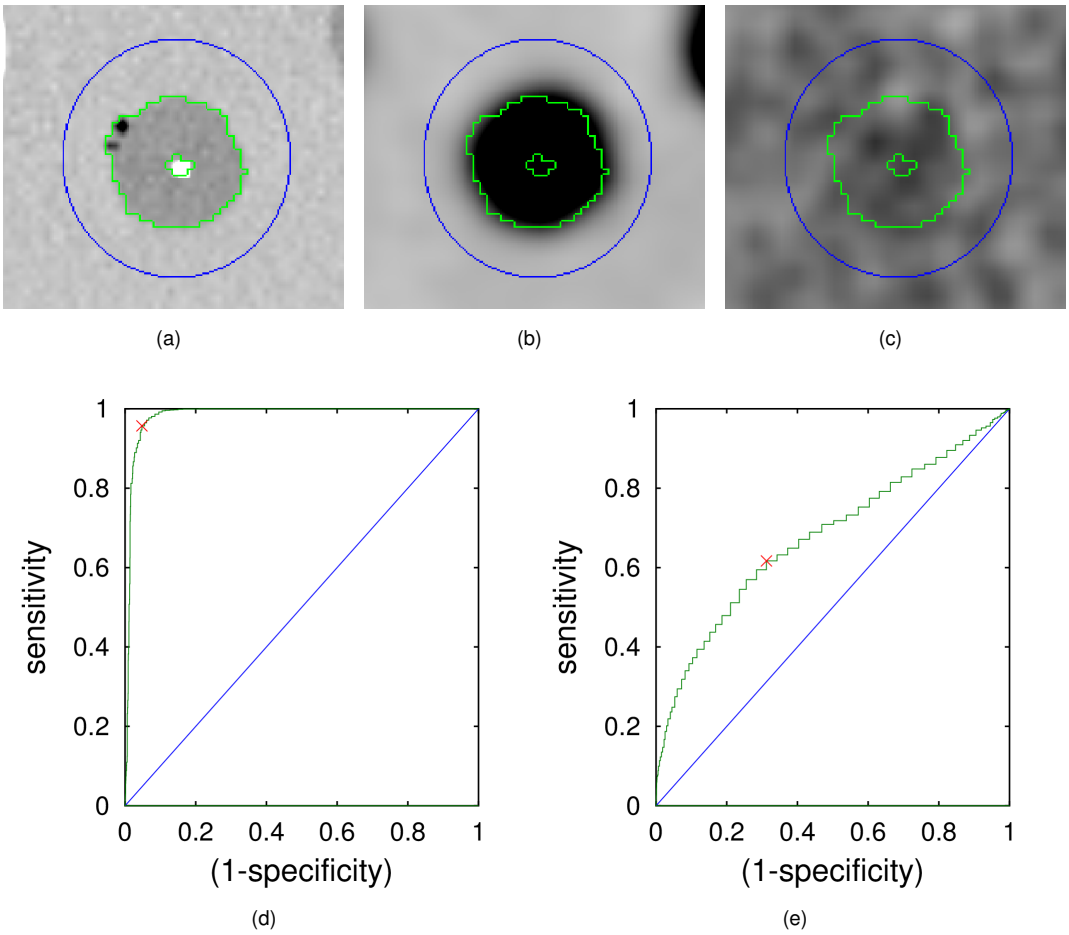


Figure 2.9.: Using a phantom experiment allows determining the ability of threshold based segmentation to segment test objects correctly (Haase et al., 2011b). In a) a CT image is shown where the boundary (green) between target volume and background can be derived from. The blue line encircles the inspected volume. In b) and c) two corresponding PET images are shown with different contrast levels. The target volume is visible in b) and the corresponding vwROC-curve in d) suggests that there is a threshold allowing to segment it almost correctly. In c) the target volume is not clearly visible and the corresponding vwROC-curve in e) suggests that there is a threshold allowing to segment it, but with limited voxel-wise sensitivity and specificity.

$$D(A,B) = \frac{2\,|A \cap B|}{|A| + |B|} \tag{2.17}$$

The Dice index is also known as the Sørensen index (Sørensen, 1948).

2.3.6. Volume and classification error

For the comparison of segmentation results in PET imaging, relative and absolute volume errors are also established methods (Schaefer et al., 2008; Tylski et al., 2010; Li et al., 2012; Cheebsumon et al., 2011b; Hofheinz et al., 2012c).

Assuming $|A|$ is the number of voxels in a segmented object A and V_{voxel} is the volume of one voxel,

allows to determine the volume V_A of a segmented object A:

$$V_A = V_{\text{voxel}}|A| \tag{2.18}$$

If volumes of segmented objects can be calculated, it is possible to determine volume differences. Assuming B is the reference object definition allows to determine the absolute volume error ΔV and the relative volume error δV of segmentation object A:

$$\Delta V = |V_A - V_B| \tag{2.19}$$

$$\delta V = \frac{\Delta V}{V_B} \tag{2.20}$$

Alternatively to volumetric error analysis the number of wrongly classified voxels may be counted using the Hamming distance d_H (Hamming, 1950):

$$d_H = |A \cup B| - |A \cap B| = |FP| + |FN| \tag{2.21}$$

Furthermore, the classification error ϵ_{Class} is the relative pendant to the Hamming distance:

$$\epsilon_{\text{Class}} = \frac{d_H}{|TP| + |FN|} = \frac{d_H}{|B|} \tag{2.22}$$

If two homologous objects should be compared, and none of them can be seen as reference, the relative volume difference $\delta\bar{V}$ is defined:

$$\delta\bar{V} = 2\frac{|V_A - V_B|}{|V_A + V_B|} \tag{2.23}$$

2.3.7. Contour distance

The contour distance of two objects A and B is theoretically defined as the mean shortest distance from every point on the contour of object A to any point on contour of object B. The contour of a segmented object could be seen as the infinitisimal small area between positively and negatively classified voxels. Furthermore an infinte number of points lies on that contour. Thus, for simplification of the contour distance metric, the contour set $K(A)$ of segment A is defined as the set of voxels $v \in A$ that have minimum one voxel $v_n \in N(v)$ in their von-Neumann neighbourhood N that fulfills $v_n \notin A$. Let $d_e(a, b)$ be the Euclidean distance between two voxels a and b. Then, the shortest Euclidean distance $d_{\text{e,min}}(a, B)$ between a voxel a and a set of voxels B is defined:

$$d_{\text{e,min}}(a, B) = \min(d_e(a, b)|b \in B) \tag{2.24}$$

Thus, the mean contour distance $\bar{d}_c(A, B)$ of two voxel sets A to B can be defined as:

$$\bar{d}_c(A, B) = \frac{\sum_{\forall a \in K(A)} d_{e,min}(a, K(B))}{|K(A)|} \tag{2.25}$$

This metric is not commutative. Thus the bilateral mean contour distance $\bar{d}_{\text{bil,c}}$ is defined, which is commutative:

$$\bar{d}_{\text{bil,c}}(A, B) = \frac{\bar{d}_c(A, B) + \bar{d}_c(B, A)}{2} \tag{2.26}$$

Alternatively, a weighted bilateral mean contour distance is defined as

$$\bar{d}_{\text{wei,bil,c}}(A, B) = \frac{|K(A)|\bar{d}_c(A, B) + |K(B)|\bar{d}_c(B, A)}{|K(A)| + |K(B)|} \tag{2.27}$$

Furthermore, the Hausdorff distance (Hausdorff, 1914) is defined as the maximum of the shortest distance of any voxel in A to any voxel in B:

$$d_h(A, B) = \max(d_{e,min}(a, B)|a \in A) \tag{2.28}$$

Again, this metric is not commutative and the corresponding commutative metric $d_{bil,h}$ is defined analogously to the bilateral mean contour distance:

$$d_{bil,h}(A, B) = \max(d_h(A, B), d_h(B, A)) \tag{2.29}$$

In the case of comparing sets of voxels using the Hausdorff distance, it can be assumed that the sets are closed sets. Thus for a more efficient computation of the metric it is not needed to find the maximum shortest distance of all voxels $a \in A$ to the set B. The maximum shortest distance between the sets will be found between voxels on the contour of both sets. Thus, the computational simplification of formula (2.28) is:

$$d_h(A, B) = \max(d_{e,min}(a, K(B))|a \in K(A)) \tag{2.30}$$

2.3.8. Centre of mass and centre of gravity distance

For the determination of volume shifts in space, two characteristic points of a segment in an image are introduced. The centre of geometry x_g and the centre of mass x_m. Both describe the position of the segment in space. For their determination, the spatial positions x_v and corresponding grey values g_v of all voxels v in a volume of interest V must be given.

$$x(V) = \frac{\sum_{v \in V} x_v g_v^i}{\sum_{v \in V} g_v^i} \tag{2.31}$$

The exponent i allows to weight grey values in this equation. If $i = 0$, grey values are not taken into account and $x = x_g$ describes the centre of geometry. If $i = 1$, $x = x_m$ the centre of mass is calculated.

Assuming there are two segments A and B allows determination of the geometry shift d_g:

$$d_g(A, B) = x_g(A) - x_g(B) \tag{2.32}$$

The shift of the centre of mass d_m is defined analogously and may be used to determine the shift of FMISO PET positive volumes during therapy (Abolmaali et al., 2010).

2.3.9. Systematisation

The large number of available methods for contour respectively segment comparison suggests that there are relationships between them. For example, equations 2.16 for the Jaccard- and 2.17 for the Dice index are related, as it will be shown in the following. The Dice index can be calculated in an alternative way by reformulating equation (2.16) using equations (2.6), (2.8) and (2.9):

$$D = \frac{2|TP|}{2|TP| + |FP| + |FN|} \tag{2.33}$$

Analogously, the Jaccard index can be formulated as:

$$J = \frac{|TP|}{|TP| + |FP| + |FN|} \tag{2.34}$$

Furthermore, expansion of this equation leads to:

$$J = (2|TP|)\left(2(|TP| + |FP| + |FN|)\right)^{-1} \tag{2.35}$$

$$= \left(\frac{2|TP|}{2|TP| + |FP| + |FN|}\right)\left(\frac{2(|TP| + |FP| + |FN|)}{2|TP| + |FP| + |FN|}\right)^{-1} \tag{2.36}$$

$$= \left(\frac{2|TP|}{2|TP| + |FP| + |FN|}\right)\left(\frac{2(2|TP| + |FP| + |FN|) - 2|TP|}{2|TP| + |FP| + |FN|}\right)^{-1} \tag{2.37}$$

$$= \left(\frac{2|TP|}{2|TP| + |FP| + |FN|}\right)\left(2 - \frac{2|TP|}{2|TP| + |FP| + |FN|}\right)^{-1} \tag{2.38}$$

Or, in short:

$$J = \frac{D}{2 - D} \tag{2.39}$$

Thus, Dice and Jaccard index are exchangeable. The Dice index is always higher than or equal to the Jaccard index. This fact is expressed in Figure 2.10.

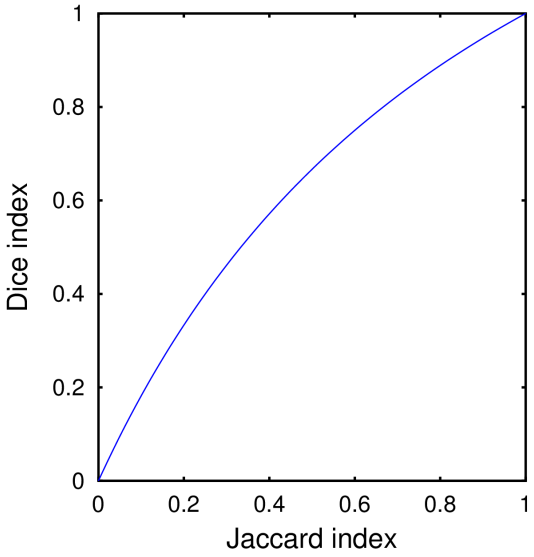


Figure 2.10.: Plotting the Jaccard index and the Dice index against each other reveals their relationship. Both are defined for the range $[0;1]$ and thus, the Dice index is always higher than or equal to the Jaccard index.

By using similar elementary transformations, other relationships with the Jaccard index can be shown, such as with sensitivity p_{TP}, the classification error ϵ_{Class}, the relative volume error δV and the Hamming distance d_H:

$$J = \frac{p_{TP}}{\epsilon_{\text{Class}} + p_{TP}} \tag{2.40}$$

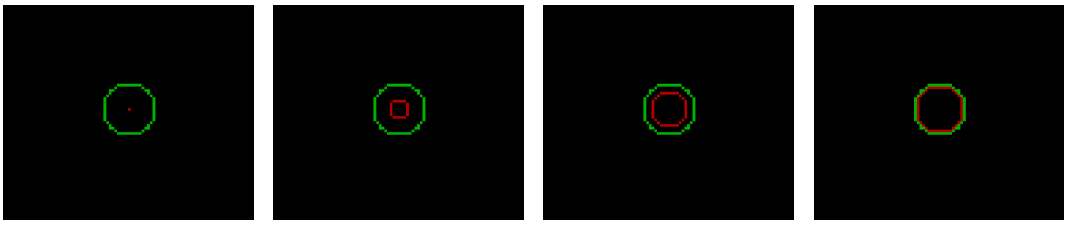
$$J = \frac{p_{TP}}{\frac{1}{1-\delta V} + 1 - p_{TP}} \tag{2.41}$$

$$J = \frac{|TP|}{d_H + |TP|} \tag{2.42}$$

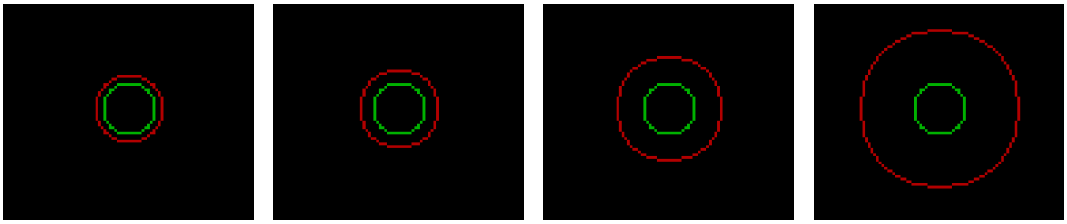
Depending on which figure of merit is used, volume overestimation and underestimation may result in different figure values. Thus, the choice of figure may be decisive depending on the purpose of the contour comparison. To visualise this, a reference spheroid contour was defined on a $1\,\text{mm} \times 1\,\text{mm} \times 1\,\text{mm}$ voxel grid and compared to a number of spheres which were smaller and larger. Some of these pairs of contours are shown in Figure 2.11. Resulting figure of merit measurements of Jaccard index, Dice index, Contour distance, volume error and relative volume error are given in Figure 2.12.

The test spheres were created by erosion and dilatation of the reference sphere and thus, there is a linear relationship between the bilateral mean contour distance and the size of the margin which was eroded or dilated. The difference between bilateral contour distance and its weighted pendant appears not relevant if the volumes are almost equal. From this point of view none of both measures delivers an advantage compared to the other. Furthermore, both contour distance measures have in common that the measurements handle over- and underestimation by a specific margin size similarily: for example eroding or dilating a volume by $5\,\text{mm}$ results in a contour distance of about $5\,\text{mm}$. However, The Jaccard

index, Dice index, volume error as well as relative volume error result in different measurements for erosion and dilatation. Thus, if contour accuracy is in focus of investigation, a contour distance metric is recommended to numerate contour differences. If accuracy of volume measurements is the aim of a study, volume error and absolute volume error shall be favorised for contour comparison.



(a) Erosion by 9 mm (b) by 6 mm (c) by 3 mm (d) by 1 mm



(e) Dilatation by 3 mm (f) by 5 mm (g) by 10 mm (h) by 20 mm

Figure 2.11.: A test sphere with 10 mm radius in a 1 mm $\times$ 1 mm $\times$ 1 mm voxel grid (green) has beed eroded (a-d) and dilated (e-h) by differently sized margins resulting in eroded and dilated spheres (red).

2.3.10. Limitations

Depending on which figure of merit was chosen, some limitations must be taken into account. Assume two identical segments, shown in Figure 2.13, which are not in the same position. The absolute volume error of both is $\Delta V = 0$ even though displacement may be obvious by visual inspection. Thus, the volume error is blind to contour displacement. Similar limitations can be shown for other figures of merit. For example it appears obvious that adding a margin to an object should not change its centre of mass. This may indeed be true for ellipsoidal objects. However, when comparing typical tumour volume definitions as they appear in real patient data sets with and without added margins, it may be observed that the centre of mass, or the centre of geometric as well, shift in space. Furthermore, the contour distance cannot be used to determine the size of the margin that was added. These limitations are also visualised in Figure 2.13.

2.3.11. Summary

In this section, figures of merit were introduced which allow determination of contour and segment differences. All of these figures of merit may be used to determine quality of segmentation algorithms as well. Depending on the purpose of the contour comparison, the right figures of merit must be chosen. In many cases it is recommended to use several figures of merit in comination to visualise contour differences

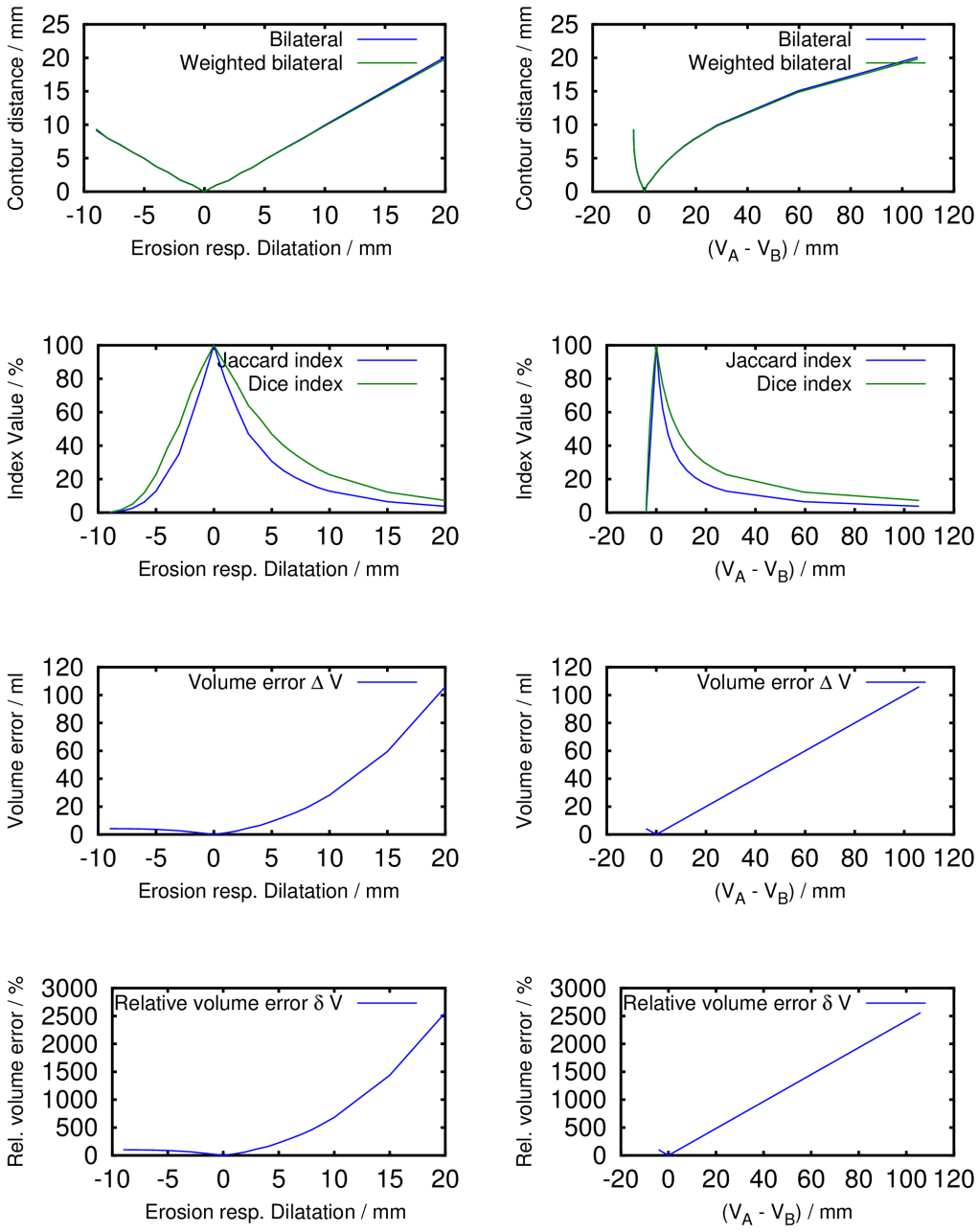
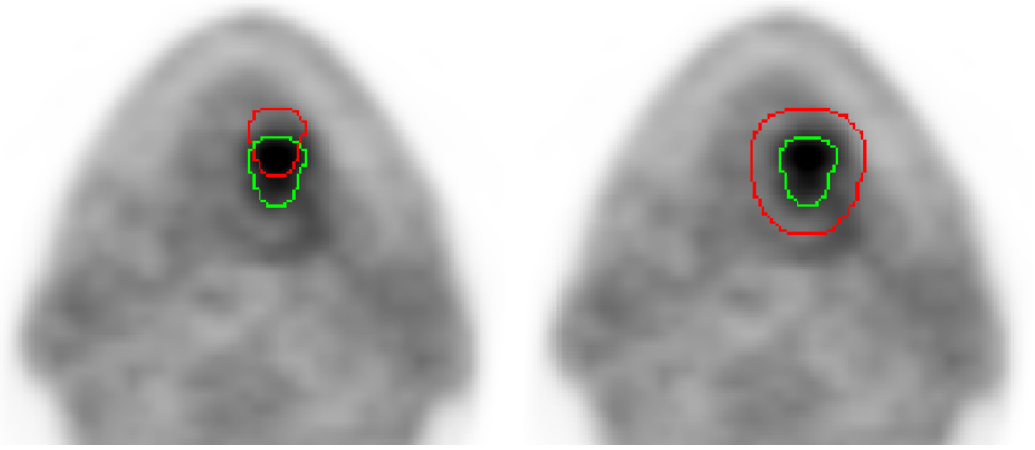


Figure 2.12.: The test spheres have been compared to the reference sphere using contour distance (top row), Jaccard index and Dice index (second row), volume error (third row) and relative volume error (bottom row). If the size of the eroded or dilated margin is given on the x-axis, the contour distance shows a linear relationship with the size of the margin. If the volume change $V_A - V_B$ is plotted on the x-axis, the volume error and the relative volume error show a linear relationship. In all the other cases, non-linear relationships are observed. Furthermore, volume over- and underestimation are numerated differently.

from more than one perspective. But it was also shown that some of them, such as Jaccard and Dice index do not deliver more insights when used together. As a general initial guess, contour comparison may always start with the Jaccard index as a relative measure of volume difference. For the purpose of



(a) Displacement (b) Dilatation

Figure 2.13.: Figures of merit for contour comparison are of limited value for answering specific questions. For example, the volume error cannot measure displacement: in a) a contour (green) was shifted in cranial direction by about $10\,\text{mm}$ resulting in a volume error $\Delta V = 0$. The centre of geometry distance $d_g = 10.9\,\text{mm}$ in this case allows estimating the shift distance. In b) the volume is increased using a margin of $10\,\text{mm}$ resulting in a centrer of geometry distance of $d_g = 1.2\,\text{mm}$. In this case, the volume error $\Delta V = 41\,\text{ml}$ is the better indicator for the volume differences.

comparisons of volume definitions which may play a role in radiotherapy treatment planning, the bilateral mean contour distance should furthermore be evaluated, even though its determination is computationally costly. Depending on the question to be answered, other measurements may be added, such as centre of geometry distance for displacement analysis or volume error of target structures to determine target volume shrinkage or growth indicating therapy progress.

2.4. Swarm based image segmentation

Within the research field of swarm intelligence, a number of algorithm classes have been established. Most popular in this field are particle swarm optimisation (PSO) and ant colony optimisation (ACO). ACO was introduced about two decades ago as a new algorithm for solving the traveling salesman problem (Colorni et al., 1992). Since then, ant-based algorithms for many theoretical combinatorial as well as real-world problems have been published (Dorigo, 2001). Because several research groups found promising results when working with artificial swarm intelligent systems, development of the research field was pushed further to build systems which were able to handle glitches and changing environments (Bonabeau et al., 2000). Analysis and segmentation of low contrast PET image data was seen as a possible working field for this group of robust algorithms and thus, a segmentation algorithm for PET imaging data has been developed (Haase, 2010). This section contains an introduction to ant-based algorithms in general and furthermore their adaption to image processing as it appears in the literature. The documentation of the actual implementation of the proposed ant-based segmentation algorithm can be found in the next chapter.

2.4.1. Ant algorithms

The class of ant-based algorithms is, as its name already suggests, adapted from principles observed in real ant colonies. Hölldobler and Wilson published a comprehensive overview on natural ants and their behaviour (Hölldobler and Wilson, 2009). Ants are able to find the shortest path, for example between nest and food source. Most commonly, ant-based algorithms are used for optimisation of algorithms solving complex problems which are not solvable in reasonable time using deterministic algorithms. Today, ant-based algorithms are applied to various problems in the field of combinatorics such as vehicle routing, graph colouring and scheduling (Dorigo et al., 2006).

In theoretical research on ant-based algorithms, convergence proves are a hot topic (Dorigo et al., 2006). Still, only for a limited number of algrithms in this field it is proven that an optimal solution to the tackled problem is found within reasonable time. For certain classes of ant-based meta-algorithms proves where published showing that the algorithms converge in at least one optimal solution with a predefined small error, if a sufficient number of iterations have been calculated (Gutjahr, 2000; Stützle and Dorigo, 2002).

Classical ant-based algorithms base on the concepts of transition rules and stigmergy which will be explained in this section. For image processing using virtual ants new paradigms, namely self-reproduction and ant ageing, are introduced as well.

2.4.2. Path-planning by transition rules

The so called double-bridge experiment performed with actual Argentine ants (Goss et al., 1989) inspired computer scientists to simulate the behaviour of social insects in the computer (Bonabeau et al., 2000). This experiment is emphasised in Figure 2.14. The first technical implementation of similar principles appeared about twenty years ago (Colorni et al., 1992). The basic principle in most ant-based algorithms are transition rules as introduced by Colorni et al. which describe the way how ants do path-planning in a graph. Ants decide where to go by interpreting properties of the environment and present pheromone intensities:

$$p_{ij} = \frac{[\tau_j]^\alpha [n_j]^\beta}{\sum_{k/i} [\tau_k]^\alpha [n_k]^\beta} \tag{2.43}$$

The probability p_{ij} of an individual ant moving from node i to node j in a graph depends on pheromone intensities τ and the value of an objective function n on node j and all nodes k that are accessible from node i. Analogously, pheromone intensities and objectives could be formulated for edges in the graph instead of nodes. This transition rule successfully allowed to simulate ants which were able to solve the traveling salesman problem (Colorni et al., 1992).

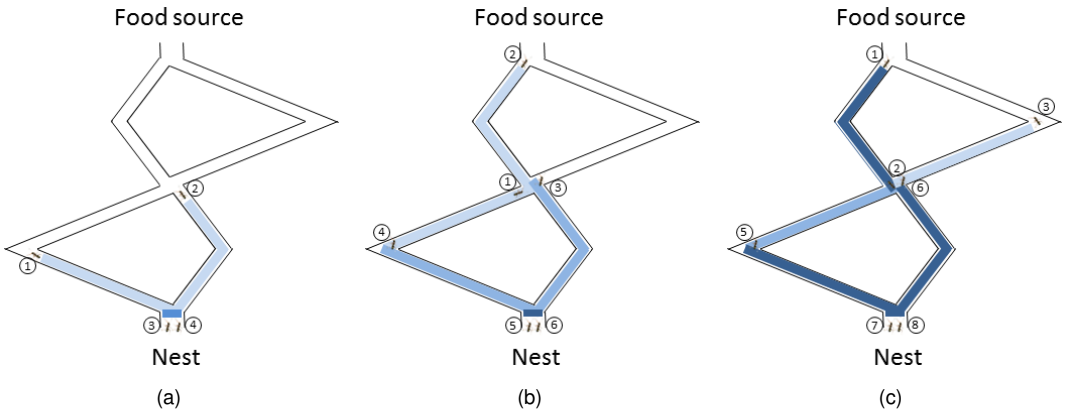


Figure 2.14.: Visualisation of the double-bridge-experiment: Ants find the shortest way from nest to the food source by marking their way with pheromone (blue). An ant that goes the shorter way and returns (for example ant number 2 in figure b) follows its own path because it has the higher pheromone intensity and increases the pheromone intensity on the path(darker blue). With a sufficient number of ants, the colony members will more and more use the shorter paths and further increase emitted pheromone there.

2.4.3. Stigmergy

The ants' transition probability does not only depend on the value of the objective function, which may be related to way length, costs and profit. It also depends on so called pheromone intensities. The perceived pheromone field is built up by the ants on their own. Thus, the ants are able to influence each other by emitting and interpreting pheromone. This nature-derived principle of simultaneous emitting and interpreting pheromone in the environment is called stigmergy. The ants use pheromone to build a cognitive map of their environment (Chialvo and Millonas, 1995). Thus, the pheromone field may lead to the solution of the given problem. By interpreting the resulting pheromone field, a solution of the problem may be retrieved.

2.4.4. Ant-based image segmentation

If ant-based algorithms are able to solve complex problems such as clustering (Handl et al., 2006), ant-based image segmentation is the next step. If the pheromone field is indeed some kind of cognitive map representing the ants environment (Chialvo and Millonas, 1995), in can be interpreted as well. Thus, solutions to the segmentation problem may be more obvious when inspecting the phermone field than when inspecting the processed image. Ant-based algorithms for edge detection gained importance during the last years (Ramos and Almeida, 2000; Fernandes et al., 2005a; Huang et al., 2008; Mullen et al., 2008; Baterina and Oppus, 2010; Jevtić et al., 2011). Using these algorithms pheromone fields are created which appear similar to the results of edge detection filters, like the Sobel-operator. But these algorithms are able to handle changing environments, because of the pheromone field which is adapted during every iteration (Ramos and Almeida, 2000). Furthermore, advanced models were developed which speed up ant-based edge detection by introducing new principles in the algorithm, such as self-reproduction (Fernandes et al., 2005a).

For medical image processing, also a number of ant-based algorithms were published. For example, for segmenting lung structures (Cerello et al., 2010), vessels in iris images (Ma et al., 2009), brain segmentation in magnetic resonance images (Huang et al., 2008; Cao et al., 2008) and detection of micro-calcifications in digital mammograms (Jevtić et al., 2011).

2.4.5. Self-reproduction and ageing

In ant-based image processing, a principle for varying the population size of the colony was introduced (Fernandes et al., 2005b). It consists of two principles called self-reproduction and ageing. Self-reproduction allows increasing the number of ants in the colony. In this probabilistic approach, every ant has a probability of reproduction p_r and a probability to survive p_s. The reproduction probability is calculated from the number of surrounding ants n. The idea is, that an ant which has no neighbours should not be allowed to reproduce, because it is probably located in a worthless place. An ant being surrounded by too many ants should not be able to reproduce as well, because there is no space for descendants and the area is obviously already explored by enough ants. Thus, an ant sitting in a half-filled environment have the highest probability to reproduce. The process of ageing using the survival probability p_s allows the colony to decrease its size. When created, every ant has maximum survival probability $p_s = 1$. With every iteration, this probability is decreased. After every iteration, ants for elimination are chosen probabilistically. Thus, if all ants have a survival probability of $p_s = 1/2$, half of the population would be eliminated randomly.

2.4.6. Summary

In this section a short review of the origin of ant-based algorithms and current research was given. The algorithm is probabilistic, every ant decides partly randomly where to go. It shall be emphasised that a given ant-based algorithm always contains explicit rules defined by its programmer, for example transition rules. But furthermore, implicit rules, which are usually not documented in detail are present as well: An ant that calculates a reproduction probability of 1 which is surrounded by other ants cannot seed a descendant. Rules like these may be present and influencing the system even though they are not implemented explicitly in the source code of the algorithm.

2.5. PET segmentation algorithms

In clinical routine, PET images may be used to create gross-tumour-volume (GTV) definitions, which are usually named GTV_{PET} according to their origin. These GTVs, independently if they were created on PET or CT images, are then dilated by adding a margin $d_{\text{GTV-CTV}}$ to a clinical target volume (CTV). These CTVs are thought to include cancerous cells, which may have invaded surrounding healthy tissue and are not visible in tomographic imaging. Thus, this volume is often extended to adjacent organs, such as lymph nodes which shall be treated as well. The CTVs are also dilated, this time using a margin

$d_{\text{CTV-PTV}}$ to define a planning-target-volume (PTV). The PTV includes the volume to irradiate including the uncertainty induced by errors such as patient positioning or organ motion during therapy. These PTVs are then transferred to radiotherapy treatment planning workstations so that medical physicists can create an individual radiotherapy treatment plan for each specific patient. A scheme showing the relationship between GTV, CTV and PTV is given in Figure 2.15 together with example contours on a CT image. This visualisation is a simplification of the actual process of contouring in clinical routine. For example, the CTV may be extended to include anatomical structures like lymph nodes, which may contain cancer cells. This extension is done one the basis of clinical experience radiooncologists gathered.

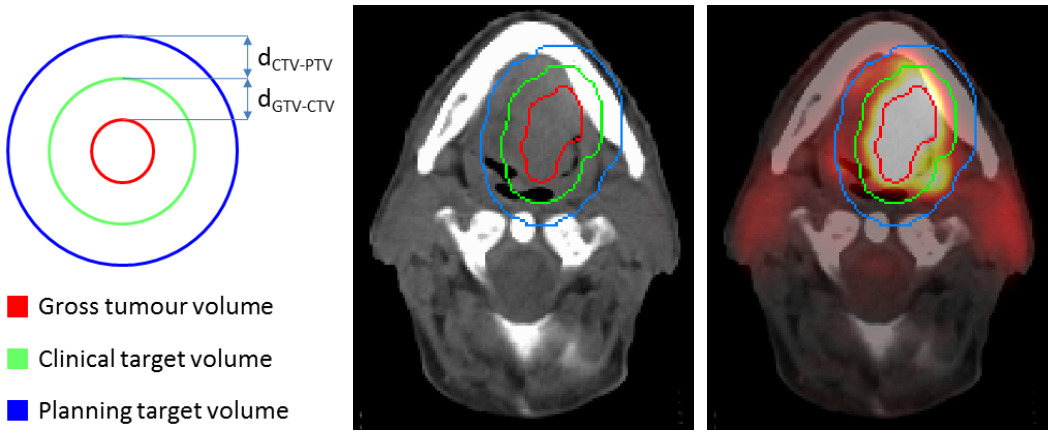


Figure 2.15.: Volume definitions in radiotherapy treatment planning use standardised names. A schematic overview is shown in a). From the gross tumour volume (GTV), that is usually defined by an experienced oncologist, first the clinical target volume (CTV) and afterwards the planning target volume (PTV) are determined by adding margins d given in millimetres to the GTV definition. In b) example volumes are given on a CT image of a head-and-neck-cancer patient. The image fusion of CT and FDG PET given in c) shows how the metabolic volume of the tumour is encircled by the three given volumes.

Even though various algorithms were published for automatic target volume delineation in PET images, in clinical practice target volumes are delineated manually by radiooncologists. Usually these contours are drawn on CT images, but it was already shown that fusion of FDG PET and CT may reveal more consistent delineation between different observers (Caldwell et al., 2001). Furthermore, in lung cancer, PET-CT-fusion for manual volume definition brings advantages if metastatic lymph nodes or atelectasis are present (Steenbakkers et al., 2005).

In FDG PET images alone, the cancer related volume of interest in pretherapeutic patients usually is distinguishable from the background by signal intensity. Thus, most automatic routines use thresholding to separate target volumes and background from each other. The technique of optimal threshold estimation varies from algorithm to algorithm: some iteratively determine a background volume and afterwards estimate a threshold based on mean background activity (Nehmeh et al., 2009; Hofheinz et al., 2012c; Schaefer et al., 2012). Other algorithms use manually defined background volumes for a similar estimation (Schaefer et al., 2008). Furthermore, using a predefined threshold based on standard uptake values (SUV) is a rather simple method which may not be applicable for investigations on different patients (Biehl et al., 2006). But according to data from Nestle et al. such a method may be the most reliable

when delineating large lesions with heterogeneous tracer accumulation (Nestle et al., 2005).

Typically validation of segmentation algorithms is done using phantom experiments and patient data sets. One way to evaluate delineation accuracy is comparing the contours created by an algorithm with manually created contours created on the same images (Nestle et al., 2005; Hofheinz et al., 2012c). Another more long-ranging approach is to compare automatically created contours with those from other imaging modalities (Daisne et al., 2004; Schaefer et al., 2008; Nehmeh et al., 2009). However, the second approach does not only validate the segmentation algorithm, but also the imaging modality itself. Only the first approach allows for comparison of the error of manually created contours with the error of automatically created contours without influence of the chosen modality. In a detailed analysis on patient data Hofheinz et al. showed that automatic contours are in mean more accurate than those from a group of observers, because the standard deviation of the relative volume error of about $13\,\%$ of the algorithm was lower than the mean standard deviation of manually-created contours which was about $27\,\%$ (Hofheinz et al., 2012c). In summary, automatic algorithms are able to imitate manually-created contours to a degree that may be more reliable than human observers, because of eliminated or reduced inter-observer-variability. But the fact that the mean relative volume error is small, for example less than $15\,\%$ (Nestle et al., 2005; Schaefer et al., 2008; Nehmeh et al., 2009; Hofheinz et al., 2012c), does not prove that a segmentation on an individual data set is accurate to any degree. Rather in all algorithms shown in Table 2.2, a substantial number of the segmented lesions had an error of more than $15\,\%$. Nestle et al. assumed that especially in large tumours with an heterogeneous activity distribution large errors may appear (Nestle et al., 2005).

2.5.1. Thresholding relative to background activity

Only one segmentation technique for FMISO PET has been published so far: relative thresholding. In the common approach, a threshold is calculated directly from background activity. This background activity can be measured in blood samples (Thorwarth et al., 2005; Nehmeh et al., 2008; Lee et al., 2008; Swanson et al., 2009). If blood samples are not available, it is straightforward to determine background activity from an appropriate volume of interest (VOI) in the PET data set itself. If the images contain large vessels, such as aorta, blood activity concentration can be measured inside it. If not, the activity concentration within contralateral neck muscle could be used to determine a reliable background activity (Abolmaali et al., 2012; Zips et al., 2012). Measuring the background activity concentration within the ipsilateral neck muscle appears also feasible (Chang et al., 2013). Alternatively, a background volume may be defined in normal tissue contralateral to the primary tumour site (Bittner et al., 2013). Finally, the absolute threshold T_{abs} for target object delineation is calculated by multiplying the mean background activity concentration $C_{background}$ with a predefined factor T_{rel}:

$$T_{\text{abs}} = C_{\text{background}} \cdot T_{\text{rel}} \tag{2.44}$$

The parameter T_{rel} depends on how background activity was determined. Some authors used $T_{\text{rel}} = 1.2$ in combination with background activity estimation from blood samples (Nehmeh et al., 2008; Swanson

Table 2.2.: Published PET segmentation algorithms, errors (Abbreviations: NSCLC: Non-small-cell-lung-cancer, HNC: head-and-neck-cancer)

Authors / publication	**Segmentation principle**	**Patients / lesions**	**Used reference standard**	**Mean relative volume error ± standard deviation**	**Number of data sets with relative volume error $< -15\,\%$ or $> 15\,\%$**
(Daisne et al., 2004)	Signal-to-noise based thresholding (Daisne et al., 2003)	29 lesions from HNC patients	CT-based contours from one observer	$26 \pm 32\,\%$	22 of 29
			MR-based contours from one observer	$27 \pm 33\,\%$	24 of 29
(Hofheinz et al., 2012c)	Iterative Background-determination, thresholding	38 lesions (from 5 liver metastases and 9 HNC patients)	PET-based contours from 9 observers	$3.7 \pm 12.7\,\%$	9 of 37
(Nehmeh et al., 2009)	Iterative Background subtraction, thresholding	6 lesions, 5 patients (4 lung, 1 axilla, 1 iliac bone)	CT-based contours from one radiologist	$10 \pm 14\,\%$	1 of 6
(Nestle et al., 2005)	Constant threshold $SUV = 2.5$	25 NSCLC lesions	PET-based contour from two observers	$-4\,\%$	1 of 8
		8 NSCLC lesions with heterogeneous FDG accumulation		$-2 \pm 23\,\%$	3 of 8
(Schaefer et al., 2008)	Contrast oriented thresholding	8 non-small-cell lung cancer lesions	CT-based	$5 \pm 10\,\%$	1 of 6

et al., 2009). Using background activity determination from a reference volume within the neck muscle, higher relative thresholds like $T_{\text{rel}} = 1.5$ (Chang et al., 2013; Bittner et al., 2013) or $T_{\text{rel}} = 1.6$ (Zips et al., 2012) appear more appropriate. Example contours generated by applying the thresholds $T_{\text{rel}} = \{1.2, 1.4, 1.6, 1.8, 2.0\}$ applying activity concentration measurements from the neck muscle as background are shown in Figure 2.16.

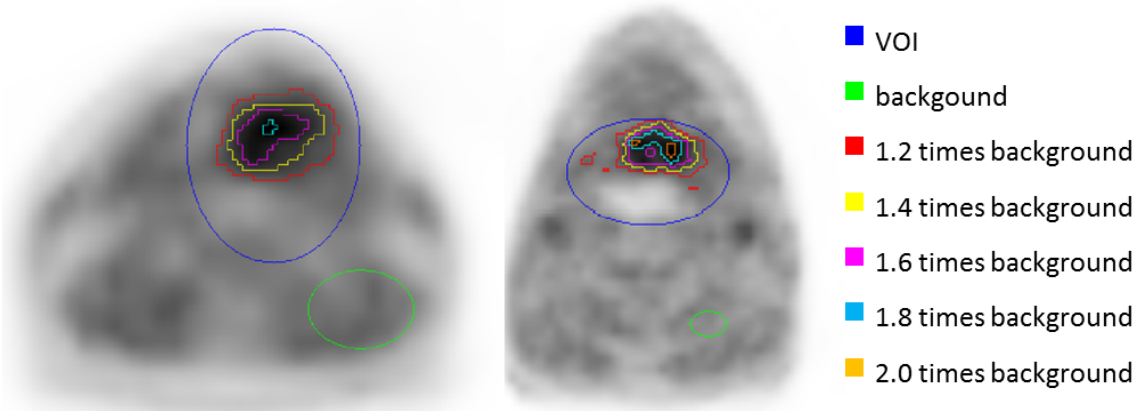


Figure 2.16.: After definition of the VOI (blue) and background volume (green), five relative thresholds were applied: 1.2 (red), 1.4 (yellow), 1.6 (pink), 1.8 (cyan) or 2.0 (orange) times background activity.

Last but not least, an approach for eliminating inter-patient differences could be to average background measurement through a whole patient cohort. Tachibana et al. determined mean SUV_{max} and standard deviation $\sigma_{SUV_{\text{max}}}$ to calculate an alternative, cohort based SUV-based threshold (Tachibana et al., 2013).

2.5.2. Threshold-based segmentation

While segmentation algorithms applicable to FMISO PET were rarely published, for FDG PET a number of algorithms were developed, validated and published so far (Daisne et al., 2003; Schaefer et al., 2008; Nehmeh et al., 2009; Schaefer et al., 2012; Hofheinz et al., 2012c,a). Some of these algorithms were documented in detail so that re-implementation is feasible. For example, one of these algorithms can be used to calculate an optimal threshold $T_{\text{opt,Schaefer}}$ from the activity concentration C_{SUV70} within a segment containing all voxels with activity concentration above $70\,\%$ of maximum concentration in the VOI and a determined background activity concentration $C_{\text{background}}$ (Schaefer et al., 2008):

$$T_{\text{opt,Schaefer}} = a \cdot C_{\text{SUV70}} + b \cdot C_{\text{background}} \tag{2.45}$$

The parameters of the algorithm were predefined $a = 0.5$ and $b = 0.5$ suggesting the optimal threshold is the mean activity concentration of the background and the temporary volume. In a later publication the authors determined the scanner- and site-dependent parameters using a phantom experiment (Schaefer et al., 2012). In this publication, $a = 0.44$ and $b = 0.453$ where determined for Biograph 16 PET/CT scanners. Applying these parameters to an FMISO PET image leads to contours as shown in Figure 2.17.

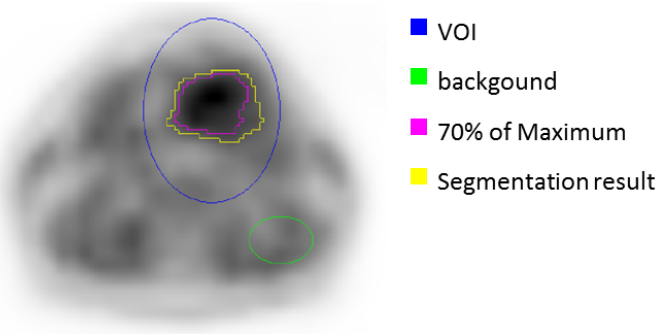


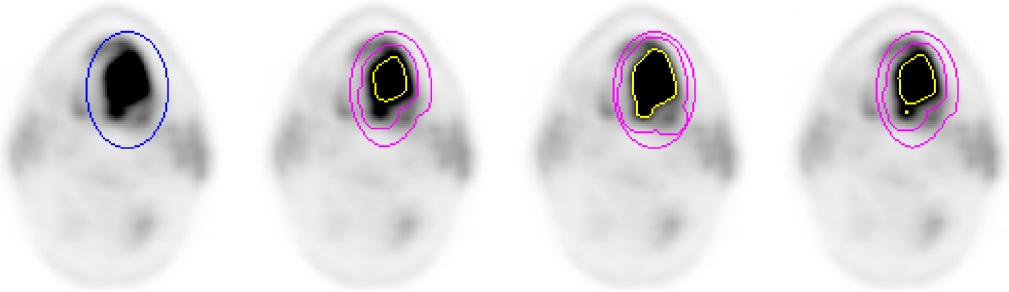
Figure 2.17.: After definition of a VOI (blue) and a background volume (green), the threshold estimation algorithm from (Schaefer et al., 2008) can be applied: Firstly, a volume including all voxels with activity concentration above $70\,\%$ of maximum activity concentration is segmented (pink). Afterwards, the mean activity concentration within this volume and activity concentration within the background volume serve to calculate a threshold. Applying this threshold to the VOI leads to the final segmentation (yellow).

A similar approach to threshold-estimation proposed by Hofheinz et al. provides a way for automatic background determination (Hofheinz et al., 2012c). As in the previous algorithm, an initial temporary segmentation is done by applying a fixed relative threshold such as $50\,\%$ of maximum activity concentration within the inspected VOI. Originating from this temporary volume, a background volume is segmented

by including all voxels within the inspected VOI which are a) distant from the temporary volume by at least one time reconstructed spatial resolution and less then 2.5 times that distance. This algorithm is supplied by Rover software (ABX, Radeberg, Germany). Exemplary contours including a visualisation of background and temporary volume are shown in Figure 2.18. The mean background activity concentration $C_{\text{background}}$ and the maximum activity concentration C_{max} in the VOI is further used to calculate a threshold $T_{\text{opt,Hofheinz}}$:

$$T_{\text{opt,Hofheinz}} = T_{rel} \cdot (C_{\text{max}}(\text{VOI}) - C_{\text{background}}) + C_{\text{background}}, \text{with } T_{\text{rel}} = 39\,\% \tag{2.46}$$

The process of threshold-estimation, temporary volume segmentation and background determination is repeated iteratively until the volume definition does not change anymore.



(a) VOI definition (b) Iteration 1 (c) Iteration 2 (d) Iteration 3

Figure 2.18.: After VOI definition (a, blue), an initial target volume contour is derived by thresholding (b, yellow). Distant to this contour, a background volume is automatically defined (b, pink). Resulting from these target and background definitions, a new threshold is calculated and applied (c, yellow). This process is iteratively repeated until the target volume definition does not change anymore.

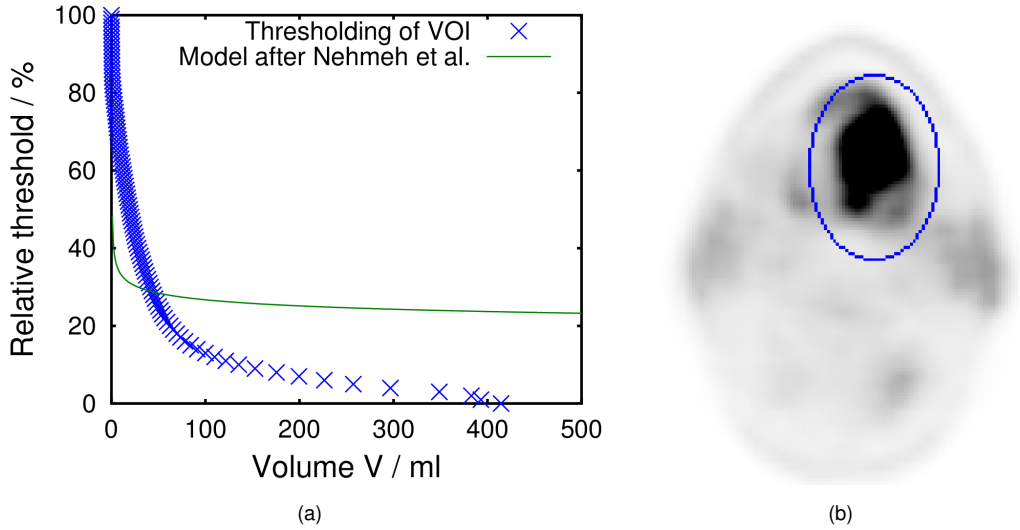
Combining the threshold estimation in equation (2.45) from Schaefer et al. and the automatic background determination approach from Hofheinz et al. leads to a hybrid algorithm which was developed and validated using phantom experiments in a multi-centre-investigation (Schaefer et al., 2012).

Alternatively to phantom experiments, Monte-Carlo simulations may be used to derive a model for estimating appropriate thresholds (Nehmeh et al., 2009). A mathematically derived model describes the relationship between PET-positive volume V and threshold $C_{\text{rel,min}}$ applied to segment the volume:

$$C_{\text{rel,min}} = a_0 + e^{a+\frac{b}{V}+c\log V}\,\% \tag{2.47}$$
$$\text{with } a_0 = 5, a = 3.568, b = 0.197, c = -0.1069$$

By applying a range of thresholds $C_{rel,min} = [0; 100]\,\%$ to a volume of interest in a specific PET data set, a curve as shown in Figure 2.19 can be generated describing the relationship between threshold and volume for this data set. According to Nehmeh et al., the optimal threshold of target volume delineation is located where this curve and the curve of the model cross. This approach is applied for two times iteratively. The first segmentation is used to determine background activity concentration from all voxels which were not defined as positive. Afterwards, the background is subtracted from all voxels in the data

set and the algorithm is applied again to generate the final target object delineation. The developers tested this algorithm using phantom experiments and it was already applied for clinical investigations on FDG PET data sets from head-and-neck-cancer patients (Lee et al., 2008). Applied to FMISO PET images, resulting contours as shown in Figure 2.20 can be retrieved.



(a) (b)

Figure 2.19.: Nehmeh et al. derived a model for an optimal relationship between threshold and volume for PET image segmentation. This relationship is plotted in green in a). The blue crosses represent corresponding volume measurements by applying one hundred thresholds between 0 and $100\,\%$ of maximum activity concentration to a volume of interest encircled in blue in a clinical FDG PET data set in b). The optimal threshold is located where blue data points cross the green line.

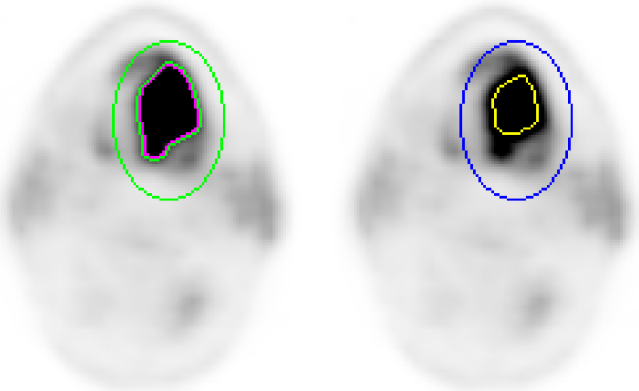
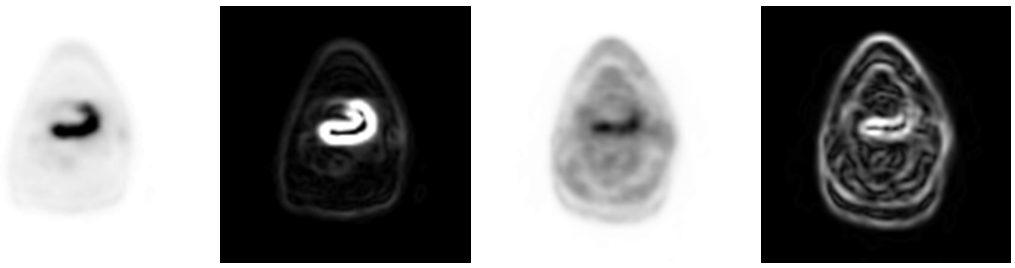


Figure 2.20.: The algorithm from (Nehmeh et al., 2009) is used to determine an initial threshold for distinguishing between target (pink) and background (green). Afterwards, the background activity concentration is temporarily subtracted and a new threshold is calculated. Finally, the target volume is thresholded (yellow) within the volume of interest (blue).

2.5.3. Gradient based segmentation methods

Aside from the established methods based on thresholding, gradient based techniques were developed for target volume delineation in PET (Geets et al., 2007). The idea is comprehensible: If target volume

and background differ in signal intensity, there must be a region between both volumes with a gradient or edge in the signal intensity. However, the fact that the contrast between target and background in FMISO PET images is about one order of magnitude lower than in FDG PET also induces that the gradient between both volumes may be less prominent. To visualise this, edge images resulting from a Sobel operator are shown in Figure 2.21. Gradient-based target volume definition in PET images may fail because the edge between target and background may not be distinguishable from other edges in the image, for example between muscle and lipid tissue.



(a) (b) (c) (d)

Figure 2.21.: To visualise the presence of edges in FDG (a) and FMISO (c) PET images, a sobel operator was applied to two PET image slices from the same head-and-neck-cancer patient. The resulting edge image corresponding to the FDG PET (b) clearly shows the presence of an edge between target volume and background. In the edge image corresponding to the FMISO PET (d), several present edges are perceivable and the edge around the target volume is not easily distinguishable from the others.

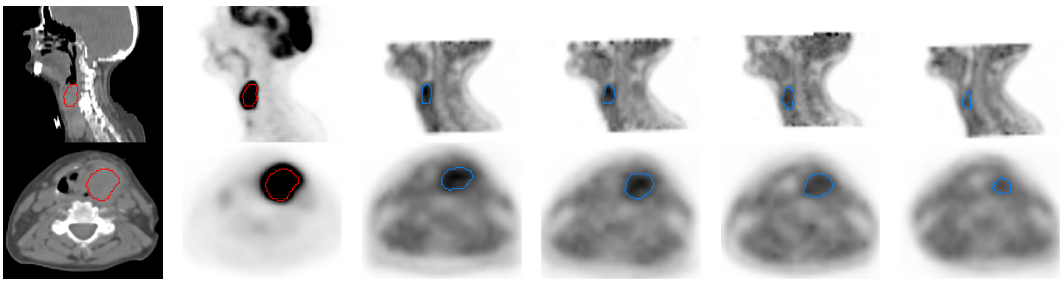
2.6. Multimodality imaging in oncology

Multimodality imaging is well known as a valuable tool for diagnostics in oncology. CT is the standard imaging technique for target volume definition for radiotherapy treatment planning, MRI adds deeper insights for soft tissue differentation and functional information. PET imaging introduces metabolic imaging. Today, there are more possible non-invasive imaging techniques available for tumour tissue differentation than actually could be used in clinical routine. Thus, research focus is how to choose the right imaging methods for therapy planning, response monitoring and follow up.

2.6.1. Biological imaging for response monitoring

Non-invasive imaging for response monitoring is a major key to making clinical decisions for individualised patient care. If therapy failure could be detected early during treatment, the affected patient could be scheduled to another therapy or different medication. However, this kind of therapy individualisation in oncology is still under research today. Many of the performed clinical studies are retrospective or observative studies used to estimate the potential of non-invasive biological imaging. FDG PET imaging may be utilised to visualise early effects of chemotherapy (Haberkorn et al., 1993). FMISO PET may allow differentiating patients with local recurrent tumours and healed ones (Eschmann et al., 2005; Zips

et al., 2012). The combination of FDG PET, FMISO PET, diffusion weighted MRI and dynamic contrast enhanced MRI may furthermore allow early response monitoring (Dirix et al., 2009). But the more imaging modalities are used, the more complex is the task of image analysis; no matter if this task is performed manually or automatically. Furthermore, for some imaging modalities such as FMISO PET, analysis becomes more complicated, the later during therapy the images were acquired: A decrease in target-to-background ratio during therapy leads to images with even more decreased contrast which are thus more complicated to analyse (Eschmann et al., 2007). An example of a patient imaged before combined radio-chemotherapy, and after weeks one, two and four is visualised in Figure 2.22. Obviously, the volume measurement is dependent on contrast in the image. Furthermore, the mean TMR measurement depends on a reliable volume definition. The fact that in FMISO PET analysis currently TMR_{max} measurements should be preffered for prognostic estimations (Zips et al., 2012), may be a result of the inability to derive accurate volume definitions.



(a) (b) (c) (d) (e) (f)

Figure 2.22.: To use FMISO PET for therapy monitoring, reliable hypoxic volume definitions are needed. The tumour volume definition (red) is visible in the CT (a) and the FDG PET (b) images. In this particular example, the FMISO-positive volume (blue) appears stable during the first three imaging time points, acquired before therapy (c) after ≈ 10 Gy (d) and ≈ 20 Gy (e). A volume shrinkage is later observed in the FMISO PET data set acquired after ≈ 55 Gy (f).

As possible alternative to segmentation of static FMISO PET images, kinetic analysis of dynamic PET imaging studies shall not be unmentioned. By imaging tracer dynamics during the first minutes after injection and the uptake several hours after injection, a two-compartment model can be applied that allows estimation of perfusion and hypoxia. These parameters correlated to local control in head-and-neck-cancer patients (Thorwarth et al., 2005). Well perfused and normoxic tumours responded well to radiotherapy in general, and hypoxic tumours, which were often worse perfused, did not and often resulted in local recurrence after a few months (Thorwarth and Alber, 2008). However, there are several approaches for analysing or modelling dynamic PET data (Thorwarth et al., 2005; van den Hoff et al., 2013) and a standardisation is not foreseeable. Furthermore, available models may be oversimplifications to describe radioresponse on voxel-level (Bentzen and Gregoire, 2011).

2.7. Biological imaging for radiotherapy treatment planning

The incorporation of PET imaging into radiotherapy treatment planning was introduced more than a decade ago, because the advantages seemed overwhelming. FDG PET based target volume definition in

lung cancer is valuable for differentating mediastinal lymph nodes from the mediastinum and atelectasis from metabolic tumour volumes (Steenbakkers et al., 2005). However, in FDG-positive lymph nodes, preoperative surgical staging may still be recommended to prevent over-staging (Poncelet et al., 2001). Furthermore, the inter-observer-variability in target definition can be reduced, if FDG PET is fused with the CT data in comparison with the CT images alone (Caldwell et al., 2001; Spratt et al., 2010). Thus, today PET/CT image fusion is recommended for accurate treatment planning in many tumour entities. However, PET-based contour definition must be developed further, as well as the handling of motion, for example through breathing (Baumann et al., 2008). Effects of the latter may be decreased through introduction of equal breathing protocols during imaging and treatment (Grgic et al., 2009). Aside from technology development, standardised procedures in this field would be valuable as well.

At the end, treatment planning based on PET imaging alone may not make the way to the clinic, because metabolic imaging cannot replace morphological imaging. Both modalities, PET and CT, deliver complementary information and thus must be used in combination for radiotherapy treatment planning (Black et al., 2004; Dirix et al., 2009; Devic et al., 2010). MRI may deliver even more information needed for radiotherapy treatment planning. For example, early response assessment in head-and-neck-cancer appears feasible using diffusion-weighed and dynamic contrast enhanced MRI (Dirix et al., 2009). Furthermore, standard clinical MRI sequences, such as Gadolinium-enhanced T1-weighted and T2-weighted, were compared to PET images and it was shown that both modalities deliver complementary but distinct information in human glioblastomas (Swanson et al., 2009).

2.7.1. Hypoxia PET in radiotherapy

While metabolic FDG PET imaging is now part of clinical routine for several years, tracers visualising other biological aspects such as apoptosis and hypoxia are part of clinical research (Thorwarth et al., 2007; Lee et al., 2008; Thorwarth and Alber, 2008, 2010). FMISO PET visualising hypoxia revealed to deliver independent information from FDG PET (Thorwarth et al., 2006) and thus, may be of value for more sophisticated radiotherapy treatment planning, especially for dose escalation (Yasuda et al., 2012). Furthermore, radiotherapy treatment planning studies were performed, showing that FMISO-based dose escalation was possible up to $85\,\mathrm{Gy}$ deposited within the hypoxic volume, without exceeding normal tissue dose limitations (Lee et al., 2008). Dose-painting-by-numbers, a technique for local modulation of deposited dose within the tumour volume, may become possible if hypoxia could be quantified adequately (Thorwarth et al., 2007) and reproducibly. Thus, target volume definition in FMISO PET as well as reproducibility of FMISO PET based contouring are the challenging tasks of current research.

The limiting contrast levels observed in FMISO PET imaging may not be surmounted using alternative hypoxia tracers such as $[^{18}\mathrm{F}]$-fluoroazomycin arabinoside (FAZA) or $[^{18}\mathrm{F}]$-2-(2-nitroimidazol-1-yl)-N-(3,3,3-trifluoropropyl)acetamide (FEF3). The latter was shown to be not superior to FMISO even though tracer distribution were similar. But contrast levels observed in PET images of rat models were lower than when using FMISO as hypoxia tracer (Mahy et al., 2008; Dubois et al., 2009). In FAZA PET small animal imaging studies, the tracer showed higher contrast levels than FMISO (Reischl et al., 2007). As well as FMISO, FAZA appears to reveal prognostic potential in head-and-neck-cancer (Mortensen et al.,

2012). Furthermore, hypoxia-directed intensity modulated radiotherapy appears feasible using FAZA PET (Grosu et al., 2007). However, the fact that the target-to-muscle-ratio target volumes in FAZA PET is just above 1.5 (Souvatzoglou et al., 2007), suggests that contouring these structures is as complicated as when using FMISO PET.

3. Materials and methods

This chapter gives an insight into the data sets and procedures used to optimise and validate the proposed ant-based segmentation algorithm. Beginning with phantom measurements but also utilising clinical FMISO PET data sets from head-and-neck-cancer patients. The processing of the data sets before segmentation is explained as well as the ant-based algorithm and its parameters.

3.1. Imaging data

3.1.1. PET phantom experimental data

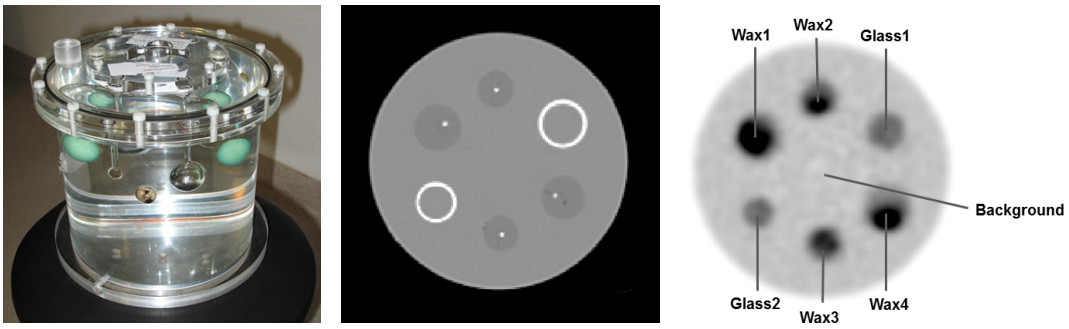
For the simulation of target volumes in PET images under different contrast levels, a phantom experiment was set up in cooperation with the Clinic and Policlinic for Nuclear Medicine of the University Hospital Carl Gustav Carus TU Dresden. The resulting DICOM data sets were used for evaluation of several segmentation algorithms. The performed PET phantom experiment simulated target objects under different contrast levels. The cylinder phantom, shown in Figure 3.1, contained a solution of $[^{18}\text{F}]$-fluorodeoxyglucose (FDG). In the cylinder, six spheres containing a solution of $[^{68}\text{Ga}]$-gallium-chloride were mounted to simulate target objects. Four of the spheres were made of wax, the remaining two spheres were glass hollow spheres filled with tracer solution. The different radionuclides were used because of the different half-life:

$$T_{1/2}(^{18}\text{F}) = 109.8\text{min}$$
$$T_{1/2}(^{68}\text{Ga}) = 67.6\text{min}$$

Due to different half-life, the contrast between target volumes and background decreases with time. The phantom was measured in a combined PET/CT scanner (Biograph 16, Siemens, Knoxville, USA) for about 10 hours. Afterwards 63 data sets were reconstructed. The frame duration of these data sets was 10 minutes. Voxel size for reconstruction was $1.3\,\text{mm} \times 1.3\,\text{mm} \times 2\,\text{mm}$.

3.1.2. Patients

To determine the performance of the proposed algorithm as well as its established alternative algorithms, application to clinical patient data sets was performed. Clinical FMISO PET data sets of 43 head-and-



(a) (b) (c)

Figure 3.1.: A cylindrical phantom is shown in a) (photography source: Dr. Andreeff M, Clinic and Policlinic for Nuclear Medicine, University Hospital Carl Gustav Carus, TU Dresden). For the performed experiment, six spheres were mounted inside the cylinder. Two were hollow spheres filled with fluids. The glass walls of these two spheres are clearly visible in the CT scan of the phantom shown in b). Four more spheres were made of wax. The naming of the six test targets is given in the PET image shown in c).

neck-cancer patients were acquired as part of a prospective trial between 2006 and 2011. First analysis on this study and more precisely on the prognostic potential of the acquired FMISO PET imaging data was already published (Zips et al., 2012). The study was approved by the Institutional Ethics board and the Federal Office for Radiation Protection. The patients were informed about the further electronic processing of the imaging and the aims of the study. All patients gave their written informed consent. From these patients FMISO PET data sets were acquired before and during radiotherapy. The imaging study layout is shown in Figure 3.2. The scans included a dynamic PET scan immediately after tracer injection and two static PET scans about two and four hours after injection. For the presented analysis in this thesis, only the static data sets acquired four hours after tracer injection were used, because the contrast between target and background was the highest in this case (Abolmaali et al., 2012). Not all patients were scanned for four times using FMISO PET because of sometimes organisational and mostly patient related issues. At the end, 43 initial scans, 35 scans after the first, 38 after the second and 38 after the fourth week of therapy had been acquired. In summary, 155 clinical FMISO PET data sets were available for analysis.

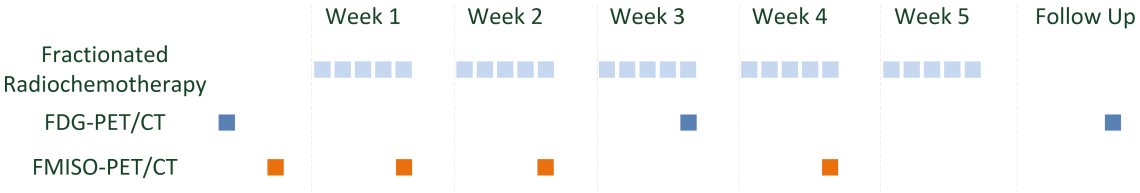


Figure 3.2.: The layout of the FMISO PET imaging study: Before therapy start, all patients were scanned using FDG- and FMISO PET/CT with a gap of at least one day between. Afterwards, during therapy further PET/CT scans were scheduled.

3.1.3. Image acquisition

PET/CT imaging was performed using a Biograph 16 PET/CT scanner. Low dose CT images were acquired for attenuation correction with 120 kV tube voltage with online tube current modulation between 20 mAs and 100 mAs. CT images were reconstructed with voxel size of 1.37 mm $\times$ 1.37 mm in x- and y-direction and 5 mm in z-direction. This resulted in images with a matrix size of 512×512 pixels. The CT images acquired in tandem with FDG PET were whole body images, including slices from the base of the skull to the upper thigh. CT images acquired together with FMISO PET, only contained one bed position around the tumour. This usually included slices from maxilla to shoulder.

The FDG and FMISO tracers were produced and delivered by the Institute of Radiopharmaceutical Cancer Research at the Helmholtz Centre Dresden-Rossendorf. One hour before the FDG PET/CT scan, the patients received an intravenous injection of 350 MBq FDG. Each bed position of the whole body scan was acquired for three minutes. The FMISO PET/CT scan used in this work was performed four hours after injection of $250 - 300$ MBq FMISO. FDG and FMISO PET images were reconstructed using the iterative OSEM reconstruction algorithm provided by the vendor of the PET/CT scanner. PET reconstruction resulted in images with 168×168 matrix size with voxel size in plane of 4.06 mm $\times$ 4.06 mm. Slice thickness as well as slice distance were 5 mm. FDG PET images were denoised using a 5 mm Full-Width-at-Half-Maximum (FWHM) Gaussian filter as typically applied in clinical scan protocols. The FMISO images were denoised using a Gaussian filter with 1 mm FWHM. Experienced clinicians from the clinics of nuclear medicine and radiology decided together that images reconstructed using this reduced denoising filter were better interpretable. After a workstation software update in 2009, after about half of the patients were scanned, only 2 mm FHWM was available for reconstruction. Thus, the reconstruction protocol of the FMISO images had to be updated as well.

3.1.4. Image registration

To allow voxel-by-voxel analysis of corresponding regions between subsequent scans, the PET data sets were registered. The procedure, visualised in Figure 3.4, consisted of four steps, performed on pairs of CT data sets acquired for attenuation correction and corresponding PET data sets:

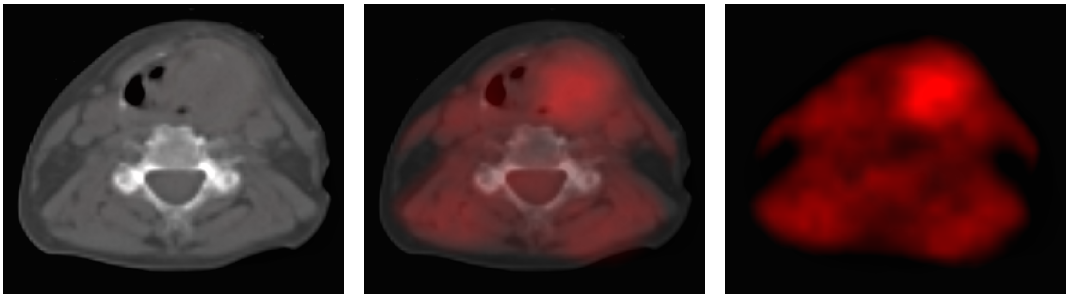
1. **CT-CT-Registration** Initially, every CT acquired during the imaging study was registered to the initial CT data set. Registration was performed in Pinnacle 9.0 software (Phillips, Eindhoven, Netherlands) by a radiological assistant. Afterwards, registration parameters were documented. Images from six patients had to be registered without the Pinnacle software because it was not available for technical reasons. In these cases the procedure started with step 2.

2. **CT-PET-Registration** The initial CT and one of the PET data sets were loaded into Rover software. The registration parameters determined in the previous step where then entered in Rover to align any PET data set to the initial CT. To adapt registration parameters from Pinnacle in Rover, the transformation along the z-axis needed to be inverted. If needed, the registration was corrected manually. In focus of this step was the region around the tumour as visible in the corresponding

initial FDG PET. Figure 3.3 shows how FMISO PET allows rough differentiation of bone, muscle and lipid tissues. Thus, registration of PET and CT was performed by aligning muscle and PET-visible bone structures. After this step, resampled images were exported to the file system.

3. **PET-PET-Registration-Approval** To check the transformation applied to an FMISO PET image, it was fused in in-house-software with the initial FDG PET using the registration parameters determine in the previous step. If a shift between structures near the tumour were visible, as exemplary shown in Figure 3.5, the registration parameters were rejected and step 2 was performed again.

4. **CT-PET-Registration-Approval** To check if the PET data sets match the initial CT an experienced radiologist loaded pairs of CT and PET image stacks in a syngo workstation (Siemens, Germany) and checked the registration visually. If this test failed, step 2 and 3 were performed again.



(a) CT (b) CT-PET-fusion (c) FMISO PET

Figure 3.3.: For manual segmentation correction, FMISO PET visible muscle structures were placed over muscle visible in the corresponding CT images. This procedure is feasible because FMISO is also accumulating in normal tissues such as muscle.

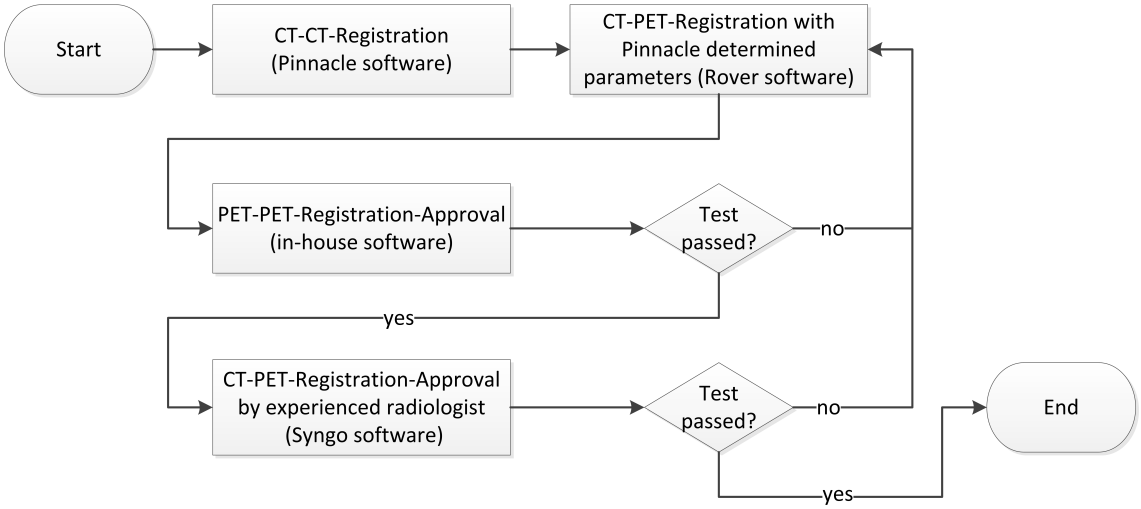
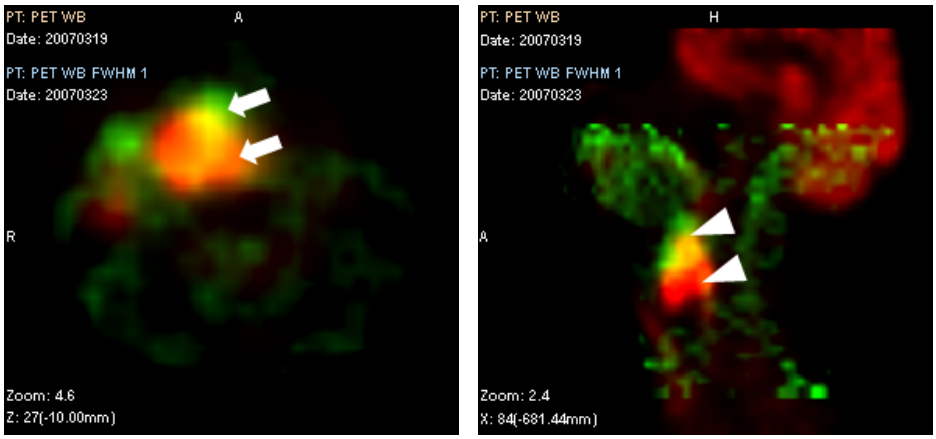


Figure 3.4.: The performed registration procedure to generate CT and PET images in the same coordinate system consisted of four steps: two initial steps for the registration and two added steps of visual registration quality approval. If one of the two tests failed, only the second registration step was performed again, because the CT-CT-registration was not checked in the quality approval routine, since this registration is standardised and part of the Pinnacle software which is a medical product. The final quality check was performed by an experienced radiologist to ensure the reliability of the registration parameters.



(a) transverse (b) sagittal

Figure 3.5.: The PET-PET-registration check was performed to evaluate the registration errors accumulated by registration of CT and PET images. If the PET images did not match together, at least on pair of PET/CT images must have been registered with questionable registration parameters. In the shown case, an apparent dorsal-ventral shift in transverse plane (a) of the primary tumour between FDG (red) and FMISO (green) PET images is explainable in sagittal plane (b): in this example, the registration between the PET images is obviously wrong in cranial-caudal-direction (white triangles).

3.1.5. Pseudonymisation and randomisation

After all PET images were registered to the initial CT, a patient data set, consisting of one CT and up to 5 PET data sets, was prepared on the filesystem. The PET data sets were pseudonymised and the name of the patient was replaced by a six digit code. In this procedure, also PET data sets from one patient retrieved new random patient names. The connection between the data sets, which was given by patient name and patient identifier, was disconnected. Thus, if the pseudonymised data sets were ordered by patient name, the data sets were ordered randomly. An observer who processes the data sets in alphabetical order would not know whether a specific data sets was acquired before or during therapy. After pseudonymisation and randomisation the data sets were prepared for contouring.

3.1.6. Manual delineations

A volume of interest (VOI) including the primary tumour and surrounding normal tissue was defined manually by an experienced radiologist by placing a spheroidal mask on the initial FDG PET data set in Rover software. The mask was placed in a position so that FDG-positive lymph nodes were excluded as far as possible. All algorithms explained in section 2.5.2 were later applied to this VOI. Another spheroidal mask was placed in cervical muscle as background definition. This mask was placed on the fused view of the initial CT and the registered initial FMISO PET scan. Example VOIs for both are shown in Figure 3.6.

For the purpose of manual contouring the registered, pseudonymised and randomised FMISO PET data sets were stored on a Syngo workstation. All data sets were processed manually by three experienced

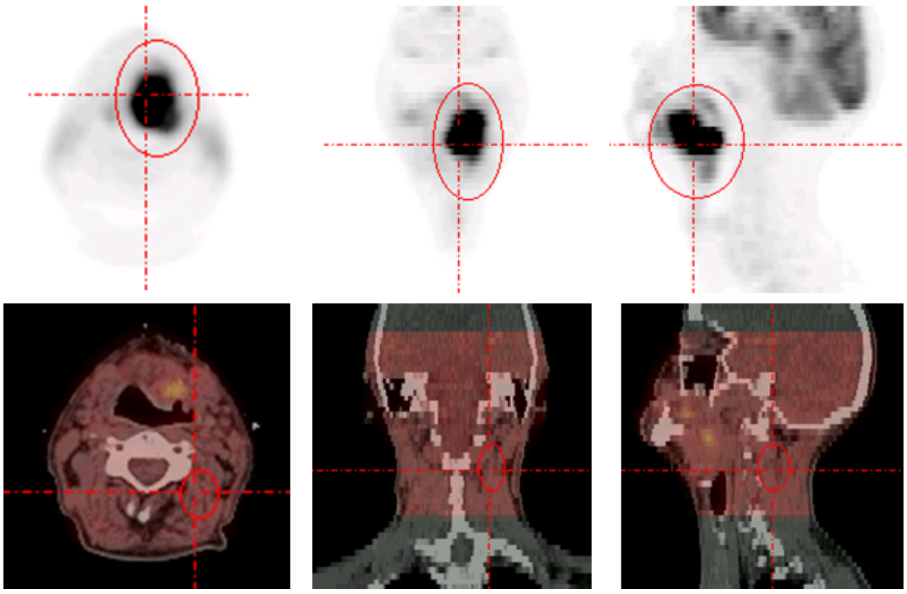


Figure 3.6.: Volume of interest around the primary tumour volume (top) and the background volume in cervical muscle tissue (bottom) were defined manually in Rover software by an experienced radiologist.

observers: a radiologist, a nuclear medicine physician and a medical physicist with experience in radiotherapy treatment planning. The observers outlined regions with PET-visible FMISO uptake related to the primary tumour. If possible, the observers excluded FMISO-positive lymph nodes. The observers were blinded to patient name, staging and therapy progress. Thus, they were not able to see if the processed data set was a scan acquired before or during therapy. After manual contouring, it was checked whether the drawn outlines were within the VOI introduced above. If not, it is obvious that the corresponding observer outlined an FMISO-positive volume which is not related to the primary tumour, but to tracer accumulating in mucous tissue or in FMISO-positive lymph nodes. In these cases, all observers discussed together if the contour should be removed from the further analysis or redrawn. The affected cases were documented.

3.2. Proposed algorithm: Ant-based contouring

The proposed ant-based algorithm for PET image segmentation, or rather contour generation, was developed earlier (Haase, 2010) and is reviewed in this chapter. The actual ant colony simulation is a part of an image processing routine which imports a PET image stack and exports a target volume definition, or rather a list of contours on transverse image slices. The simulation itself consists of two loops. The inner loop handles all ants. For every ant actions like motion, ageing and self-reproduction need to be performed. Furthermore, new ants seeded in the PET volume and some ants are selected to be eliminated. The outer loop repeats these actions for a given number of iterations. The procedure is visualised in figure 3.7 and the subprocesses are explained in detail in the following sections.

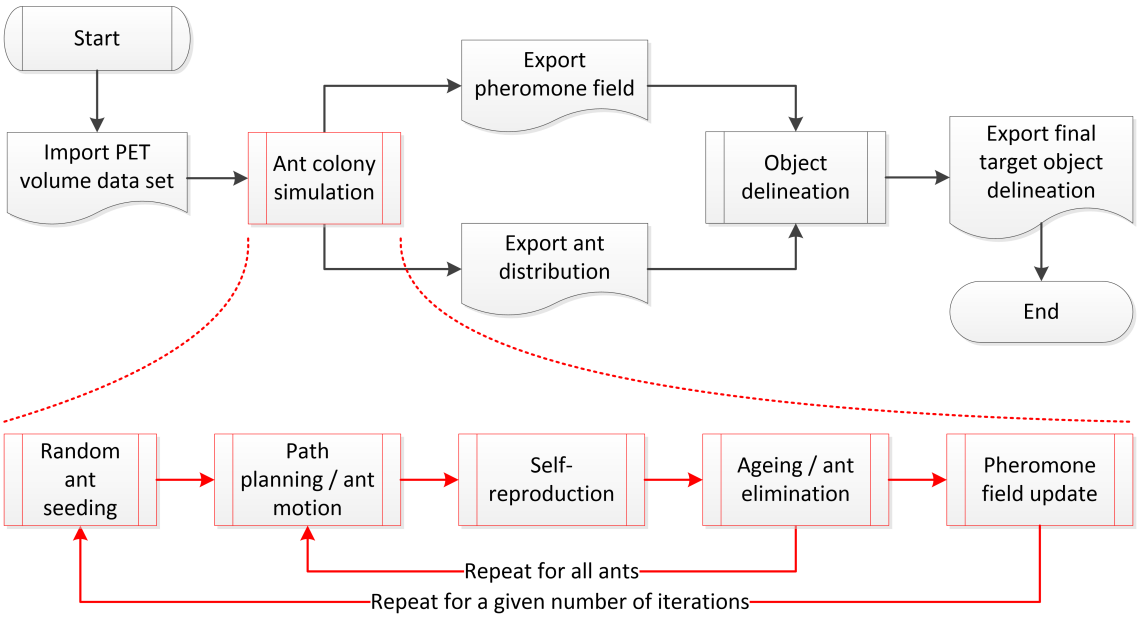


Figure 3.7.: The ant-based image segmentation algorithm is visualised schematically. The ant colony simulation is only a part of the whole algorithm which consists of two inner loops handling the single ants of the colony for a given number of iterations. This figure is adapted from (Haase et al., 2012a).

3.2.1. Ant path planning and motion

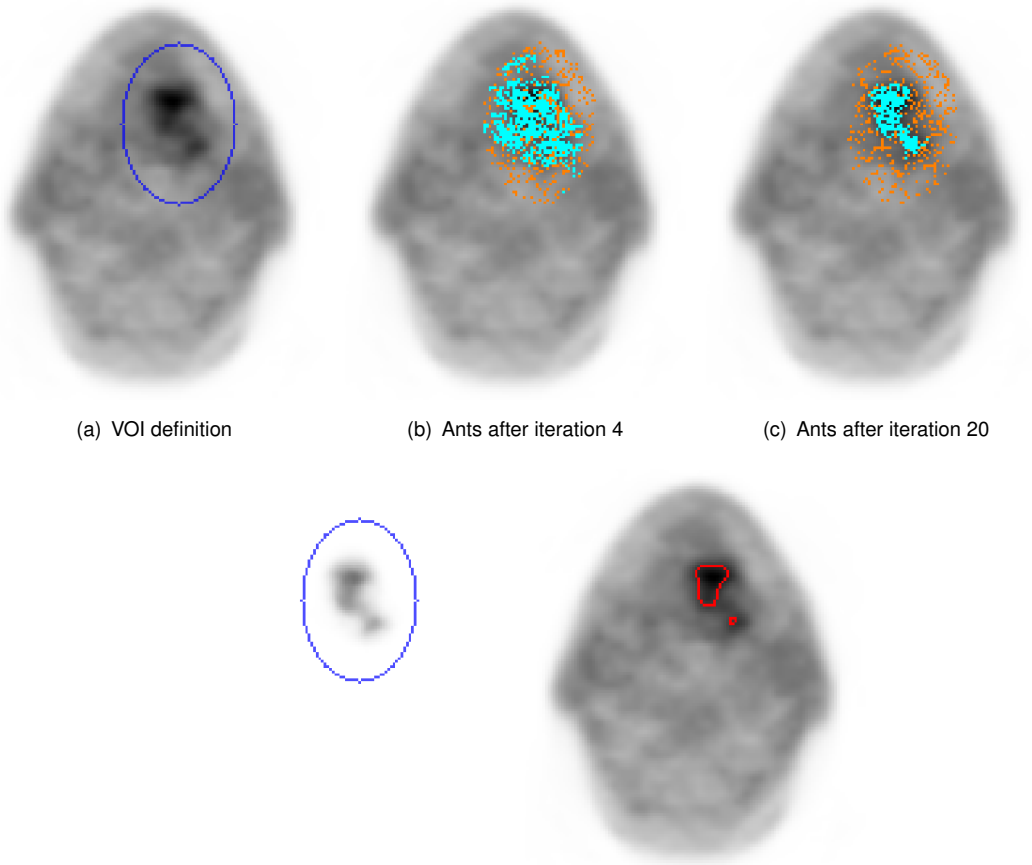
As introduced in section 2.4.2, ants calculate a probability of every possible location they could go to by doing a single step. The given formula for ants path planning can be adapted to PET image processing by defining activity concentration C of the voxels as the objective function. Thus, ants travel along the gradient in the PET image to a local maximum. Ants also consider the pheromone intensities of the voxels in their Moore-neighbourhood. Thus, the adapted formula for path planning is:

$$p_{ij} = \frac{[\tau_j]^\alpha [C_j]^\beta}{\sum_{k \in M(i)} [\tau_k]^\alpha [C_k]^\beta} \tag{3.1}$$

This formula allows the ants to overcome local optima in the PET image, if the optima are not present in the pheromone field because of pheromone diffusion. Furthermore, the parameter configuration of α and β should allow the ants to perceive the pheromone field.

3.2.2. Self-reproduction

The aim of the ant motion in the PET volume is to fill the target object with ants. To assure that the whole object could be filled with ants, the number of ants must be higher than the number of voxels within the target object. The fact that this number is unknown, denies to define the population size a priori. Thus, the colony must be allowed to vary the population size on its own. An approach for self-regulation of the population size was earlier proposed (Fernandes et al., 2005b,a): The principle called self-reproduction allows increasing the size of the population. For the proposed ant-based algorithm, self-reproduction



(a) VOI definition (b) Ants after iteration 4 (c) Ants after iteration 20

(d) Pheromone field after it. 20 (e) Final contour

Figure 3.8.: Ant-based delineation deserves a volume of interest (a) to randomly seed scout ants (orange) which produce worker ants (blue) in regions with increased signal intensity. Initially worker ants appear in regions outside the target volume (b). After 20 iterations, the worker ants concentrate in the target volume (c). From the pheromone field (d), the ants use for indirect communication, a target volume definition can be derived from (e).

rules were defined to allow only ants in the vicinity of the target volume to reproduce. This is thought to increase the process of accumulation inside this volume. Thus, the probability to reproduce p_r, C of an ant sitting on voxel i depends on the activity concentration C_i and activity concentration threshold $C_{\text{min}}(t)$:

$$p_{r,C} = \begin{cases} 1, \text{ if } C_i > C_{\text{min}}(t) \\ 0, \text{ else.} \end{cases} \tag{3.2}$$

with

$$C_{\text{min}}(t) = max_{T=1...t_{max,\bar{C}}}(\bar{C}_T). \tag{3.3}$$

The reproduction threshold $C_{\text{min}}(t)$ ensures that only ants sitting on voxels with activity concentration above the average can reproduce. This threshold is calculated as the maximum of the mean activity

measured in any previous iteration. Thus, this threshold may change during every iteration, but it can only increase. It does so, if after an iteration the mean activity concentration of all voxels being omitted by ants is higher than the reproduction threshold used during the iteration.

Another similar reproduction rule leading to the probability $p_{r,\tau}$ depends on the intensity values stored in the pheromone field. The corresponding threshold allowing ants to reproduce increases from iteration to iteration, until the activity concentration threshold introduced in the last paragraph does no longer increase:

$$p_{r,\tau} = \begin{cases} 1, \text{ if } \tau_{i,t} > P_\tau \cdot max_i(\tau_i, t) argmax_{T=1\ldots t}(C_{\text{min}}(T)) \\ 0, \text{ else.} \end{cases} \tag{3.4}$$

Both reproduction rules leading to the probabilities $p_{r,C}$ and $p_{r,\tau}$ are then used to compute the final reproduction probability p_r used for selecting ants for self-reproduction:

$$p_r = p_{r,C} \cdot p_{r,\tau} \tag{3.5}$$

The probability to reproduce p_r obviously is binary. It will always be either 0 or 1. As an alternative, a probabilistic rule was also experimentally tested in the proposed algorithm but delivered less reproducible volume definitions (Haase et al., 2011a).

The counterpart to self-reproduction is called ageing and allows to decrease the population size. Every ant is created with a survival probability $p_s = 1$. Thus, it will survive the first iteration. With every further iteration t, its probability to survive decreases by $\Delta p_s \in [0; 1]$:

$$p_s(t+1) = p_s(t) - \Delta p_s \tag{3.6}$$

Depending on the parameter Δp_s the maximum lifetime of an ant is limited to $\frac{1}{\Delta p_s}$ iterations.

3.2.3. Pheromone field update

The pheromone field which is generated during the simulation is a crucial data structure, because the final target object delineation is derived from it. The genesis of the field is an iterative process which updates the field during every iteration of the ant-simulation. The update of the pheromone field is separated into three subprocesses: emission, evaporation and diffusion. Initially, the amount of pheromone $\Delta\tau$ a single ant deposits is calculated from the activity concentration C_i of the voxel i where the ant is located and the mean activity concentration $\bar{C}_t$ of all voxels being occupied by ants at the iteration t:

$$\Delta\tau_i(t) = \max(0, C_i - \bar{C}(t)) \tag{3.7}$$

Thus, only ants on voxels with activity above the average emit pheromone. Pheromone emission is performed simultaneously with evaporation in one step. Evaporation is the process which slowly decreases

pheromone intensities on a voxel where no ant emits new pheromone. It furthermore diminishes intense changes to the pheromone field. The prospective pheromone intensity τ' after the current iteration is calculated from the prior pheromone intensity τ and the pheromone change $\Delta\tau$. The weighting between both values is achieved using the evaporation coefficient $\rho \in [0; 1]$ which was introduced by (Colorni et al., 1992):

$$\tau_i'(t) = (1 - \rho)\tau_i(t) + \rho\Delta\tau_i(t) \tag{3.8}$$

Afterwards, when all ants were allowed to emit pheromone, the pheromone field is blurred using a mean average filter applied to the von-Neumann-neighbourhood N including the current voxel. This process simulates diffusion of pheromone between voxels and is derived from (Ji et al., 2008):

$$\tau_i(t + 1) = \sum_{k \in N(i)} \tau_k'(t)/7 \tag{3.9}$$

3.2.4. Algorithm parameters and corresponding consequences

The proposed algorithm includes a number of parameters, which may influence final segmentation results. Table 3.1 gives an overview on the parameters, mandatory ranges and default values. If ∞ is given as boundary of a range, it is theoretically the boundary, but of course in practice a specific number is used. The influences of parameter changes are afterwards explained in detail in this section.

Table 3.1.: Overview on the parameters of the ant-based segmentation algorithm and corresponding default values and useful ranges.

Parameter	default value	comment
path planning weights α and β	caste dependent	in $[0, 1]$
evaporation constant ρ	0.1	in $[0, 1]$
pheromone threshold for self-reproduction P	0.015	in $[0, \infty]$
number of ants seeded in every iteration n_{seed}	$4\,\%$ of inspected voxels	
survival probability decrement δp_s	0.33	in $[0, 1]$
pheromone threshold for segmentation $\tau_m in$	90	in $[0, 255]$
number of smoothing operations n_s	1	in $[0, \infty]$

3.2.4.1. Path planning weights α and β

The parameters α and β used in the transition rule in equation (3.1) can be adapted to induce varying ant behaviour. For example, by setting $\alpha = 1$ and $\beta = 0$ the rule is simplyfied to

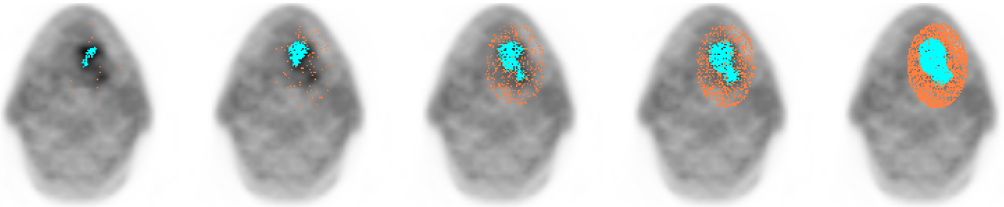
$$p_{ij} = \frac{\tau_j}{\sum_{k \in M(i)} \tau_k}. \tag{3.10}$$

Ants following this rule ignore activity concentration in the PET image and only perceive pheromone intensities. Analogously, if $\alpha = 0$ and $\beta = 1$ ants will follow the gradient in the PET image stack and ignore pheromone intensities. Randomly moving ants can also be simulated by setting $\alpha = 0$ and $\beta = 0$, because it results in equal probabilities $p_{ij} = \frac{1}{n}$ with n representing the number of voxels

where the ant could go to. Thus, ants move randomly in this case. A more detailed comparison of what happens, if transition rules were manipulated by changing α and β, was published earlier (Haase et al., 2012a). In summary, randomly walking ants were also able to delineate the target volumes, but with reduced contour reproducibility. Finally, it is worthwhile to split the ant colony in two castes, scout ants with $\alpha = 0$ and $\beta = 1$ for exploration and worker ants with $\alpha = 1$ and $\beta = 0$ for exploitation of image regions with increased signal.

3.2.4.2. Colony population size

As stated above in section 3.2.2, the ant colony is allowed and able to change its population size by self-reproduction and ageing. However, the process behind involves two parameters which need to be determined for an optimal configuration of the algorithm: a) the number n_{seed} of scout ants seeded at the beginning of every iteration, b) the aging parameter Δp_s. The first effect is shown in Figure 3.9 and was analysed earlier in detail (Haase, 2010). In short, there is a number of ants which needs to be seeded at least. It depends on the number of voxels being inspected and thus, in general the number of seeded scout ants $n_{\text{seed,rel}}$ is given in percent in relation to the number of inspected voxels. In phantom experiments, this number corresponded to $1.3\,\%$ of voxels within the volume of interest. If the number of seeded ants is higher, more computational effort is needed, but it did effect segmentation results only to a small degree (Haase, 2010). However, when processing PET data sets from patients, it was observed that this number was too small in some cases. Thus, the proposed ant-based algorithm now has $4\,\%$ of inspected voxels as default configuration to ensure that these data sets also may be processed correctly as well.



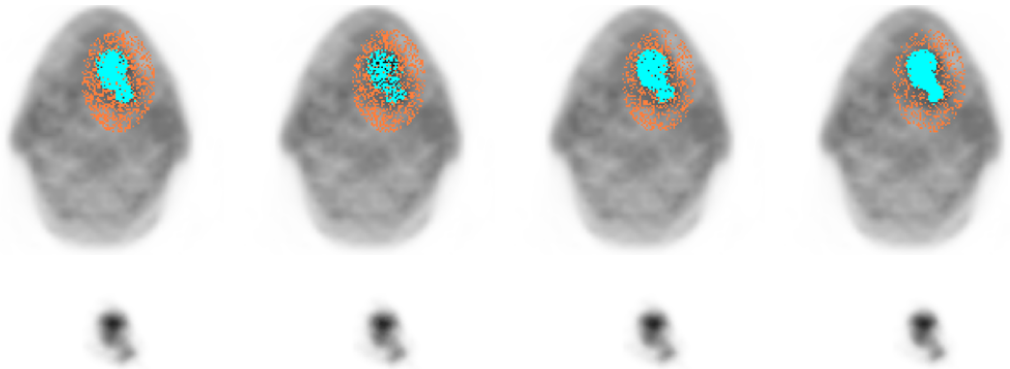
(a) Ant distribution created with $n_{\text{seed,rel}} = \{0.1, 1, 4, 10, 50\}\,\%$

(b) Corresponding pheromone fields created with $n_{\text{seed,rel}} = \{0.1, 1, 4, 10, 50\}\,\%$

Figure 3.9.: The number of seeded scout ants influences the ant distribution and pheromone field. If not enough ants are seeded, the ants cannot explore the whole volume and may miss to mark subvolumes of the target volume.

The second effect, corresponding to the decrease of survival probability of the ants in each iteration, the aging parameter Δp_s, was analysed in detail earlier as well (Haase et al., 2011a). In principle, volume

measurements were not dependent on Δp_s as long as the savage reproduction rule was utilised, as it is standard in the proposed algorithm. As shown in Figure 3.10, neither changes in the pheromone field nor in the ant distribution are obvious when applying different values for Δp_s. A reason may be seen in the fact that longer living ants block reproduction of new ants. Thus, if elimination of ants through aging is inhibited by low Δp_s values, reproduction is not possible as well and at the end, the population size stabilises anyhow. However, this is not the case when using a probabilistic reproduction rule. When applying it, $\Delta p_s < 0.05$ is recommended (Haase et al., 2011a).

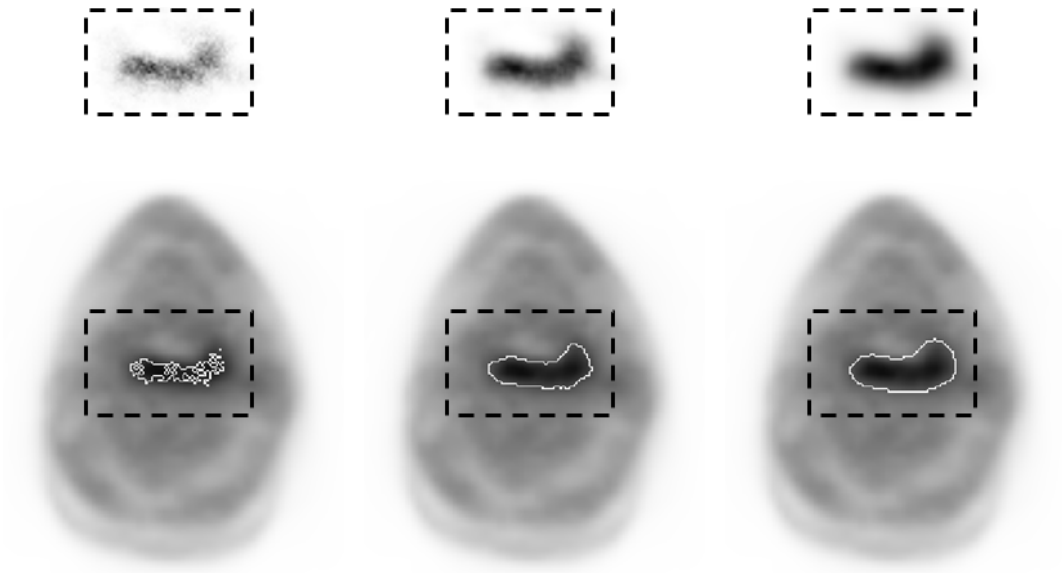


(a) $\Delta p_s = 0.001$ (b) $\Delta p_s = 0.01$ (c) $\Delta p_s = 0.1$ (d) $\Delta p_s = 1$

Figure 3.10.: When varying the aging parameter Δp_s, changes in ant distribution and pheromone field are not obvious. The exact value of the parameter may not be very relevant to the final segmentation result as well.

3.2.4.3. Pheromone diffusion by local averaging

As introduced in section 3.2.3, the pheromone field is locally averaged after every iteration using a mean average filter applied to the three dimensional Moore neighbourhood of every voxel in the pheromone field. This process is parameterised by the number n_s of iterations this filter is being applied. Example pheromone fields and finally resulting contours are given in Figure 3.11. Obviously, the smoothing operation is needed to spread pheromone into regions where no ant emitted pheromone. Disabling the pheromone diffusion by setting $n_s = 0$ leads to unplain contours because the pheromone field has several local maxima. Smoothing the pheromone field too much, for example by setting $n_s = 5$, the pheromone field no longer expresses the target volume edges correctly. Thus, as default value for the proposed algorithm, $n_s = 1$ was defined.



(a) (b) (c)

Figure 3.11.: The pheromone field is locally averaged after every iteration using a mean average filter. If the filter is not applied (a), the final pheromone field and corresponding target volume definition appears inappropriate. A single averaging operation (b) leads to a visually accurate volume definition. By applying the filter several times, such as 5 times in (c), the pheromone field loses shape and the final contour loses spatial accuracy. Thus, a single application of the filter is default in the algorithm.

3.3. Practical software implementation

The best algorithm is useless, if there is no handy user interface to operate it. For this reason, the proposed ant-based segmentation algorithm was embedded in the pre-existent Geisterr software which is maintained at the OncoRay. The software was earlier developed for image analysis and rather image fusion and registration. Within this project the established algorithms for automatic PET-based contouring given in section 2.5.2 were embedded in the software as well. Thus, it is possible to oppose segmentation results of the various algorithms directly. Through the fact that the software already allowed image fusion, it is also possible to compare contours from different image stacks, such as contours outlining FDG-positive volumes in comparison to volumes which were FMISO-positive. Further research on biologically-adapted radiotherapy treatment planning may profit from these impressions of multimodal image inspection. It is possible to compare anatomic, metabolic and hypoxia imaging directly using a powerful software. Including further imaging modalities such as diffusion weighted MRI appears feasible without a need for changes to the software. Aside from purely visual inspection of image stacks and contours visualising different physiological aspects of the patient and his cancerous disease, it is possible to do statistical analysis of the imaging data: Histograms of subvolumes allow assembling facts which may be important for therapy response monitoring, prognosis and predicton of possible side-effects. Last but not least, the methods for contour comparison, given in section 2.3, are available in the software as well. Thus, measuring volume changes during therapy, differences in metabolic and hypoxic volumes and reproducibility measurements of contours are available within the same software. A screenshot of the software visualising the tools for contouring and contour comparison is shown in Figure 3.12.

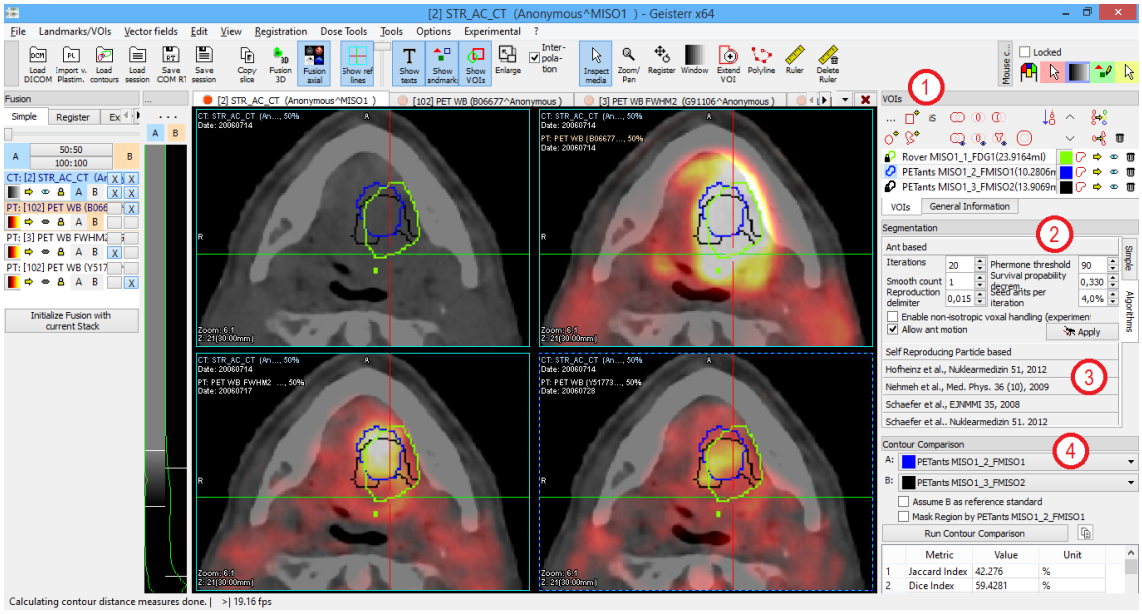


Figure 3.12.: The screenshot of the Geisterr software gives a visual impression of working with image stacks and contours visualising different biological properties. The clinical images in the top row visualise CT and FDG-PET/CT images of a HNC patient before therapy. In the row below, FMISO-PET/CT image fusions are given which show the therapy progress. The left image shows the FMISO tracer distribution before therapy and the image on the right tracer distribution after the first week of combined radiochemotherapy. A list of volumes of interest (1) is available on the right as well as the user interface of the ant-based segmentation algorithm (2). Alternative segmentation algorithms are available (3) as well as tools for contour comparison (4). In this particular case, the operator can see, that the FMISO-positive volumes before and after the first therapy week overlap with a Jaccard index of $42\,\%$.

4. Experiments

4.1. Introduction

Ant algorithms, and most swarm intelligence algorithms in general, suffer from the fact that the performance of the algorithm cannot be proven or it is at least very hard to prove. Furthermore, in application to real-world problems, such as medical image analysis, a prove of the correctness or optimality of the achieved results is impossible, because a gold standard for comparison is mandatory for validation but missing. Alternatively, instead of proving or verifying properties of an algorithm, its validation is standard in clinical research. For this purpose, phantom experiments are performed to demonstrate abilities of an algorithm. The first section of this chapter covers a performed phantom experiment with PET measurements of simulated target objects under conditions with decreasing contrast. This experiment serves as a prove of principle and shows that ant-based delineation brings advantages at low contrast levels compared to standard thresholding techniques. Phantom experiments usually do not simulate clinical patient measurements in every detail. Thus, for further validation of the algorithm, it is worthwhile to apply it to clinical measurements. For this purpose, the second section in this chapter documents the performed steps to optimise the proposed algorithm to be applicable to patient measurements. More precisely this includes the parameter optimisation of the algorithm followed by cross-validation. The third section of this chapter covers a comprehensive analysis of the proposed algorithm combined with the optimised parameter configuration in comparison with other established segmentation algorithms in the field of PET image segmentation. By comparison with manually-created contours from experienced observers, the performance of the different algorithms were reviewed in detail. As a fourth experiment, contour reproducibility on repeated FMISO PET imaging was determined. This experiment was performed to investigate whether observers and algorithms are influenced by local signal changes, which may be a result of noise.

4.2. Ant-based contouring: a phantom study

4.2.1. Introduction

Studying upcoming segmentation algorithms for medical imaging usually starts with phantom experiments. These experiments can be divided into two groups: simulated and measured phantom experiments. Application of the proposed algorithm to simulated phantom data was in focused in published

earlier work (Haase, 2010; Haase et al., 2010). The application to a measured PET phantom experiment was also presented earlier (Haase et al., 2011b) and will be documented in detail in this section. The aim of the experiment was to study how well the pheromone field could be segmented by thresholding in comparison to the original PET data set. Receiver operating characteristic (ROC) analysis was utilised to determine the degree of ability to segment the data set using any threshold. To study this measurement in detail PET phantom data sets with different contrast levels were used to value contrast dependencies.

4.2.2. Materials and methods

The phantom experiment introduced in section 3.1.1 was used to determine the voxel-wise ROC (vwROC) curves of segmentation using thresholding. For analysis, a reference definition of ground truth is needed. Therefore, a binary image stack as shown in Figure 4.1 was provided where positive voxels were white and negative voxels black. This reference image stack was defined by thresholding the initial PET data set with a threshold that lead to a volume definition that matched to the target spheres visible in the corresponding CT data set.

(a) (b) (c)

Figure 4.1.: To allow voxel-wise ROC analysis, a definition of ground truth was derived from the acquired CT image stack (a) and the initial PET data set (b). By thresholding the PET data set with a threshold that leads to a volume definition that matches to the CT-visible spheroidal target volumes, a binary image stack as reference ground truth was derived (c).

To draw vwROC-curves as introduced in section 2.3.4 of the measured PET data set and pheromone fields derived from the ant-based segmentation approach, a free parameter must be defined. For this experiment, the threshold to segment the image stacks was utilised as the free parameter. Thus, a range of thresholds were applied to the PET data set and the corresponding pheromone fields. The thresholds were uniformly distributed between minimum and maximum activity concentration and pheromone intensity, respectively. The fact that the reference ground truth was given as a binary image stack, allowed determination of voxel-wise sensitivity and specificity. To determine vwROC for every single target object, the image space was divided into six cuboid VOIs as shown in Figure 4.2. Furthermore, ant-based pheromone field generation was applied to these six VOIs separately.

The area under the vwROC-curve A_{ROC} serves as the estimation of the degree of segmentability. If $A_{\mathrm{ROC}} \approx 1$, a data point exists, which represents the threshold allowing segmenting the volume of

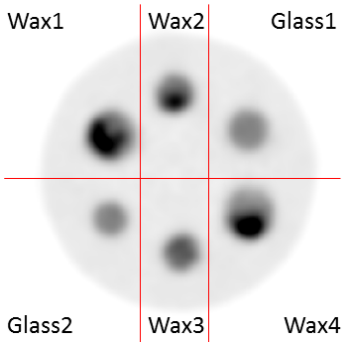


Figure 4.2.: The PET phantom has been divided into six cuboid VOIs, each containing a target sphere, background activity and air in the surrounding volume.

interest in a way that maximises sensitivity $p_{TP} \approx 1$ and specificity $p_{TN} \approx 1$. To plot the resulting A_{ROC} measurements against the contrast between activity concentration in the target volume and the background, two contrast definitions introduced in section 2.2 were used: The target-to-background-ratio derived from mean activity concentration C in target and background

$$TBR = \frac{C_{\text{target}}}{C_{\text{background}}} \tag{4.1}$$

and the contrast-to-noise-ratio:

$$CNR = \frac{C_{\text{target}} - C_{\text{background}}}{\sigma_{\text{background}}} \tag{4.2}$$

Figure 4.3 visualises from which VOIs the mean activity concentration measurements were derived from to determine TBR and CNR. For the measurement of mean activity concentration within the spheres a margin within the spheres was left out to minimise the effect of spill-in to the measurements. It can be assumed that there is a relationship between contrast in the PET image stack and A_{ROC}. As also visible in Figure 4.3, the contrast between target spheres and background obviously fall with time during the measurements. Thus, the degree of applicability of thresholding to segment the spheres decreases with falling contrast: at a certain level of contrast, target volume and background are not differentiable by a threshold.

Before application of the segmentation algorithm, a vwROC curve for the inspected VOI of the PET volume was drawn and the corresponding A_{ROC} value was determined. During application of the ant-based algorithm to the PET image stacks, vwROC curves of the pheromone field were drawn after every iteration. Corresponding A_{ROC} values were collected. After 25 iterations, the pheromone field with the highest A_{ROC} value was determined and the corresponding value further processed. Furthermore, ant-distributions for worker and scout ants and pheromone fields of every iteration were stored for later analysis.

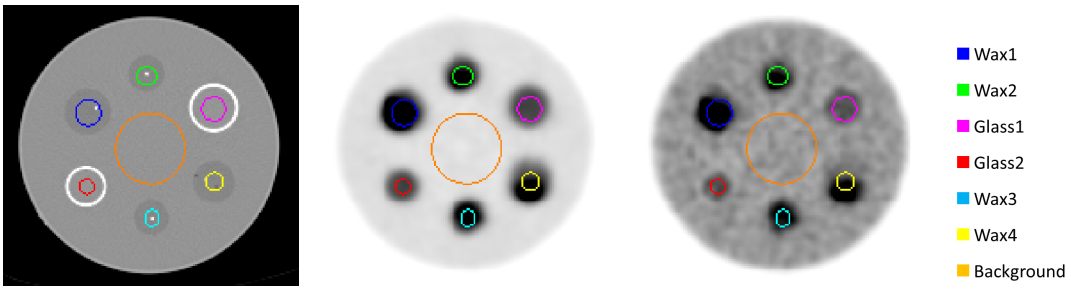


Figure 4.3.: The coloured spheroidal volumes were defined to measure mean activity concentration as basis for TBR and CNR calculation. The VOIs were placed on a fused view of CT (a) and PET (b,c) to place them near the centre of the object with a margin to the real sphere boundary. This margin was introduced to reduce the effect of spill-in to the measurement of the mean activity concentration.

4.2.3. Results

Application of the ant-based algorithm to PET image stacks under different contrast levels lead to pheromone fields with different contrast levels. Three example pheromone fields with corresponding ROC curves are shown in Figure 4.4. It is obvious that the contrast between target volume and background was higher in the pheromone field compared to the PET image stack. But this was only observed, as long as there is enough contrast between target and background. When the contrast fall below a certain level, the ants were no longer able to differentiate the target from the background and thus emit pheromone to target and background. In the shown example with the lowest contrast, the ants were only able to differentiate between target and glass walls.

Plotting the resulting A_{ROC} values against the TBR measurements lead to the diagram shown in Figure 4.5. As long as $TBR > 3$ in any target sphere, the A_{ROC} is approximately 1 pointing to the fact that for the original PET image stack as well as for the pheromone field a threshold exists that allowed to distinguish between target and background volume. Thus, theoretically it was possible to segment the given PET image stacks resulting in a volume definition with sensitivity and specificity of about $100\,\%$. However, at latest when TBR falls below 2, A_{ROC} decreases when thresholding the PET image stack. The A_{ROC} measurements of the corresponding pheromone fields were higher in all tested cases when $TBR < 3$. Thus, the pheromone field could be segmented with higher accuracy. The glass spheres and wax sphere 3 were still thresholdable with $A_{\text{ROC}} \approx 1$ if $TBR \approx 1.5$. The other wax spheres lead to lower to A_{ROC} measurements. This may be a result of heterogeneous activity distribution observed in these spheres.

Plots of the resulting A_{ROC} values against CNR are given in Figure 4.6. The plots visualise a relationship similar to the relationship between A_{ROC} and TBR. Thus, if $CNR > 25$ the PET image stacks as well as the corresponding pheromone fields could be segmented accurately using a specific threshold. The A_{ROC} measurements from the pheromone fields are again higher compared to the PET image stacks, when $CNR < 25$. However, when CNR fall below 5, A_{ROC} decreases dramatically.

Finally, processing the PET image to a pheromone field can be seen as contrast enhancement. Thus,

further processing of higher contrasted pheromone fields has obviously advantages for delineation compared to processing the original PET image.

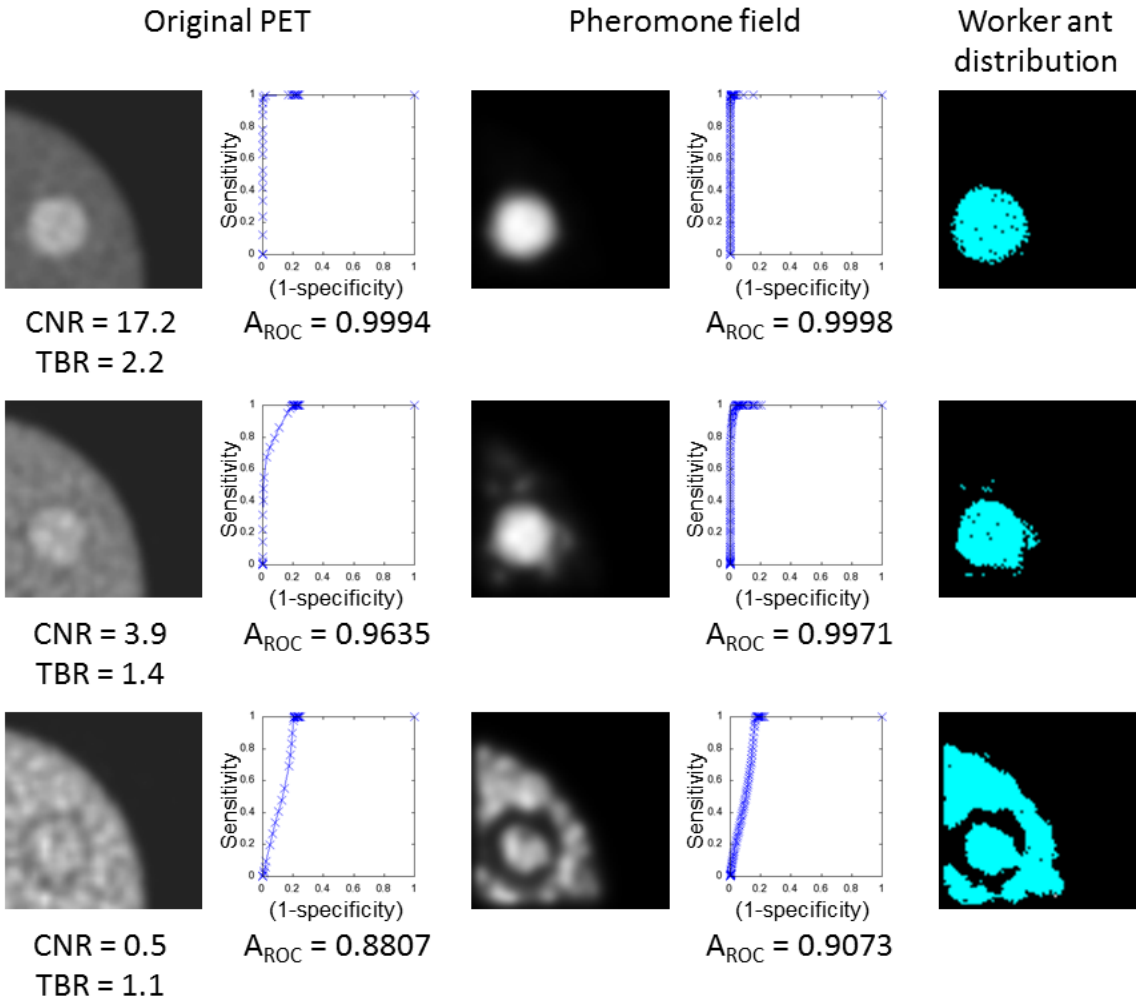
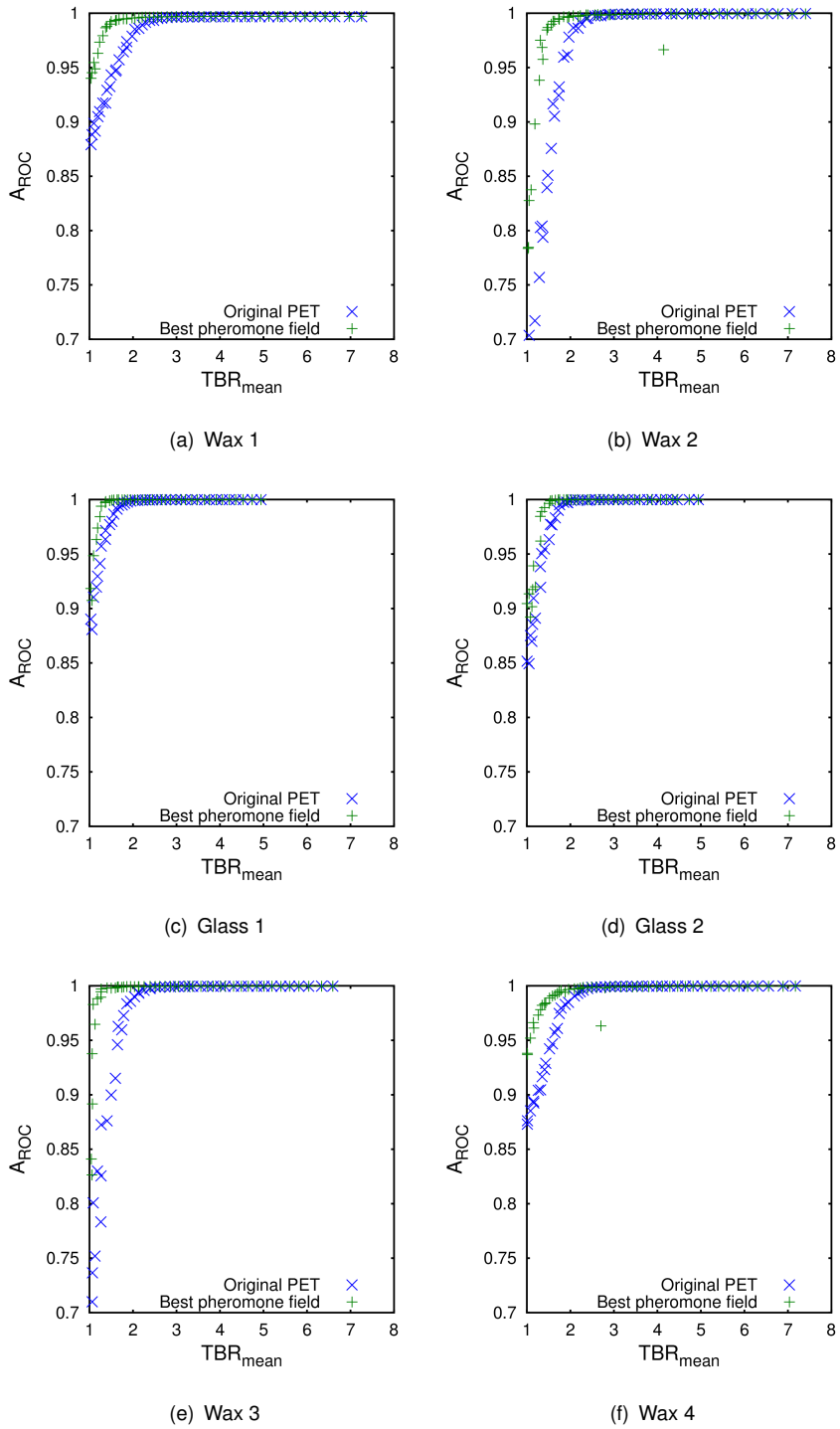


Figure 4.4.: An image representing glass sphere 1 at different contrast levels shows that the area A_{ROC} under the ROC curve decreases with falling contrast. This effect was also present - but retarded - in the corresponding pheromone fields. The ant-based approach appeared to increase contrast from the original PET image stack to the pheromone field as long as the contrast was high enough. The last example data set (bottom line) shows a case where neither for the PET image stack nor for the pheromone field a threshold existed which allowed differentiation between target and background.



(a) Wax 1 (b) Wax 2

(c) Glass 1 (d) Glass 2

(e) Wax 3 (f) Wax 4

Figure 4.5.: The resulting A_{ROC} values were plotted against the measured TBR values. When TBR falls below 2, the A_{ROC} resulting from the PET image stacks decreased. The same effect was observed for the pheromone fields but at a lower TBR of about 1.5. The single outliers in b) and f) illustrate that the results of the algorithm are not deterministic. Double-checking contours by repeatedly executing the algorithm is worthwhile.

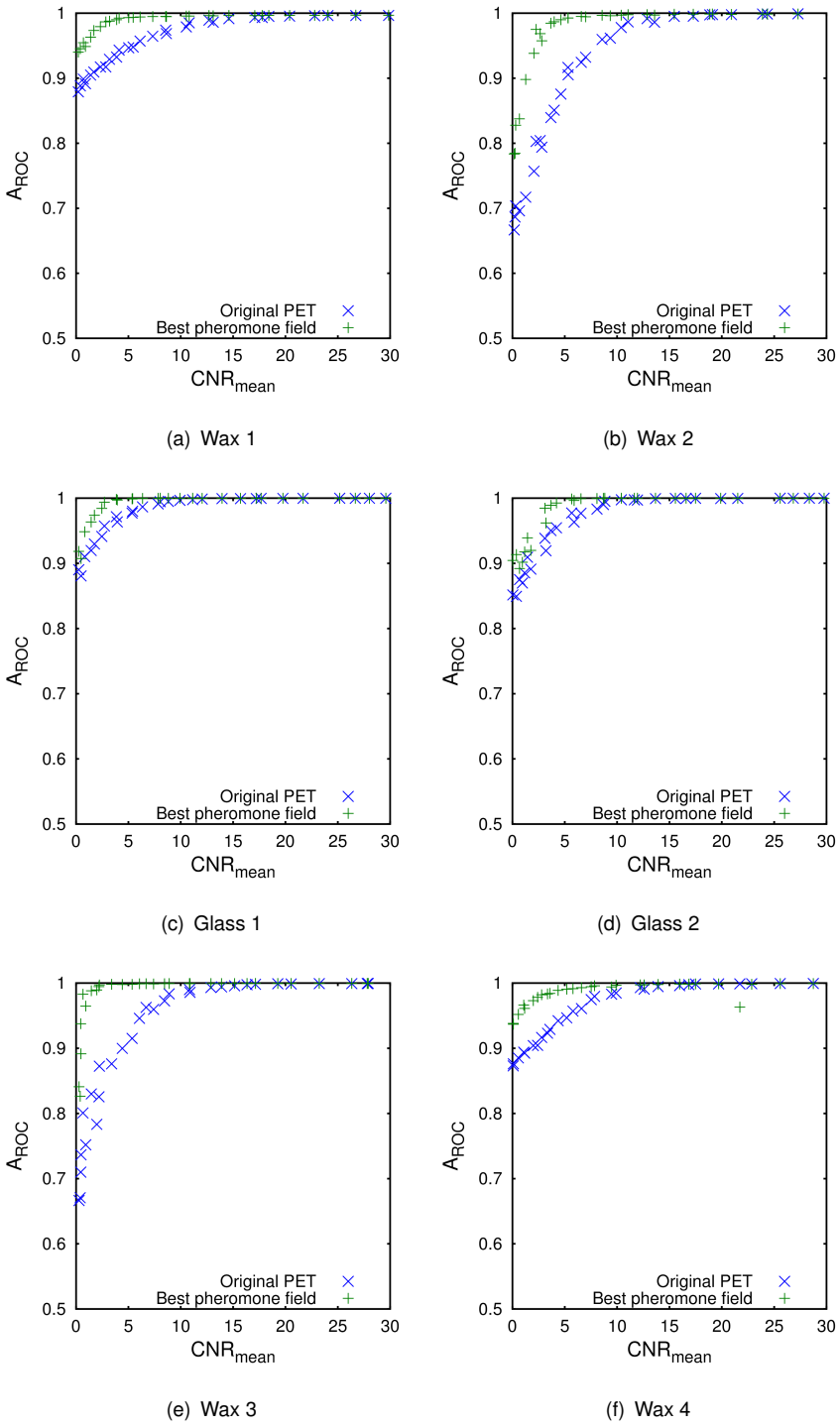


Figure 4.6.: The resulting A_{ROC} values were plotted against the measured CNR values. As shown for TBR, when CNR fall below 25, the A_{ROC} resulting from the PET image stacks decreased. When CNR fall below 5, representing a noise level of $20\,\%$, an accurate segmentation became impossible using the PET image stack as well as the pheromone field. The shown graphs do not visualise the whole space of measured values. The x-axis has been cut, because the CNR values ranged up to about 120 but the corresponding A_{ROC} did not deliver further insights.

4.3. Calibration and optimisation of the proposed algorithm

4.3.1. Introduction

Segmentation algorithms applied in PET imaging deserve calibration (Nehmeh et al., 2009; Schaefer et al., 2012). However, building a calibration PET phantom for low contrast conditions as observed in FMISO PET is complicated. On the one hand, glass spheres disturb accurate delineation if TBR is low (Hofheinz et al., 2010). On the other hand, simulation of target volumes using radioactive wax spheres without inactive walls as proposed by (Bazañez-Borgert et al., 2008) is challenging, too (Haase et al., 2010). Thus, calibrating the proposed algorithm using clinical patient data sets and manually-created contours from experienced observers is worthwhile. As shown in section 3.2.4, not all parameters of the algorithm are worth considering for calibration. But the threshold used for segmenting the pheromone field obviously is crucial for accurate delineation. Thus, determination of the optimal threshold will be in focus of this section. The concept of cross-validation will afterwards reveal if the optimal threshold sustains and allows application of the algorithm to a group of clinical PET data sets which were not used for optimisation.

4.3.2. Materials and methods

Before applying the threshold to the pheromone field, the field was normalised to an integer in the range $[0; 255]$ representing the range between minimum and maximum pheromone intensity. Thus, the optimal threshold $\tau_{\text{min,opt}}$ must be in the same range. To test various thresholds τ_{min} all 155 clinical FMISO PET data sets from the 43 patients introduced in section 3.1.2 were processed by the proposed algorithm. Afterwards, the resulting pheromone fields were segmented by 17 thresholds $\tau_{\text{min,i}} = 10, 20, ..., 170$ used to create 17 segments per PET data set. Those segments were compared to the segments manually-created by the three experienced observers introduced in section 3.1.6. This comparison was only applied to data sets, where all three observers concordantly outlined any target volume. As criterion for the contour quality, the Jaccard index and the bilateral mean contour distance as introduced in sections 2.3.5 and 2.3.7 were utilised. For the figure of merit Jaccard index, the optimal value is the maximum, for the bilateral mean contour distance, the metric was minimised. The contour comparison was applied to automatically-created contours versus manually-created contours and manually-created contours versus each other. For every pheromone threshold $\tau_{\text{min,i}}$, the mean Jaccard index J_i and bilateral mean contour distance $d_{\text{bil,c,i}}$ were determined. To get a detailed view on relationships between automatically-created contours and contours from individual observers, the figures of merit were averaged for every single observer as well as for all observers together. This experiment was applied to determine whether the optimal tested threshold τ_i resulted in a figure of merit value which a) was better than the reference value taken from the pair-wise comparison of manually-created contours and b) was still applicable if the optimal threshold was applied to another group of data sets that were not taken into account for determination of the optimum. For this cross-validation procedure, the patient population was separated into two groups: The test cohort containing the first 25 patients included in the study and the validation cohort containing the remaining 18 patients. From the first 25 patients, target volumes in 59 FMISO PET data

sets were outlined concordantly by the observers. Rather, all observers outlined any object in the vicinity of the primary tumour. The validation cohort contained 35 PET data sets with contours applicable for this analysis.

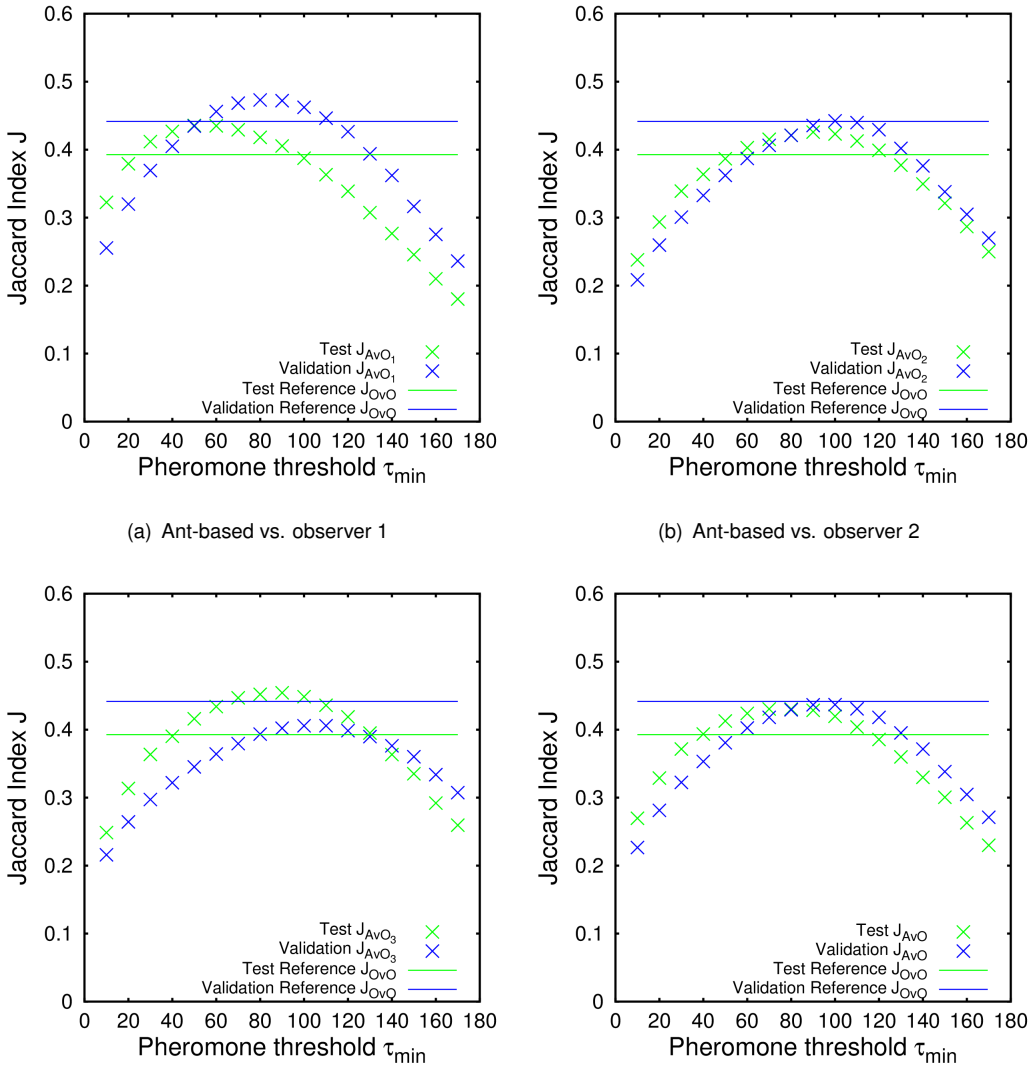
After determination of the optimal threshold, resulting Jaccard index and contour distance measurements of observer-observer- (OvO) and algorithm-observer-comparisons (AvO) were analysed to determine if the differences were significant. A p-value < 0.05 resulting from the paired t-test was defined as criterion for significance.

4.3.3. Results

The reference contours determined by the observers were compared with each other and resulted in $J_{\text{OvO}} = 0.39 \pm 0.11$ for the test group and $J_{\text{OvO}} = 0.44 \pm 0.14$ for the validation group. The corresponding measurements for the contour distance were $d_{\text{OvO}} = 5.0 \pm 1.6$ mm for the test group and $d_{\text{OvO}} = 3.8 \pm 1.3$ mm for the validation group. Thus, the inter-observer-variability of manually-created contours is comparable in both groups.

An overview on Jaccard index measurements of the comparison for every single observer and all observers averaged is given in Figure 4.7. It is obvious that for the test cohort there was always an optimal pheromone threshold τ_{min} leading to a maximum J_{AvO} measurement which lied above the reference value J_{OvO} resulting from pair-wise comparison of manually-created contours. Optimal thresholds and corresponding Jaccard index values as well as contour distance measurements are given in Table 4.1. However, in the validation cohort the mean J_{AvO} was lower than J_{OvO}, indicating that the inter-observer-differences were lower than differences between automatically- and manually-created contours. Furthermore, the optimum threshold varied between 50 and 90 depending on which observer served as reference. If all the Jaccard index measurements from the comparison with manuallay-created contours were averaged, the optimum threshold $\tau_{\text{min}} = 70$ resulted in $J_{\text{AvO}} = 0.43 \pm 0.14$ above the reference $J_{\text{OvO}} = 0.39 \pm 0.11$. The corresponding value from the validation cohort was $J_{AvO} = 0.42 \pm 0.13$ and also lied below the reference value $J_{\text{OvO}} = 0.44 \pm 0.14$. The observed small differences between algorithm-observer-comparison and observer-observer-comparison became more obvious, when error bars were added as shown in Figure 4.8. In the other plots, error bars were removed to increase diagram interpretability.

The respective analysis of the bilateral mean contour distance, displayed in Figure 4.9, revealed an alternative view on the contour differences. First of all, the contour distance between contours from observer 1 and automatically-generated contours were more distant than the manually-created contours to each other. For the other two observers, the algorithm was able to create contours which were less distant to their contours than the inter-observer-variability between all three observers. Thus, in the average case, the automatically-created contours which were more similar to the contours of observer 2 and 3 than to observer 1. Furthermore, the optimal threshold in the test group for matching contours from observers 2 and 3 $\tau_{\text{opt,min}} = 100$ while for observer 1 the optimal threshold was $\tau_{\text{opt,min}} = 70$. Finally, in the validation cohort none of the determined optimal thresholds τ_{min} obviously sustained. In all



(a) Ant-based vs. observer 1

(b) Ant-based vs. observer 2

(c) Ant-based vs. observer 3

(d) Ant-based vs. mean observers

Figure 4.7.: For determination of an optimal pheromone threshold τ_{min} for the ant-based algorithm, various τ_{min} values have been tested. The resulting Jaccard index measurements from comparisons with manually-created contours suggest different thresholds to simulate manually-created contours depending on which observer shall be simulated. The horizontal lines represent reference Jaccard index measurement resulting from determination of inter-observer-variability. In all cases, the optimal τ_{min} value in the test cohort was higher than the reference Jaccard index. However the determined thresholds were not applicable to the validation cohort to outperform the reference.

cases, the contour distance between automatically- and manually-created contours was higher than the reference value $d_{bil,c} = 3.8\,\text{mm}$.

The fact that differences between observer-observer-comparisons and algorithm-observer-comparisons were smaller than the standard deviation within the test and validation cohort, suggests that the differences may not be significant. Furthermore, the optimal threshold depended on the chosen reference. As a compromise between the different optima, $\tau_{min} = 90$ was defined as further used pheromone threshold. After application of this threshold for volume determination, the differences between observer-

Table 4.1.: Optimal pheromone thresholds and corresponding Jaccard index and bilateral contour distance values in test and validation group depending on used reference

Reference	$\tau_{\text{min,opt}}$	J_{test}	**>ref**	$J_{\text{validation}}$	**>ref**
Observer 1	50	0.44 ± 0.17	yes	0.43 ± 0.14	no
Observer 2	90	0.43 ± 0.15	yes	0.44 ± 0.15	no
Observer 3	90	0.45 ± 0.14	yes	0.40 ± 0.16	no
Observers averaged	70	0.43 ± 0.14	yes	0.42 ± 0.13	no

Reference	$\tau_{\text{min,opt}}$	$d_{\text{bil,c,test}}$ / **mm**	**< ref**	$d_{\text{bil,c,validation}}$ / **mm**	**< ref**
Observer 1	70	5.5 ± 2.2	no	4.4 ± 1.8	no
Observer 2	100	4.5 ± 1.8	yes	4.2 ± 2.1	no
Observer 3	100	4.0 ± 1.3	yes	4.3 ± 1.9	no
Observers averaged	100	4.7 ± 1.97	yes	4.3 ± 1.9	no

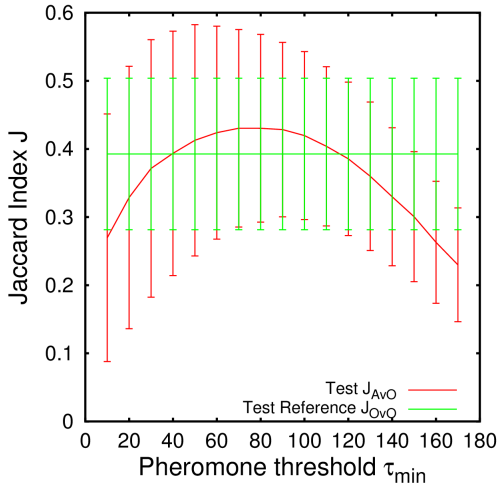
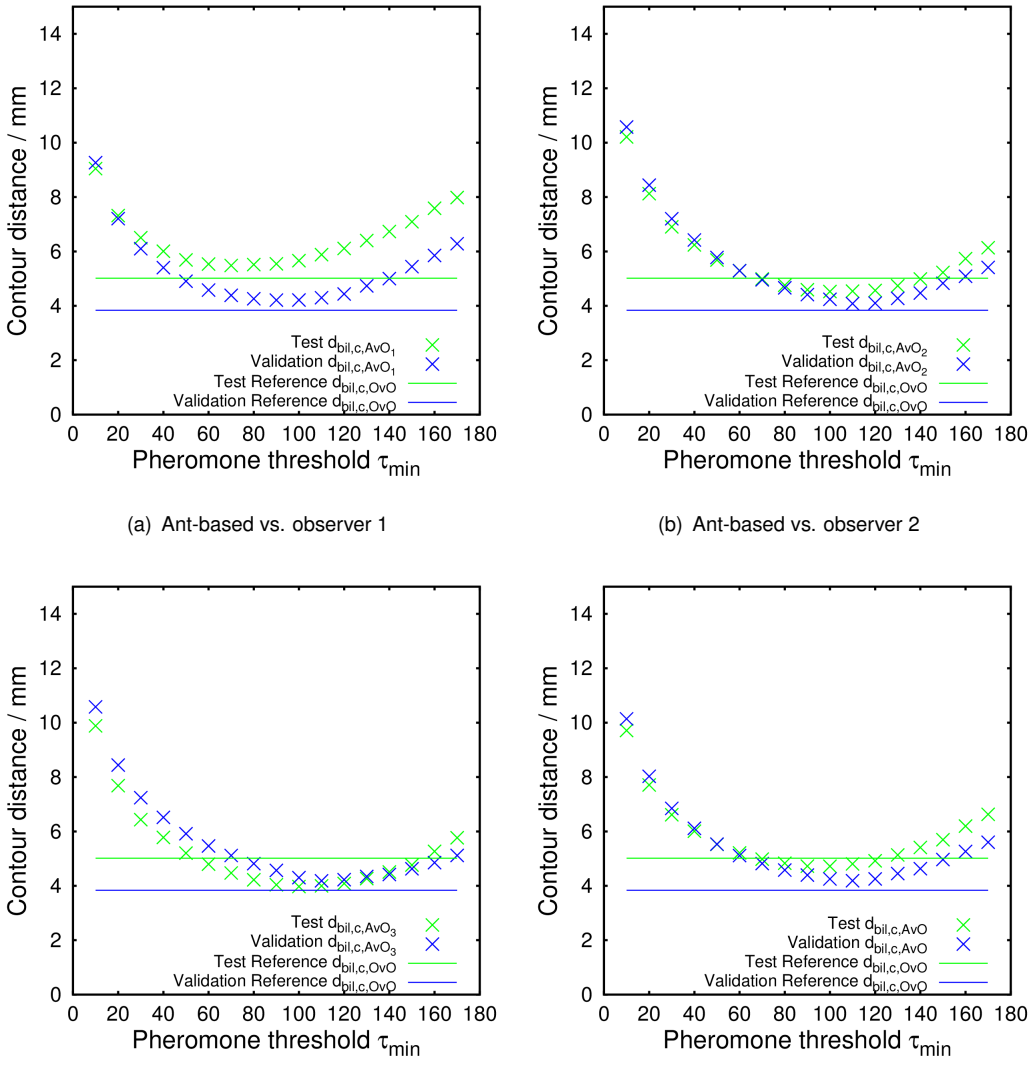


Figure 4.8.: The application of different τ_{min} values leads to Jaccard index values with varying error. The error bars visualise the range of Jaccard index measurements determined in the whole test group of the contouring study.

observer-comparison and algorithm-observer-comparison were analysed to determine if differences were significant. The paired t-test revealed a p-value of 0.39 for Jaccard index and 0.19 for the contour distance metric. Thus, differences were not significant.



(a) Ant-based vs. observer 1

(b) Ant-based vs. observer 2

(c) Ant-based vs. observer 3

(d) Ant-based vs. mean observers

Figure 4.9.: Determination of optimal τ_{min} values using the bilateral mean contour distance revealed that in no case the algorithm clearly outperformed the observers. For observer 1 and all observers averaged the optimal τ_{min} value appeared applicable to some degree, even though the reference values displayed as horizontal lines were not exceeded.

4.4. Comparison with other algorithms

4.4.1. Introduction

After the principle of the ant-based segmentation algorithm has been proven feasible and the algorithm has been optimised to handle clinical data sets from head-and-neck-cancer patients, it must be compared to established standard PET segmentation algorithms. For this purpose, the algorithms introduced in section 2.5.2 as well as the ant-based approach were compared to the contours drawn by the experienced observers introduced in section 3.1.6. Aim of the comparison was to determine whether the new algorithm creates contours comparable to the experienced observers and if it outperforms the established

algorithms in accurate volume and target-to-muscle-ratio estimation. These two properties were chosen, because these parameters may be relevant for patient prognosis determination (Zips et al., 2012).

4.4.2. Materials and methods

The applied segmentation algorithms and corresponding abbreviations are given in Table 4.2. All algorithms were applied to all 155 FMISO PET data sets introduced in section 3.1.2.

Table 4.2.: Applied segmentation algorithms and used software packages

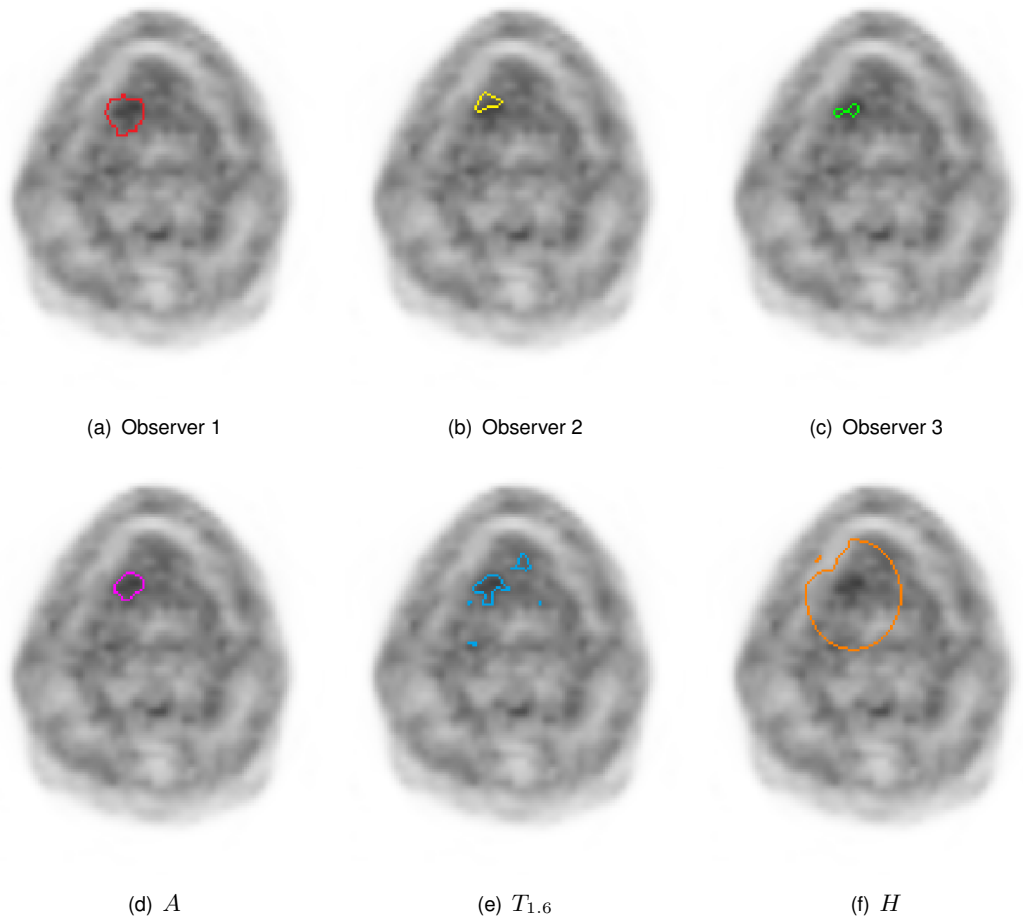
Contour source	Applied software	Contour abbreviation
Ant-based delineation	Geisterr (in-house implementation)	A
Hofheinz et al. (2012c)	Rover (ABX, Radeberg, Germany)	H
Nehmeh et al. (2009)	in-house implementation	N
Schaefer et al. (2008)	Geisterr (in-house implementation)	S_{2008}
Schaefer et al. (2012)	Geisterr (in-house implementation)	S_{2012}
1.2 times background	Geisterr (in-house implementation)	$T_{1.2}$
1.4 times background	Geisterr (in-house implementation)	$T_{1.4}$
1.6 times background	Geisterr (in-house implementation)	$T_{1.6}$
1.8 times background	Geisterr (in-house implementation)	$T_{1.8}$
2.0 times background	Geisterr (in-house implementation)	$T_{2.0}$
Observer 1	Syngo (Siemens, Forchheim, Germany)	O_1
Observer 2	Syngo (Siemens, Forchheim, Germany)	O_2
Observer 3	Syngo (Siemens, Forchheim, Germany)	O_3

Furthermore, the contours drawn by the experienced observers, as introduced in section 3.1.6, were used as reference standard to determine errors in volume V and mean target-to-muscle-ratio TMR estimation. While the determination of TMR_{max} is trivial if a rough volume definition such as a VOI is given, determination of the mean ratio TMR is challenging and depends on a reliable segmentation algorithm. V and TMR were also analysed, because both did not play a role in the optimisation of the algorithm presented in the previous section. For error analysis, the relative and absolute volume and TMR errors were determined. For statistical analysis of differences in volume and TMR estimation, the paired t-test was applied. A p-value of 0.05 was defined as criterion for significant differences.

4.4.3. Results

As an overview on the derived segments respectively contours, Table 4.3 contains derived volume measurements from the observers in comparison to the volume measurements delivered by the applied algorithms. Corresponding TMR measurements are given in Table 4.4. Both tables only contain data sets for which all observers outlined a target object within the defined VOI. Thus, only 94 of 155 PET data sets were taken into account. From the remaining 61 data sets, 29 were not outlined by any observer, in 24 data sets a target volume was defined by two observers and in 8 cases only a single observer delineated a target volume. To allow a reliable comparison of contour-pairs, only those cases were analysed, for which all three contours from the experienced observers were available.

Firstly, not all algorithms delivered contours for all PET data sets. Thus the number n of available contours for comparison shall be emphasised. The algorithms H, $T_{1.8}$ and $T_{2.0}$ did not deliver contours at all in about $20\,\%$ of the cases which were marked FMISO-positive concordantly by the observers. Secondly, the differences between volume measurements of observers and algorithms were significant in all cases except for A and $T_{1.6}$. The H algorithm also showed non-significant differences, but only in TMR estimation. Thus, only these three algorithms appeared able to deliver volume estimations that may be related to volume measurements done by the observers. Differences in volume estimation varied by about 20 to $50\,$ml depending on which pair of observers or observer-algorithm pair was analysed. The increased mean volume differences between observers and algorithm H were obvious. The reason could be that in some cases, the H algorithm delivered a volume definition that was very similar to the VOI definition. An example for such a case is shown in Figure 4.10. Thus, for a detailed inspection of the volume differences between observers and algorithms, Bland-Altman plots for pair-wise volume comparison of the three observers and comparisons of observers with these three algorithms are given in Figure 4.11. The volume errors appear not very systematic, the only obvious and reasonable point was that volume errors were smaller when delineating smaller volumes.



(a) Observer 1 (b) Observer 2 (c) Observer 3

(d) A (e) $T_{1.6}$ (f) H

Figure 4.10.: Example contours from a patient after therapy week 4 outline a small FMISO-positive volume. Only algorithm H was not able to distinguish between target volume and background. In this case H outlined almost the whole volume of interest as FMISO-positive.

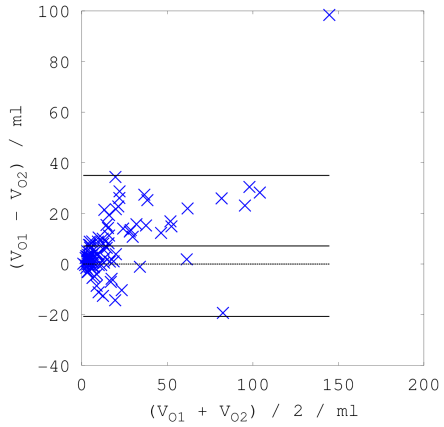
Table 4.3.: Volume measurements of tested segmentation algorithms in comparison

Algorithm	n	$V_{\text{algorithm}}/\,\text{ml}$	$V_O/\,\text{ml}$	p-**value**	$\Delta V/\,\text{ml}$	$\delta V/\%$
A	94	16.5 ± 15.4	18.4 ± 21.0	0.480	-1.9 ± 14.7	43 ± 119
H	79	45.1 ± 76.0	19.9 ± 22.2	0.005	25.2 ± 76.4	612 ± 1597
N	94	121.0 ± 112.9	18.4 ± 21.0	0.000	102.6 ± 108.3	1403 ± 1927
S_{2008}	94	45.0 ± 33.0	18.4 ± 21.0	0.000	26.6 ± 25.7	369 ± 505
S_{2012}	94	79.8 ± 74.9	18.4 ± 21.0	0.000	61.4 ± 71.8	950 ± 1500
$T_{1.2}$	94	71.3 ± 50.1	18.4 ± 21.0	0.000	53.0 ± 37.1	483 ± 509
$T_{1.4}$	94	37.1 ± 32.2	18.4 ± 21.0	0.000	18.7 ± 18.2	147 ± 207
$T_{1.6}$	93	20.3 ± 22.1	18.5 ± 21.0	0.572	1.8 ± 11.6	8 ± 75
$T_{1.8}$	82	12.6 ± 15.8	20.4 ± 21.7	0.010	-7.8 ± 14.9	-43 ± 41
$T_{2.0}$	72	7.7 ± 11.3	22.4 ± 22.4	0.000	-14.7 ± 19.0	-69 ± 29

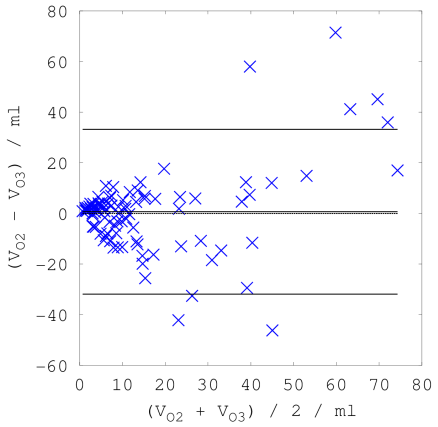
Table 4.4.: TMR_A measurements of the respective algorithms and mean TMR_O measurements of the experienced observers of tested segmentation algorithms in comparison

Algorithm	n	TMR_A	TMR_O	p-value	ΔTMR	δTMR
A	94	1.86 ± 0.49	1.80 ± 0.30	0.312	0.06 ± 0.22	2 ± 9
H	79	1.88 ± 0.66	1.81 ± 0.31	0.377	0.07 ± 0.40	1 ± 21
N	94	1.48 ± 0.52	1.80 ± 0.30	0.000	-0.32 ± 0.30	-19 ± 15
S_{2008}	94	1.64 ± 0.39	1.80 ± 0.30	0.002	-0.16 ± 0.15	-9 ± 7
S_{2012}	94	1.64 ± 0.39	1.80 ± 0.30	0.002	-0.16 ± 0.15	-9 ± 7
$T_{1.2}$	94	1.47 ± 0.14	1.80 ± 0.30	0.000	-0.33 ± 0.19	-17 ± 7
$T_{1.4}$	94	1.66 ± 0.16	1.80 ± 0.30	0.000	-0.14 ± 0.16	-6 ± 7
$T_{1.6}$	93	1.83 ± 0.18	1.80 ± 0.30	0.508	0.02 ± 0.14	2 ± 7
$T_{1.8}$	82	2.01 ± 0.19	1.85 ± 0.28	0.000	0.16 ± 0.12	9 ± 7
$T_{2.0}$	72	2.20 ± 0.19	1.89 ± 0.27	0.000	0.30 ± 0.12	16 ± 7

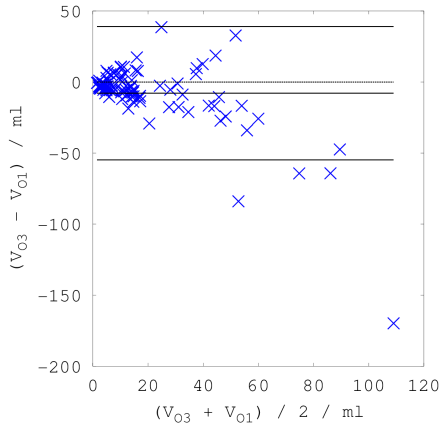
Also TMR measurements were significantly different except for A, H and $T_{1.6}$. Corresponding Bland-Altman plots showing the differences in TMR estimation, given in Figure 4.12, reveal systematic errors for all three algorithms with different markedness. The H algorithm underestimated TMR in most data sets with $TMR < 1.7$ and overestimated most of the remaining data sets. The same observation is true but less intense for the ant-based approach A. The thresholding algorithm $T_{1.6}$ showed a reverse systematic error: For target volumes with $TMR < 1.75$ the TMR was overestimated by tendency, the remaining data sets were underestimated in many cases. The fact that the thresholding algorithm $T_{1.6}$ was not able to determine TMR correctly if TMR is low appears logical when regarding its implementation. The algorithm defines voxels with $TMR < 1.6$ as negative. Thus, the mean average of all voxels enclosed by this thresholding technique cannot be below 1.6. To further emphasise the consequences for data sets with $TMR < 1.6$ as determined by the observers, the data set was restricted. Figure 4.13 shows differences in volume and TMR estimation by the algorithms A and $T_{1.6}$ for the 23 data sets where observers outlined volumes with $TMR < 1.6$. In these low-contrast PET images, ant-based delineation broadly showed smaller errors than the thresholding algorithm $T_{1.6}$.



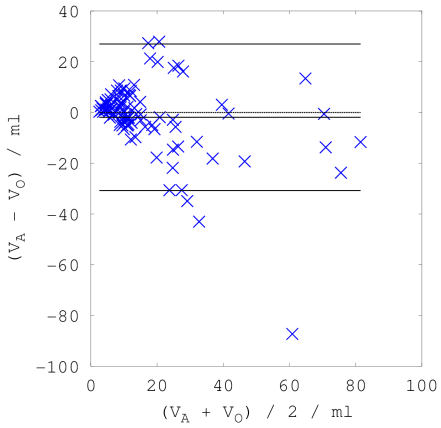
(a) Observers 1 vs. 2



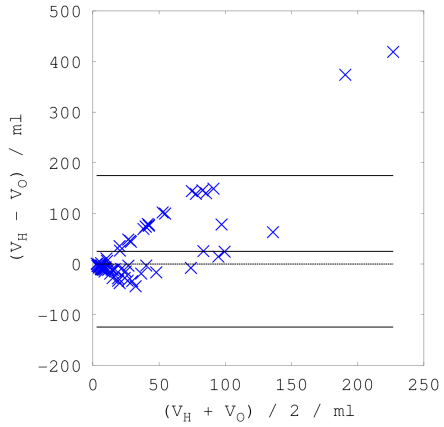
(b) Observers 2 vs. 3



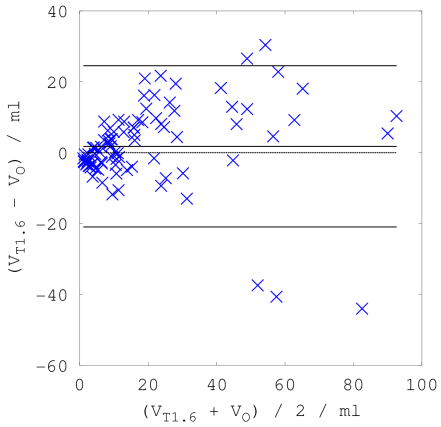
(c) Observers 3 vs. 1



(d) Ant-based vs. averaged observers

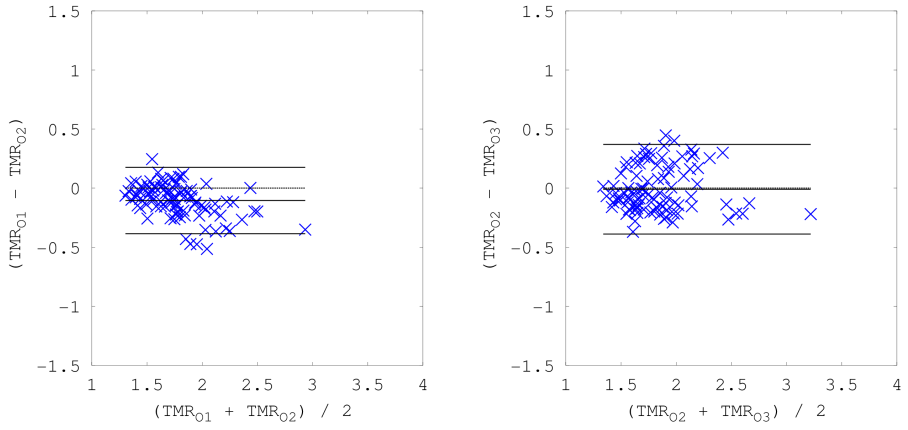


(e) (Hofheinz et al., 2012c) vs. averaged observers



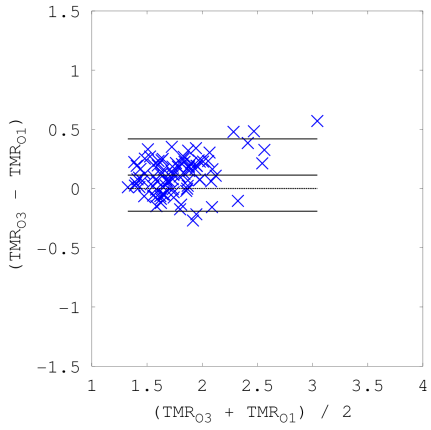
(f) Threshold $T_{1.6}$ vs. averaged observers

Figure 4.11.: Bland-Altman plots of observers volume determinations pair-wise (a-c) and comparison between observers and algorithms (d-f) did not reveal obvious systematic errors, but that larger volumes may lead to larger volume errors. The differences between algorithm H and observers were measurably higher than in the other five comparisons.

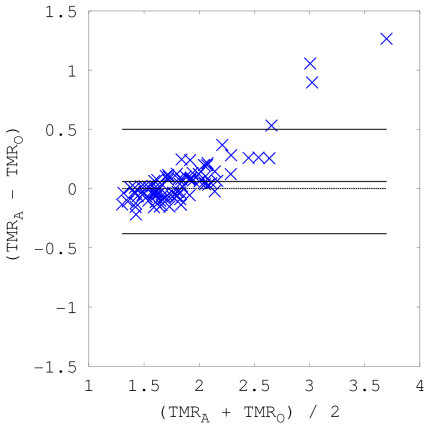


(a) Observers 1 vs. 2

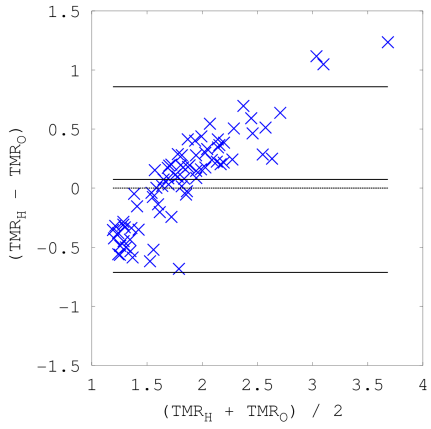
(b) Observers 2 vs. 3



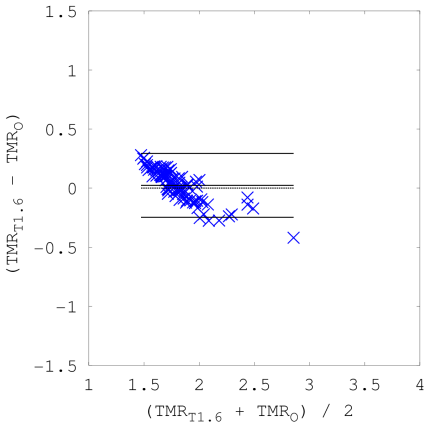
(c) Observers 3 vs. 1



(d) A vs. averaged observers

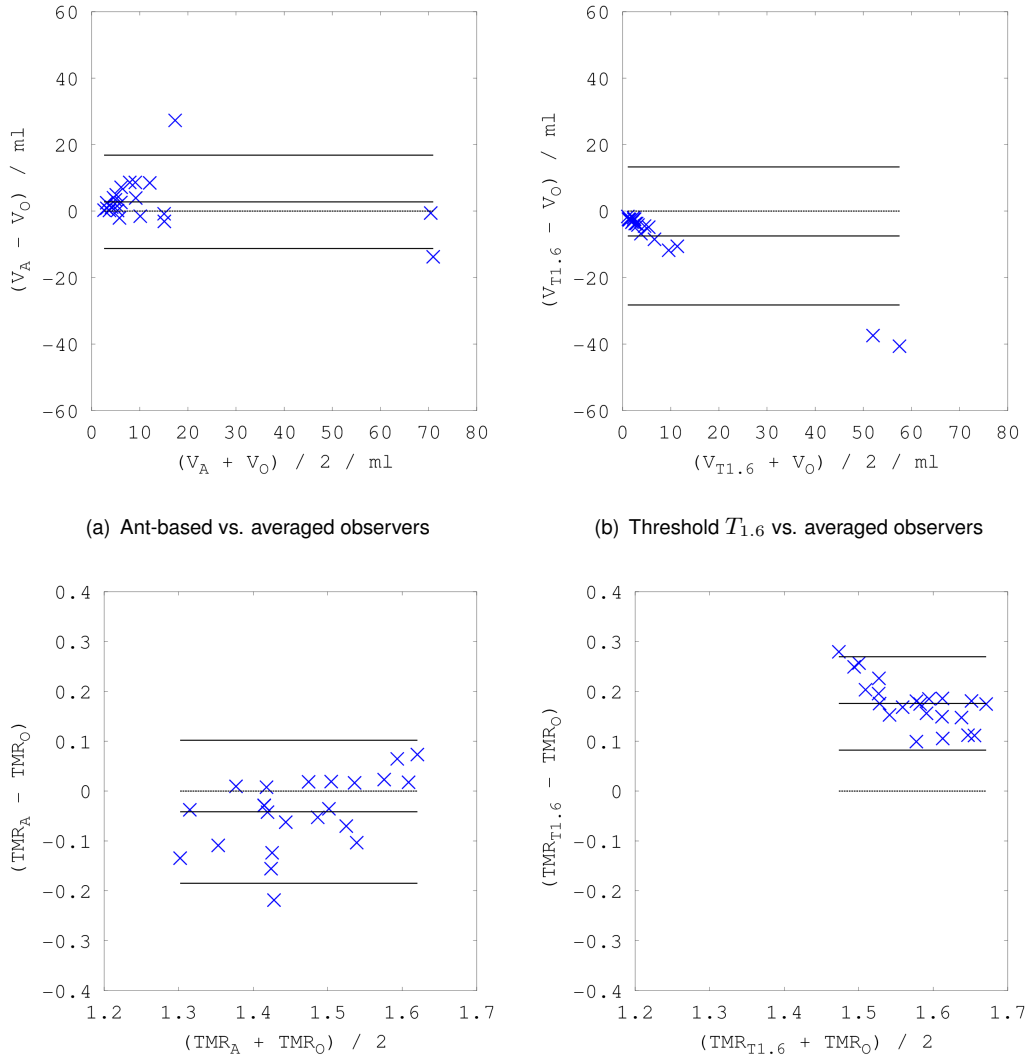


(e) H vs. averaged observers



(f) $T_{1.6}$ vs. averaged observers

Figure 4.12.: Bland-Altman plots of observers volume determinations pair-wise (a-c) do not show a systematic error while the comparisons between observers and algorithms (d-f) do: The algorithms A and H show a systematic error in TMR estimation which appears to be related with TMR itself. The thresholding algorithm $T_{1.6}$ shows a similar but reverse systematic error.



(a) Ant-based vs. averaged observers

(b) Threshold $T_{1.6}$ vs. averaged observers

(c) Ant-based vs. averaged observers

(d) Threshold $T_{1.6}$ vs. averaged observers

Figure 4.13.: If the volume and TMR difference analysis is limited to data sets where the observers estimated $TMR < 1.6$, the thresholding algorithm $T_{1.6}$ leads to volume underestimation and TMR overestimation in all cases. Ant-based delineation was more accurate in these low-contrast data sets.

4.5. Contour reproducibility in subsequent FMISO PET imaging

4.5.1. Introduction

Research projects in the field of hypoxia imaging using FMISO PET must face the controversy on FMISO PET reproducibility present in the scientific literature. Still, it is not finally clear if FMISO PET measurements are reproducible. In one study high reproducibility was only measured in a half of the investigated patient data sets (Nehmeh et al., 2008). In another study, high reproducibility of FMISO PET imaging derived parameters in all cases of eleven investigated patients was observed (Okamoto et al., 2013). Both

investigations relied partly on measuring the voxel-wise Pearson correlation coefficient, even though it may not be applicable for method-comparison studies (Westgard and Hunt, 1973; Altman and Bland, 1983). Alternatively to voxel-wise comparison of signal intensities contour comparison may be used to determine the reliability of FMISO PET imaging (Haase et al., 2013b). The experiment in this section is to estimate reproducibility of FMISO PET based contours in subsequent FMISO PET imaging. Further emphasis is taken on differences between manually- and automatically-created contours.

4.5.2. Motivation and related work

In the discussion about FMISO PET reproducibility most prominently Nehmeh et al. are cited with a report showing FMISO PET is of limited reproducibility. The authors compared subsequent FMISO PET scans of patients with head and neck carcinomas acquired within 2-4 days and found strong voxel-wise correlation only in 6 of 13 patients. Thus, the conclusion was that tracer accumulation was not stable in the other cases (Nehmeh et al., 2008). Okamoto et al. did a comparable analysis with the result that FMISO uptake was reproducible (Okamoto et al., 2013). However, Schwartz et al. showed that the use of the Pearson correlation coefficient may not be suitable for this kind of investigation (Schwartz et al., 2011). About four decades ago, Westgard and Hunt stated “The correlation coefficient [...] is of no practical use in the statistical analysis of comparison data.” (Westgard and Hunt, 1973). This fact is still not present in the scientific community encouraging Dr. Westgard in republishing his view on the field of statistical method-comparison repeatedly (Westgard, 1998, 2008). To investigate the applicability of Pearson correlation coefficient (PCC) measurements for voxel-wise PET analysis, a phantom experiment published earlier (Haase et al., 2012b) and introduced in section 3.1.1 shall shortly be reviewed from another perspective. Two subsequent measurements of a cylinder containing a solution with $[^{18}\text{F}]$-fluorodeoxyglucose as shown in Figure 4.14 were compared to determine Pearsons' coefficient r. Within the analysed region of interest, activity concentration $C_{PET1} = 5450 \pm 296\,\text{Bq}/\text{ml}$ and $C_{PET2} = 5427 \pm 322\,\text{Bq}/\text{ml}$ were measured in the two images, respectively. A noise level of $5\,\%$ is present and from the first to the second scan, the mean activity concentration decreases slightly. Visually the PET measurement appeared reproducible, even though the noise is perceivable in the images. However, the scatter plot and Bland-Altman-plots shown in Figure 4.15 suggest no linear relationship between both scans. Pearsons' $r = -0,044$ supports this observation by suggesting the absence of correlation. Thus, volumes with homogeneous activity concentration may appear visually reproducible even though there is no linear relationship of grey values of voxels present. This example shows the inapplicability of voxel-wise PCC to subsequent PET measurements for the determination of reproducibility.

Furthermore, the comparison of hypoxic target volume contours created by experienced observers and the proposed algorithm may allow determining the reliability of FMISO PET. Thus, focus in this analysis is inter-observer-variability, intra-observer-variability between subsequent FMISO PET scans, corresponding measurements from the ant-based algorithm, using thresholding and contour comparison between algorithms and observers.

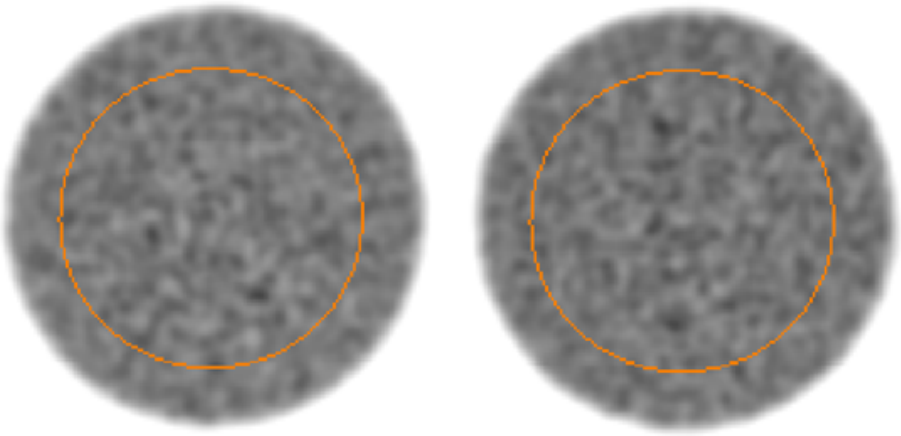


Figure 4.14.: In a PET hantom experiment, two subsequent measurements of the same phantom are compared with each other. Visually obviously both images show the same object. The voxel values within the orange outlined region of interest were analysed statistically.

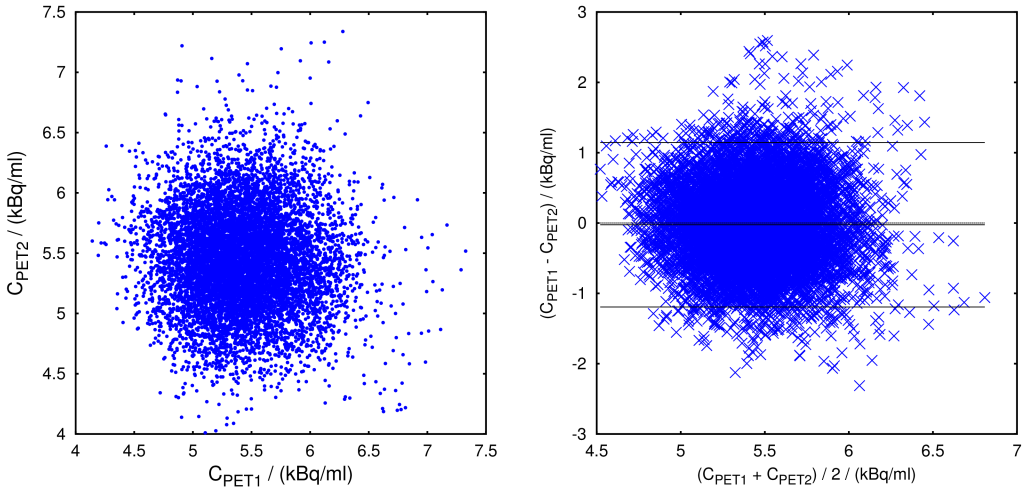


Figure 4.15.: Scatter plot and Bland-Altman plot show the relationship between voxels of both PET measurements. There is no linear correlation. The present variability is a result of noise and decay.

4.5.3. Materials and methods

Contour comparison for the determination of changes between subsequent FMISO PET scans can be done using a subgroup analysis of the patient group introduced in section 3.1.2. To determine contour differences on subsequent FMISO PET scans, contour comparison was applied to the two FMISO PET scans acquired before and after the first therapy week. This experimental setup is similar to the methodology (Nehmeh et al., 2008) performed with the difference that the average time between the scans is 7 days in our case and 3 days in the study by Nehmeh et al.. In our study patients were undergoing the first fractions of radiotherapy in the meantime. During this first therapy week, limited effects like reoxygeneation through changing perfusion may influence FMISO imaging only to a small degree. Furthermore, if contours were similar before and after the first week even though therapy already started, the repro-

ducibility of these contours is underlined. In this section, subgroup determination, figures of merit and statistical analysis are described.

4.5.3.1. Patients, PET data sets and contours

For the analysis of contour changes, a reliable set of FMISO PET data set pairs was needed. How suitable patients from the whole patient cohort of 43 patients were composed, is visualised in Figure 4.16. Firstly, all patients who didn't receive the second FMISO PET scan after the first therapy week were sorted out. From the remaining 36 patients, availability of contours from experienced observers had to be taken into account, because only full sets of contours should be analysed to get an equitable view on inter-observer-variability. Thus, cases where less than three observers outlined a target volume in both FMISO PET scans were sorted out. Finally, 27 FMISO PET data set pairs were available for contour comparison. For simplicity, the PET data sets are named FMISO1 and FMISO2 in the following.

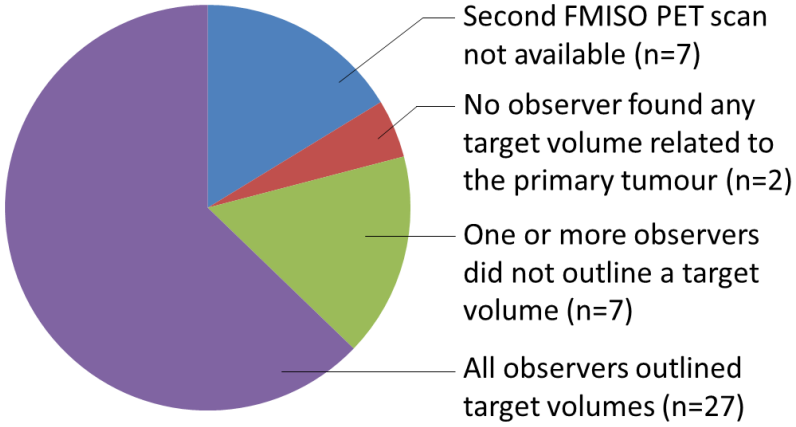


Figure 4.16.: Subgroup creation for FMISO contour reproducibility analysis

4.5.3.2. Image analysis

As basis for contour comparison, the manually-created contours introduced in section 3.1.6 were re-analysed. Furthermore, contours from the two most promising approaches for automatic contouring in the previous experiment, ant-based A and thresholding $T_{1.6}$, were used for automatic contour generation.

For the comprehensive determination of contour differences between FMISO1 and FMISO2 following approaches were used: Firstly, the volume of the manually- and automatically-created segments were compared. Secondly, the mean target-to-muscle ratio TMR and the TMR_{max} derived from the volume definitions were compared. Thirdly, contour comparison using Jaccard index, bilateral contour distance and absolute as well as the relative volume error were analysed. The four figures of merit for contour similarity respectively contour differences were used to determine differences between observers and algorithms on the single FMISO PET data sets FMISO1 and FMISO2 and to determine differences between intra-observer-variability and intra-algorithm-variability when processing subsequent FMISO PET data sets. Furthermore, effects on volume and target-to-muscle-ratio measurements were evaluated.

To determine, whether measured parameters or comparison results were significantly different, the paired t-test was used. A p-value < 0.05 was introduced as criterion for significant differences.

4.5.4. Results

Contour parameters volume V, mean TMR and TMR_{max} derived from FMISO1 and FMISO2 are given in Table 4.5. Furthermore, resulting p-values from the paired t-test are given. Even though differences between FMISO1 and FMISO2 PET data sets were observed, for example volume measurements from observer 3 decreased from approximately 18 ml to 16 ml, none of the measured volume differences were shown to be significant. Furthermore, mean TMR and TMR_{max} measurements did not vary significantly between FMISO1 and FMISO2 as well.

Table 4.5.: Image parameters were derived from contours drawn manually and automatically on subsequent FMISO PET before and after the first week of combined radiochemotherapy. Neither volume nor target-to-muscle-ratio TMR measurements were significantly different between both scans.

Measurement	FMISO1	FMISO2	p-value
Volume by observer 1	23.68 ± 24.92	25.49 ± 38.45	0.838
Volume by observer 2	16.56 ± 18.69	17.16 ± 23.93	0.919
Volume by observer 3	17.71 ± 14.85	16.17 ± 16.75	0.722
Volume ant-based A	15.99 ± 13.17	15.75 ± 14.13	0.950
Volume by $T_{1.6}$	23.12 ± 22.53	21.24 ± 21.36	0.754
TMR by observer 1	1.75 ± 0.22	1.80 ± 0.30	0.509
TMR by observer 2	1.85 ± 0.27	1.93 ± 0.39	0.422
TMR by observer 3	1.85 ± 0.26	1.91 ± 0.38	0.528
TMR_{max} by observers	2.54 ± 0.57	2.63 ± 0.86	0.630
TMR ant-based A	1.91 ± 0.41	1.97 ± 0.60	0.680
TMR by $T_{1.6}$	1.84 ± 0.15	1.86 ± 0.21	0.731

Table 4.6 shows the results from contour comparison within the two FMISO PET scans. Firstly, the contour differences generated by observers were comparable to differences measured when comparing observers segments O with those generated by the ant-based algorithm A or the thresholding approach $T_{1.6}$. The measured differences were not significant as expressed by high p-values. One tested figure of merit comparing differences in the FMISO2 data set, the relative volume error $\delta V_{\text{O vs. T1.6}} = 55.78 \pm 36.19\,\%$, was significantly decreased compared to relative volume error observed between the observers $\delta V_{\text{O vs. O}} = 107.64 \pm 82.56\,\%$. In the other cases, the observed differences allow to state that automatically-created contours were of comparable quality as manually-created contours. In summary, independent from the chosen approach for contouring, either manually or automatic using the algorithms A or $T_{1.6}$, contour distances of about 5 mm and volume errors of about 11 ml were observed. None of the two tested algorithms delivered a clear advantage for contour quality in comparison to each other or to manual delineation.

The comparison of contour-pairs from the two subsequent FMISO PET scans delivered another insight. The results of the contour comparison using Jaccard index, bilateral mean contour distance, absolute and relative volume error are given in Table 4.7. The intra-observer- and intra-algorithm-variability of contours between the PET scans is documented in this table. Corresponding boxplots are given in Figure 4.17. The intra-observer comparison results suggest more different contours in both scans compared

Table 4.6.: Comparing contours from observers and algorithms drawn on the FMISO1 and FMISO2 data sets. The figures of merit deliver inter-observer-differences and observer-algorithm differences which were not significant except the relative volume error in FMISO2: In this case, the thresholding algorithm $T_{1.6}$ showed significantly smaller differences than the inter-observer-differences.

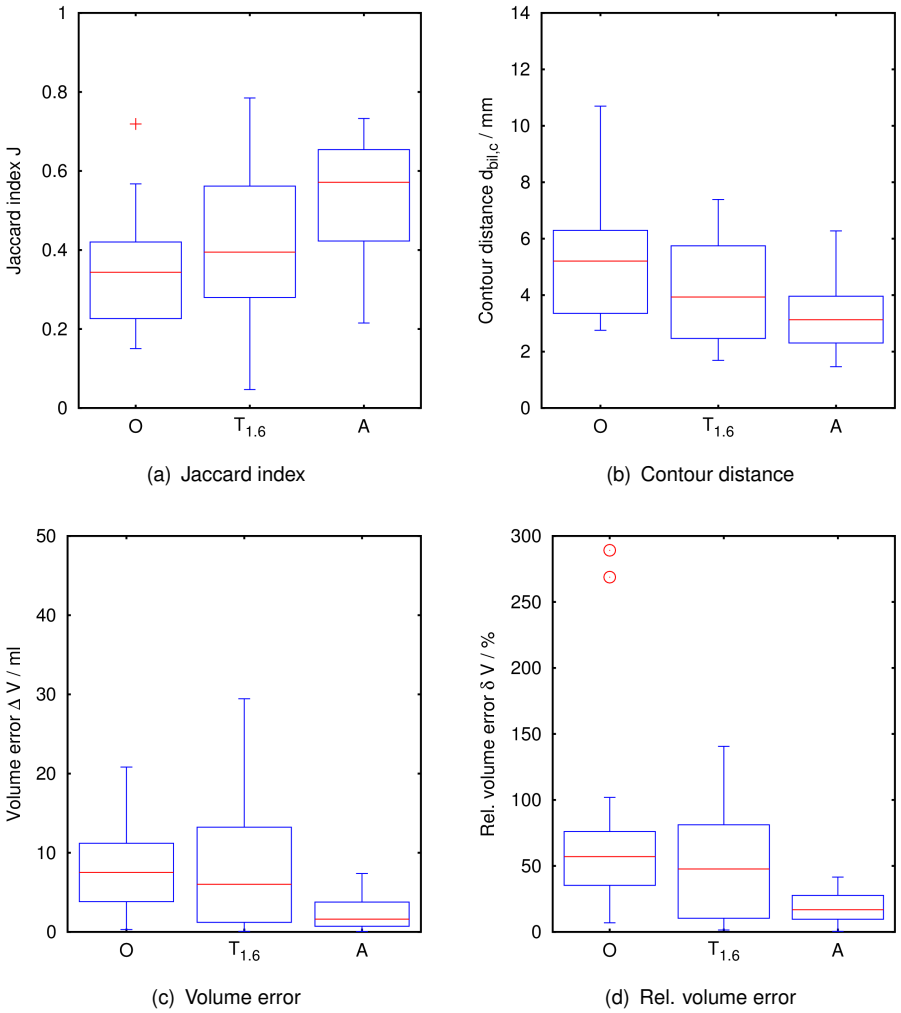
FMISO1 vs. FMISO1	O vs. O	O vs. $T_{1.6}$	p	O vs. A	p
Jaccard index J	0.43 ± 0.14	0.47 ± 0.15	0.294	0.46 ± 0.11	0.390
Contour dist. $d_{bil,c}$/ mm	4.75 ± 2.00	4.82 ± 1.85	0.888	4.59 ± 2.12	0.779
Volume error ΔV/ ml	10.53 ± 8.85	7.29 ± 6.43	0.129	8.23 ± 9.70	0.366
Rel. vol. error δV/%	123.52 ± 188.85	49.94 ± 38.23	0.053	52.27 ± 58.21	0.067

FMISO2 vs. FMISO2	O vs. O	O vs. $T_{1.6}$	p	O vs. A	p
Jaccard index J	0.43 ± 0.15	0.45 ± 0.14	0.595	0.42 ± 0.13	0.879
Contour dist. $d_{bil,c}$/ mm	4.31 ± 1.79	4.96 ± 1.98	0.209	4.70 ± 2.48	0.510
Volume error ΔV/ ml	12.73 ± 21.55	9.99 ± 11.47	0.562	10.88 ± 18.15	0.735
Rel. vol. error δV/%	107.64 ± 82.56	55.78 ± 36.19	0.004	77.63 ± 105.53	0.250

to the algorithms. An example set of contoured images visualising this fact is given in Figure 4.18: While observers interpret local signal changes in the image and drew different contours in both scans, the thresholding and the ant-based approach were robust against those local changes and created similar contours. However, depending on which figure of merit is analysed for the thresholding algorithm $T_{1.6}$ these differences were signifcant or not. As shown for the inter-observer comparison, the contour distances were about 5 mm and absolute volume errors about 10 ml. The ant-based segmentation algorithm showed an improvement for contour similarity. When comparing contour differences between observers and ant-based delineation contour distance and volume error were significantly smaller with 3 mm and 3 ml, respectively. To visualise differences in volume measurements of the observers and algorithms between the two PET scans, Bland-Altman plots are given in Figure 4.19. Obviously, the volume differences between the PET scans are minimised using the ant-based segmentation algorithm. This suggests that changes within the irradiated tumour volume during the first week of combined radiochemotherapy influence ant-based contours only to a limited degree.

Table 4.7.: When comparing intra-observer-differences between FMISO1 and FMISO2 with differences of the two algorithms A and $T_{1.6}$, smaller volume differences were observed for both algorithms. However, only the contour comparison results of algorithm A between FMISO1 and FMISO2 were concordantly significantly improved to the intra-observer contour similarity. The thresholding algorithm $T_{1.6}$ revealed significant differences only in the bilateral mean contour distance metric.

FMISO1 vs. FMISO2	O vs. O	$T_{1.6}$ vs. $T_{1.6}$	p	A vs. A	p
Jaccard index J	0.34 ± 0.14	0.42 ± 0.20	0.112	0.54 ± 0.14	0.000
Contour dist. $d_{bil,c}$/ mm	5.31 ± 1.98	4.19 ± 1.66	0.029	3.26 ± 1.24	0.000
Volume error ΔV/ ml	10.70 ± 15.68	10.00 ± 13.41	0.860	2.61 ± 2.18	0.010
Rel. vol. error δV/%	96.99 ± 125.19	101.46 ± 267.07	0.937	18.00 ± 11.78	0.002



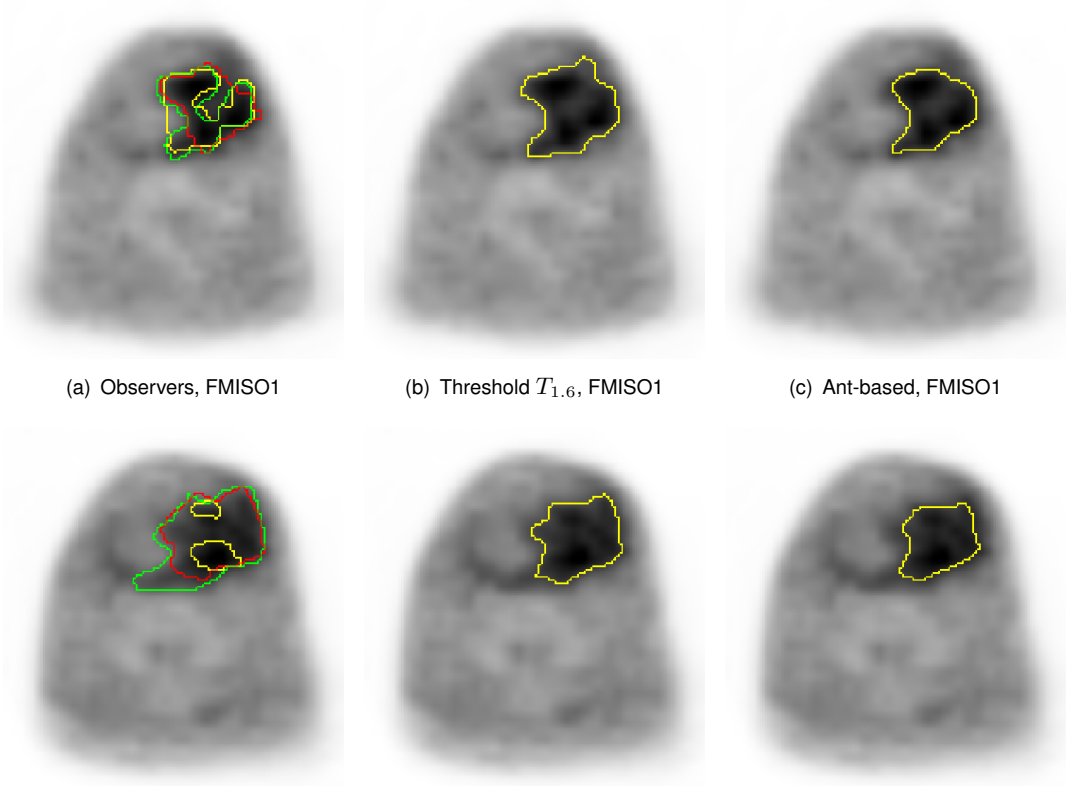
(a) Jaccard index

(b) Contour distance

(c) Volume error

(d) Rel. volume error

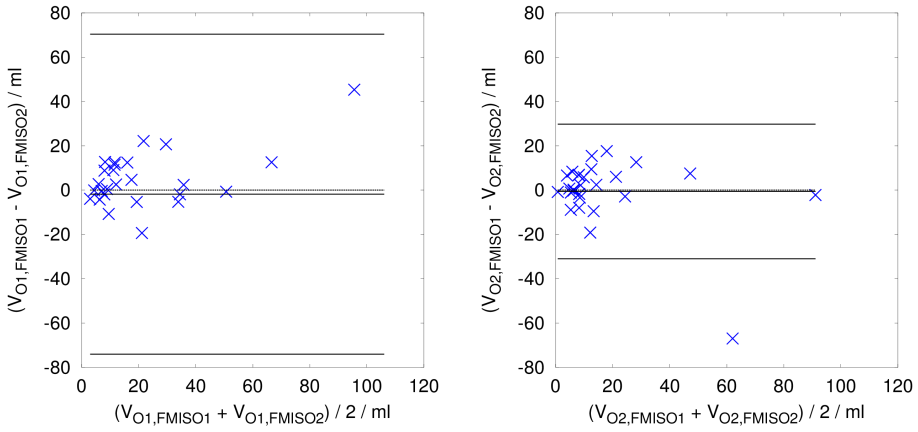
Figure 4.17.: Boxplots summarising the figures of merit used for contour comparison between the two FMISO PET scans reveal that both algorithms improve contour-reproducibility compared to the observers.



(a) Observers, FMISO1 (b) Threshold $T_{1.6}$, FMISO1 (c) Ant-based, FMISO1

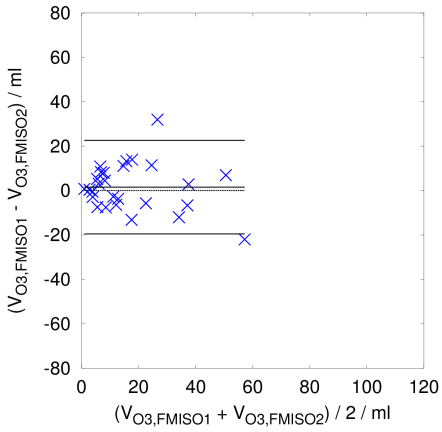
(d) Observers, FMISO2 (e) Threshold $T_{1.6}$, FMISO2 (f) Ant-based, FMISO2

Figure 4.18.: In the given pair of FMISO1 (a-c) and FMISO2 (d-f) data sets corresponding contours of the observers and algorithms are shown. Contours visually appear more reproducible if the contours were generated automatically by the algorithms $T_{1.6}$ and A.

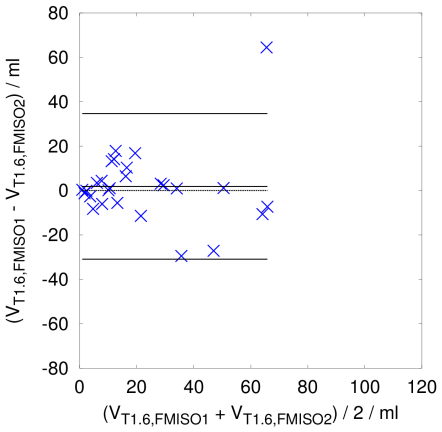


(a) Observer 1

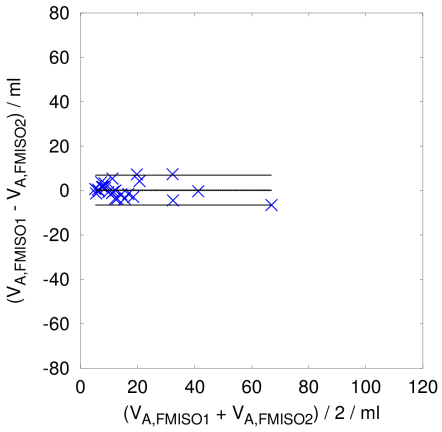
(b) Observer 2



(c) Observer 3



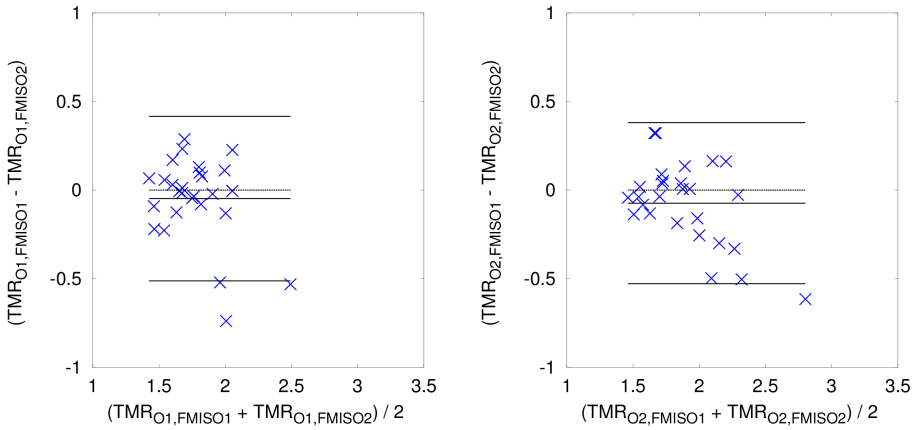
(d) Threshold $T_{1.6}$



(e) Ant-based

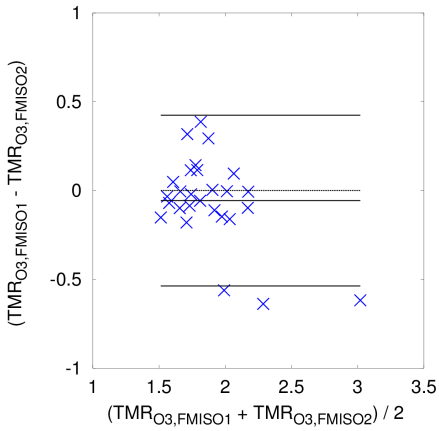
Figure 4.19.: The observed volume differences between subsequent FMISO PET scans and manually-created contours depended on the observer who did the delineation. The thresholding algorithm $T_{1.6}$ showed volume differences similar to the volume differences observed between manually-created contours. However, the ant-based algorithm A minimised the volume error observed between the contours from both scans.

Comparing the TMR measurements between the subsequent FMISO PET scans revealed increased TMR changes determined by the ant-based approach as shown in the Bland-Altman plots in Figure 4.20. While the observers concordantly determined TMR changes between -0.5 and 0.5, the ant-based approach determined slightly larger variances. Thus, the decreased volume differences of algorithm A shown in the previous paragraph lead to increased TMR differences. The example pair of PET data sets and corresponding contours given in Figure 4.21 may explain this observation: While observers and the thresholding approach $T_{1.6}$ determined volume changes between the subsequent FMISO PET scans, the ant-based approach outlined more similar target volumes. Thus, changes in mean TMR within the FMISO-positive volume were detectable more distinct using the ant-based approach compared to the thresholding algorithm $T_{1.6}$ and the observers.

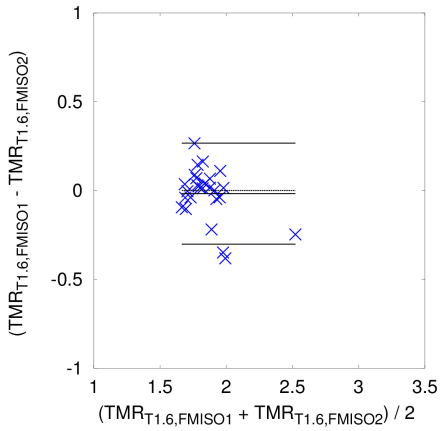


(a) Observer 1

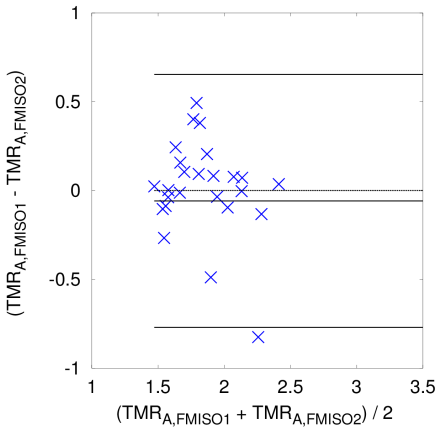
(b) Observer 2



(c) Observer 3

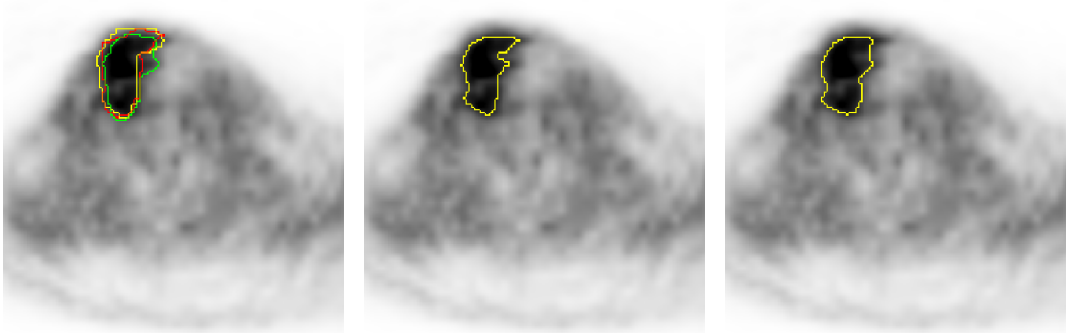


(d) Threshold $T_{1.6}$

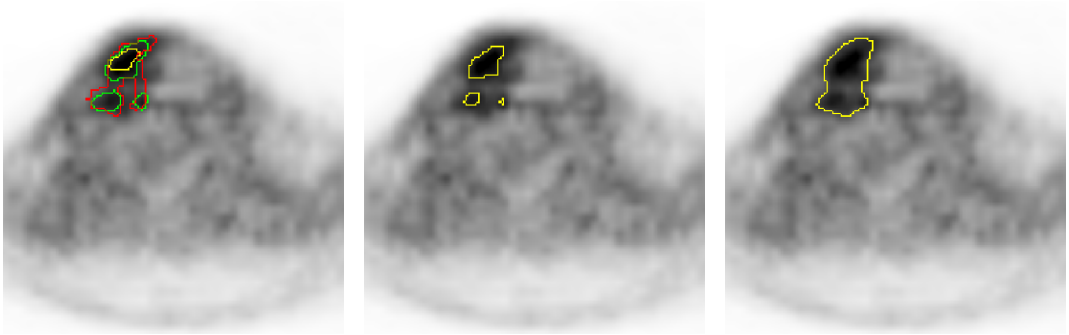


(e) Ant-based

Figure 4.20.: Bland-Altman plots visualise the differences in TMR estimation. While observers and the ant-based algorithm A revealed differences in TMR estimation of usually not more than 0.5, algorithm $T_{1.6}$ showed decreased differences in TMR estimation up to approximately 0.3.



(a) Observers, FMISO1 (b) Threshold $T_{1.6}$, FMISO1 (c) Ant-based, FMISO1



(d) Observers, FMISO2 (e) Threshold $T_{1.6}$, FMISO2 (f) Ant-based, FMISO2

Figure 4.21.: In this pair of clinical FMISO PET data sets, contour reproducibility appears improved using the algorithm A. In this case the three observers measured almost constant TMR measurements of 1.62, 1.88 and 1.54 for FMISO1 and 1.59, 1.84 and 1.58 for FMISO2, respectively. The thresholding algorithm $T_{1.6}$ measured a TMR slightly decreasing from 1.81 in FMISO1 to 1.74 in FMISO2. However, algorithm A revealed TMR decreasing from 1.75 in FMISO1 to 1.51 in FMISO2. Thus, in this case, only algorithm A was able to value the decrease of TMR within the target volume. This is related to a smaller volume error of the ant-based approach between the two PET scans.

5. Discussion

This chapter summarises findings from the performed experiments and points out how the results may be interpreted. Experimental procedures are justified in detail. Furthermore, limitations for using the proposed ant-based segmentation algorithm are given.

5.1. Phantom experiment

The phantom experiment introduced in section 3.1.1 allowed to simulate various contrast levels in the resulting PET data sets without interacting with the phantom itself. Usually, those phantom measurements are interrupted by the experimenter to increase the background activity manually. This includes repositioning of the phantom. In the applied setup this interruption is not needed, because background and target volumes contain radionuclides with a different half-life. Thus, the contrast changes with time without the need for interaction. This advantage of unnecessary registration of the single PET data sets allows voxel-by-voxel analysis and direct comparison of contours created on any PET data set of the phantom.

However, some disadvantages must be discussed. Firstly, the introduction of two radionuclides in one phantom creates a situation which is not relevant in clinical practice. Through different mean positron ranges of the both nuclides, $2.9\,$mm for $[^{68}Ga]$-gallium and $0.6\,$mm for $[^{18}\mathrm{F}]$-flourine, a gradient between target and background may appear which would not be observed in clinical images. However, the fact that the intrinsic spatial resolution of the PET scanner is approximately $5\,$mm suggests that the effect may not be detectable in the images. The effects of scattering as introduced in section 2.1.3 may be overwhelming. Additionally, the gradient around a solid glass-sphere would be different from the gradient in clinical images, because the glass sphere is a inactive region inducing changes on the objects boundary in the image, because of the spill-in effect. Furthermore, under low contrast conditions, the glass spheres became visible in the PET images as shown in Figure 5.1. The reason is that only if background and target show similar activity concentration, the inactive glass walls become visible as valley within the activity distribution. Otherwise the spill-over effect introduced in section 2.1.4 hides the glass under the gradient of the adjacent volumes. Thus, a phantom experiment utilising glass spheres may not be the optimal approach for calibrating or evaluating a segmentation algorithm (Hofheinz et al., 2010; van den Hoff and Hofheinz, 2012). Secondly, the wax spheres used to overcome the glass-wall related issues were also not optimal target-simulating objects. The approach introduced by (Bazañez-Borgert

et al., 2008) suffers from a complicated production process (Haase et al., 2010). When the solution of gallium-chloride is mixed with hot wax, the solved substance anneals and introduces small crumbs of gallium-chloride. These crumbs may later appear as hot regions in the PET images, as shown in Figure 5.2. Furthermore, the wax spheres were put into a refrigerator to accelerate the hardening-process. If hardening happens too fast, the wax spheres may contain breaks, which may be seen as inactive regions in the PET images.

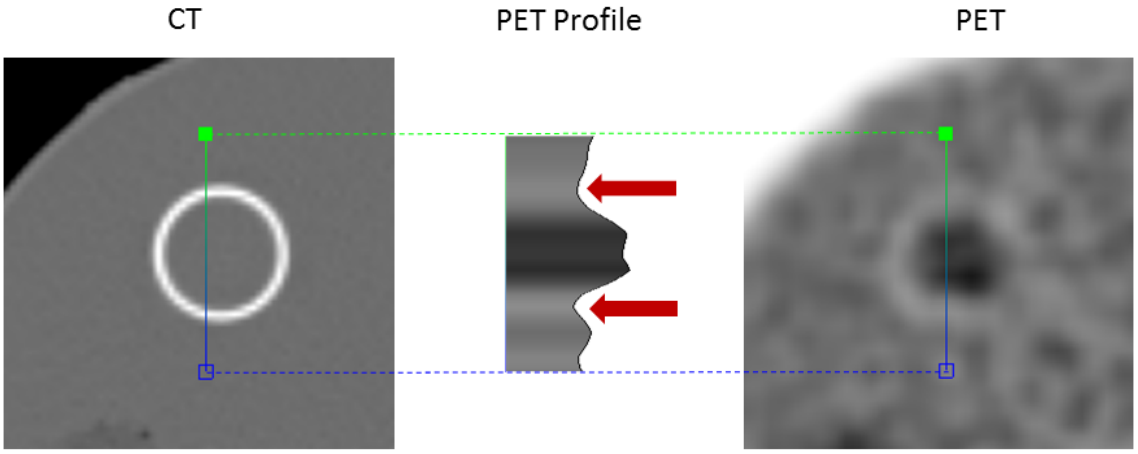


Figure 5.1.: When glass spheres are used in a PET phantom to simulate target volumes, the spheres introduce a inactive region between target volume and background. This inactive region is visible within the PET images and profiles along a line through the target volume (red arrows), when the contrast between target and background is low. Especially in such a case the influence of the glass sphere to a segmentation result must be taken into account.

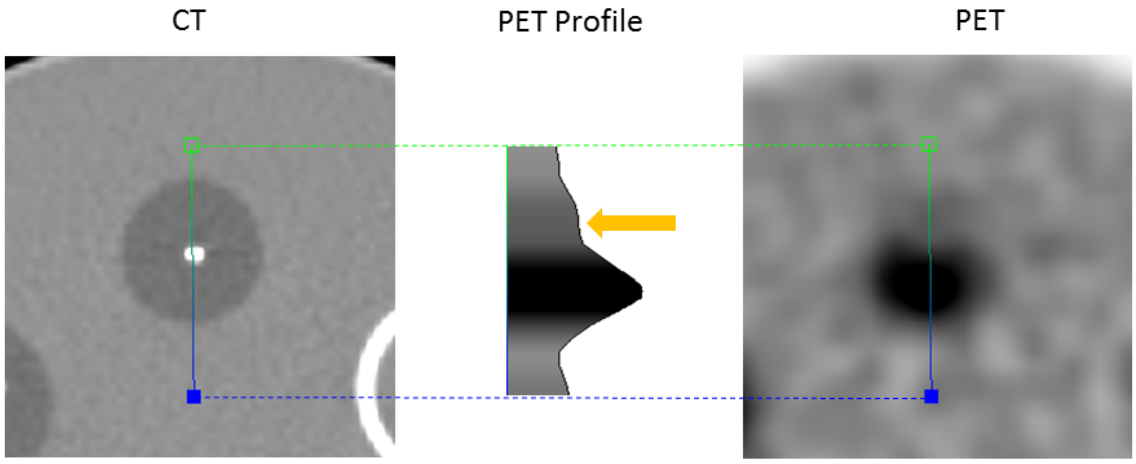


Figure 5.2.: Utilising wax spheres in a PET phantom experiment eliminates the influence of inactive regions between target and background volumes. However, producing wax spheres with a homogeneous activity distribution is a complicated process. In the visualised case, a wax sphere is shown that includes a hot region and a region within the wax sphere where the activity concentration is hardly higher than the background (orange arrow).

Finally, the influence of the usage of different nuclides was assumed to be not relevant for PET imaging, while the imponderables introduced by glass spheres and wax spheres were assumed to be more relevant. Thus, for a clinically acceptable evaluation of a new algorithm, comparisons of contours created on patient data sets under clinically relevant conditions, or more precisely created by experienced physicians

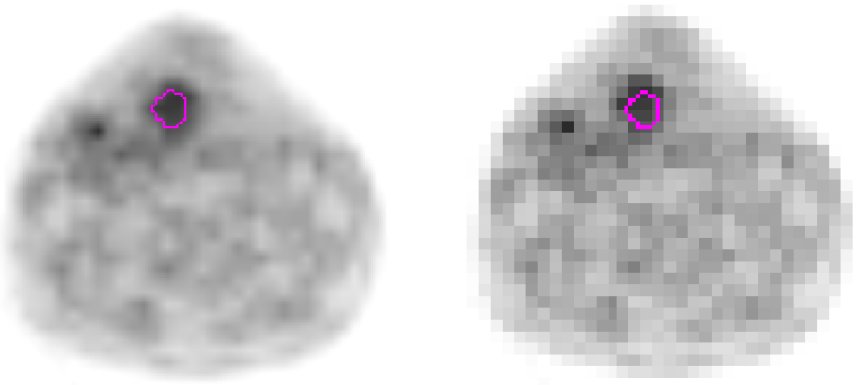
and medical physicists on a clinical research workstation, was focussed in the presented analysis.

5.2. Registration and resampling procedure

Before the experienced observers were allowed to draw target volume outlines on the PET data sets, the data sets needed to be registered and resampled. Only registered data sets allow analysing spatial shifts of volumes or volume changes like growth or shrinkage during therapy and voxel-by-voxel analysis. However, registration is a process which may result in PET images which are dislocated to a small degree. The time consuming process of CT-CT-registration and afterwards correcting the registration parameters by visual inspection of the corresponding CT-PET-fusion, introduced in section 3.1.4, was chosen to minimise the effect of registration errors. The possible alternative of doing only CT-CT-registration and application of the registration parameters to the PET images was seen as possible accumulation of registration errors between three pairs of images. If CT and PET images, which were acquired in tandem, are misregistered because the patient moved between CT and PET scan, there could be a registration error of some millimetres. If this happens in two PET/CT scans and CT-CT-registrations result in a registration error of some millimetres, these three registration errors may be accumulated resulting in a clearly visible misregistration and hindering the desired analysis. In the applied approach, every single PET image was registered to the initial CT image and furthermore, all PET images were visually checked for registration errors with the initial CT and the initial FDG PET image. This apparently complicated process of registration quality assurance was seen as the best practically available procedure for image registration in this study. One limitation of the further usage of the registered images should be given: voxel-by-voxel analysis of objects located distant to the primary tumour may be impossible using these registered data sets, because the registration was focussed on the primary tumour and directly surrounding tissues. Misalignment of distant FMISO-positive lymph nodes was obvious in some data sets. Analysis of lymph nodes was not performed in this study and thus, registration accuracy for them was not relevant.

Furthermore, the resampling technique must be taken into account to determine influences on the image data. Resampling is the process of transferring an image from one coordinate system and matrix size to another system and matrix. Resampling includes recalculation of all voxel values by interpolation. Thus, if voxel values, such as SUV, should be determined, in general non-resampled data should be preffered, to eliminate errors introduced by interpolation. This may indeed be more relevant, if maximum values are of interest (Haase et al., 2014). Figure 5.3 shows an example FMISO PET data set where changes in SUV_{max} were observed when resampling the PET image to the coordinate system of its corresponding CT.

However, when accurate contouring is of primary interest and delineation software only supports outlining on voxel-size, it is recommended to resample the data to get more accurate contours. Most of the software packages available for delineation anyhow display interpolated views on the PET images. From this point of view it is not relevant if an observer outlines resampled images that are viewed using interpolation or non-resampled images viewed using interpolation. For the observers eye, both images may not be differentiable. Furthermore, if the observers outline images in an interpolated view, the algorithm should



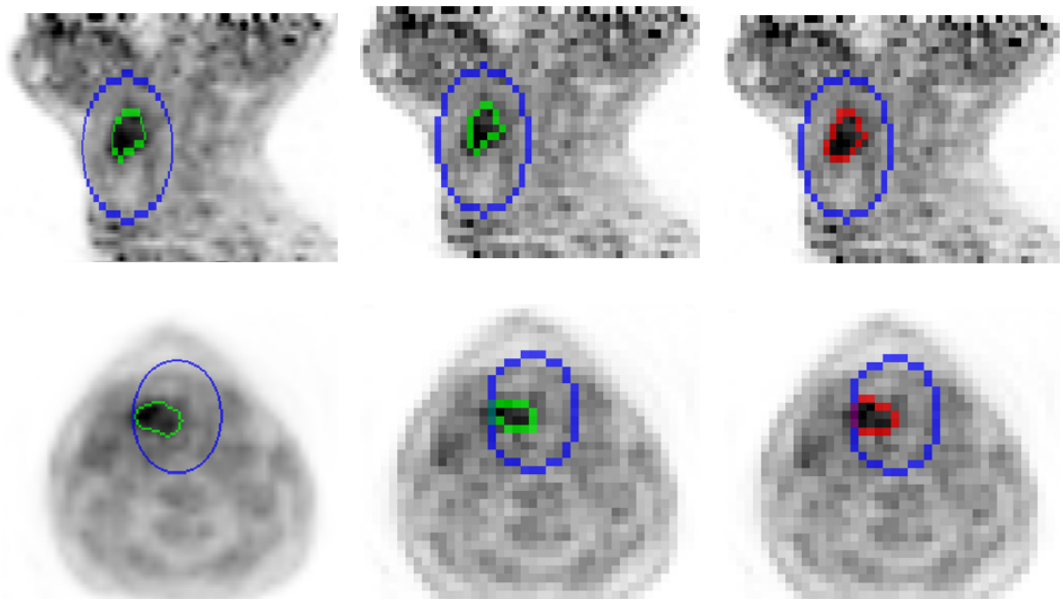
(a) (b)

Figure 5.3.: When comparing voxel statistics of resampled and non-resampled PET data sets, it is obvious that values differ between these images. The ant-based determined target volume in an FMISO PET image shown in this example (pink outline) has a $SUV_{\text{mean,r}} = 2.09 \pm 0.15$ in the resampled case (a) and $SUV_{\text{mean,nr}} = 2.07 \pm 0.25$ in the non-resampled case (b). The $SUV_{\text{max,r}} = 2.57$ in the resampled image is decreased to a small degree compared to the non-resampled image with $SUV_{\text{max,nr}} = 2.74$. Thus, especially the SUV_{max} may be influenced by resampling.

be applied to a similar situation to make contours from observers and algorithm comparable. Otherwise, the automatically generated contours would have a rectangular shape as shown in Figure 5.4.

If accurate voxel statistics for non-resampled PET data sets are needed in future research, the software presented in section 3.3, which was developed as part of this thesis, allows this analysis. Transforming contours from a small voxel size to a large voxel size introduces some kind of information-loss, but this may be the best practical way to determine SUV_{max} from data actually delivered by the clinical scanner. How contours may look like if they are transformed from the CT image coordinate system with smaller voxel size to the PET image coordinate system with larger voxels, is shown in Figure 5.4. However, generating the target volume definitions is also possible in the coordinate system of the PET image but results in unnatural appearing target volume definitions with a rectangular shape. Depending on the focussed issue, for example determination of maximum versus mean voxel values, it must be discussed, to which coordinate system the segmentation algorithm should be applied and if contour transformation introduces information loss or not. Depending on the purpose of the analysis, it would also be an asset to disable resampling. In general, voxel value statistics should be performed in non-resampled image data, because resampling changes the voxel values in a certain VOI, which were acquired by the scanner. For contour comparison and evaluation of segmentation algorithms, resampled data with decreased voxel size should be preferred.

The applied combination of various software packages introduced in sections 3.1.4 and 3.1.6 appears overstated. However, for every single task, registration, resampling, manual contouring and contour comparison, a decision on which software should be used were made carefully. For the initial step of automatic CT-CT-registration, Pinnacle was used, because it was the only available clinically certified software which allowed registration and viewing of registration parameters. Only in this software it was possible to read out the rigid registration parameters to enter them in other software. For resampling the



(a) Contour in CT image coordinate system

(b) Contour transformed from CT to PET image coordinate system

(c) Contour created in PET image coordinate system

Figure 5.4.: Two contours in blue and green generated on resampled PET data in the coordinate system of the CT are shown in figure (a). When this contour is transferred to the coordinate system of the PET (b), a different segmentation result may be present as compared to contours which were created in the coordinate system of the PET image (c).

images, Rover was used because at that time, it was the only in-house available software that allowed resampling and storing the resulting images in DICOM file format for further processing. For ease-of-use, Rover was also used for manual registration correction, pseudonymisation and quality assurance. The experienced radiologist viewed the image fusion of all PET images with the initial CT in Rover. The check if FMISO PET and FDG PET images were registered sensefully, was performed in the Geisterr software presented in section 3.3, because checking PET-PET image fusion deserves special viewing options such as non-standard colour tables. For manual definition of the volume of interest (VOI), Rover was used as well, because this step was done by the experienced radiologist, who also did the registration check. Thus, it was possible to do both steps during the same session. Furthermore Rover allowed exporting the VOI-definitions in a non-proprietary file format, which can be read by other registration software. Manual delineation of all FMISO PET data sets was then performed on Syngo workstations, because two of the observers were already familiar with the software. Automatic contouring was performed in Rover with the algorithm proposed by (Hofheinz et al., 2012c), because Rover already included an implementation of the published algorithm. All the other established PET segmentation algorithms introduced in section 2.5.2, as well as the proposed ant-based method, were implemented in Geisterr software as well to have all available in one place. Contour comparison for determination of the figures of merit introduced in section 2.3 was also implemented in the same software.

5.3. Optimisation and calibration

Even though segmentation algorithms are under development for years, there is still controversy about if and how accurate calibration can be done. On the one side, algorithms were published which need no calibration (Hofheinz et al., 2012c). In such a case, the calibration is not needed, because the algorithm iteratively determines needed parameters, such as background activity, from the image. Calibration from phantom experiments could introduce additional error sources, such as inactive glass walls (Hofheinz et al., 2010; van den Hoff and Hofheinz, 2012). From this point of view, a robust algorithm could indeed be in no need of calibration. However, on the other side, algorithms were published for which a calibration is strongly recommended. A multicentre calibration performed by Schaefer et al. of a threshold-based algorithm revealed quite different parameter settings depending on which scanner and protocol was utilised (Schaefer et al., 2012). But also in this study it was shown that the calibration of an algorithm for scanners identical in construction leads to parameter settings with only small differences.

The detailed calibration of the τ_{min} parameter of the proposed algorithm in section 4.3 revealed that different methods of calibration, especially the choice of figures of merit, may lead to different parameter settings. The reason is that volume-over- and -underestimation is weighted differently by different figures of merit as shown in section 2.3.9. When defining $\tau_{\text{min,opt}} = 90$ as compromise the contour distance was seen as more important than the Jaccard index, because of this effect. The calibration of the τ_{min} parameter furthermore showed that its value cannot be determined exactly, because it depends on inspected patient data sets and the reference generating observers used as reference. The Jaccard index measurements differed only to a small degree, even if the pheromone threshold was increased or decreased by 10. Furthermore, choosing different pheromone threshold τ_{min} values such as 80, 90 or 100 lead to contour differences which were about 1 mm (Haase et al., 2013b). Contour differences of the ant-based approach in subsequent FMISO PET scans, presented in section 4.5, were smaller compared to contour differences between observers, which were about 5 mm. Also, automatic contours on different PET scans, which were different by about 3 mm. Thus, the recommended pheromone threshold $\tau_{\text{min}} = 90$ can be used certainly, because its determination and evaluation was done in very detail.

The calibration furthermore showed that contour errors and uncertainties of 6 to 10 mm must still be expected when working with FMISO PET. This is independent from the applied contouring technique: For both, manually-created contours as well as automatically-created contours, cases were observed where such differences between target volume definitions were present. For a clinical application of FMISO PET-based derived contours, double-checking by experienced observers is recommended. Furthermore, the measured contour distances can guide appropriate safety-margins, if FMISO PET-based contours may play a role in radiotherapy planning. In head-and-neck-cancer anyhow margins between CTV and PTV of about 8 mm are used. If an algorithm reveals large contour differences when applied, for example to subsequent FMISO PET scans, the margin could be increased to cover more of the volume of uncertain FMISO signal. However, reducing margins is in interest for many radiooncologists. For this purpose, FMISO PET based radiotherapy treatment planning may not be recommended due to the observed uncertainties.

The test and validation cohorts used in the optimisation procedure were of different sizes, which may affect the resulting optimal threshold used for contouring. However, the used separation was defined by the study protocol of the exploratory trial (Zips et al., 2012). At the time of writing this thesis, the trial and image analysis is still on-going. However, the proposed algorithm and its calibration procedure may be repeated as soon as the last patient data set was processed in the study. For a more robust cross-validation analysis, larger groups would be needed anyhow. However, the fact that contour differences between algorithm and observers were not significantly different from differences between observers may still be present in a larger patient cohort. Contrary, if observers learned to outline more consistently after more patients were imaged, the calibration of the algorithm would not essentially suffer from differences between the observers. The calibration could deliver parameters leading to contours even more similar to manually-created contours. Thus, a more accurate automatic delineation could be possible.

Furthermore, the proposed algorithm was optimised and validated for the FMISO PET images acquired at one site. It may not be applicable to images from other sites, because imaging and reconstruction parameters may affect segmentation results. Thus, standardisation of PET image acquisition parameters is worthwhile to harmonise segmentation results (Cheebsumon et al., 2011b,a).

5.4. Comparison with other algorithms

The comparison of the proposed algorithm revealed an alternative segmentation algorithm to the proposed ant-based technique. The thresholding algorithm $T_{1.6}$ which applies a threshold of 1.6 times background activity concentration, also delivered reliable volume definitions. However, it must be highlighted that applying a standardised threshold to all data sets may deliver systematically biased image-derived parameters, as shown in section 4.4. It cannot be assumed that a single threshold is reliably applicable to different patients (Biehl et al., 2006). Furthermore as also shown in section 4.4, the $T_{1.6}$ approach delivered underestimated volume definitions if the target-to-background ratio as determined by the observers was below 1.6. The consequence is obvious: volumes as defined by observers with mean activity concentration below 1.6 times background activity, cannot be delineated using this thresholding approach. But such volumes were observed in about a quarter of the cases where observers consistently defined an FMISO-positive volume. These low contrast volumes cannot be explained with noise effects: usually noise levels of about $20\,\%$ are observed in the background volume of FMISO PET scans in the scanner setting applied (Abolmaali et al., 2012). Small regions with $TMR = 1.2$ may be a result of noise. But for example a massy target volume with mean $TMR = 1.4$ is obviously distinguishable from a noise-related structure. Thus, especially for low-contrast data sets with $TMR < 1.6$, ant-based delineation is recommended. Furthermore, if the $T_{1.6}$ approach underestimates volumes with low contrast, but in average determines correct volume measurements, the consequence is that target volumes with high contrast are overestimated. However, a similar effect was observed using the ant-based approach: volumes with high contrast were underestimated. Thus, FMISO PET based volume definitions with $TMR > 2.5$ should be reviewed carefully. In the present study, only a hand full of data sets with volume definitions with $TMR > 2.5$ were observed. Alternatively, depending on the measured TMR

value, segmentation results from other algorithms, such as the $T_{1.6}$ approach, should be considered for further volume analysis.

A similar analysis of reliability of ant-based contours and therefrom derived parameters shall be mentioned: it is possible to use the mean TMR of volume definitions for distinguishing between contours that were defined FMISO-positive and -negative consistently by the three observers. By ROC analysis, a threshold $TMR_{min} = 1.4$ was defined to detect FMISO-negative cases with sensitivity of $89\,\%$ and specificity of $83\,\%$ (Haase et al., 2013a). Thus, if the ant-based approach delivers a volume definition with $TMR < 1.4$, it should be rejected as invalid. In such cases the probability is high that the observers would not outline a target volume. Such an approach of contour analysis and -postprocessing may lead to quality assurance methods applied in individual cases. Checking derived contours for meaningfulness in such a way may also increase trust of clinicians in modern intelligent algorithms like the proposed one.

5.5. Contour reproducibility

The most prominent observed advantage of ant-based segmentation was contour-reproducibility in subsequent FMISO PET imaging. Contours derived from FMISO PET before and after the first week of therapy revealed highest similarity using the ant-based approach. Observers and the $T_{1.6}$ thresholding approach were significantly outperformed by means of contour reproducibility in this analysis. However, there may be biological changes to tumour volumes and normal tissue during the first week of radiochemotherapy. If these changes should be measured using a segmentation algorithm, it is not worthwhile to recommend an algorithm that oversees these changes. However, applied contour comparisons showed that contour differences between algorithm and observers were not significantly different from differences between the observers. This was true for both subsequent FMISO PET data sets. Thus, the signal changes during early therapy may not be quantified savely by manual contouring because of the inter-observer-variability, but could be determined more reliably using automatic routines such as the ant-based approach.

On the other hand, local signal changes should not be misinterpreted as radiation-induced biological effects. Physical effects like the partial volume effect and noise may lead to visually moving hot spots between the imaging timepoints. However, these changes should not induce a large change in volume measurements. Furthermore, a huge change of the hypoxic volume would not be easy to explain. However, in a previous analysis of the first 25 patients, in 5 cases, an increase of the hypoxic volume by more than $100\,\%$ was observed using a fixed threshold of 1.6 times background activity concentration (Zips et al., 2012). These changes could be explained by radiation-induced changes of tumour perfusion and thus reoxygeneation of the hypoxic volume. But also image analysis related effects, such as signal changes in the background volume can lead to increasing or decreasing hypoxic volume measurements. Thus, a background-indepentend method should be preferred. In the other 20 cases the hypoxic volume stayed almost stable or decreased during the first therapy week. A change of mean or maximum tracer uptake would be easier to explain, because reoxygenation of tumour tissue may appear during the first

week of radiochemotherapy. Furthermore, differences in volume estimation cannot be explained by registration errors. But differences in contour distance measurements may be related to registration errors. In both figures of merit, volume error as well as contour distance, similar observations were made: the algorithms contours are more reproducible between both PET scans than manually-created contours.

Finally, the thresholding approach $T_{1.6}$ depends not just on the definition of the volume of interest but also on the definition of the background volume. Thus, for its application, two user-related error sources are present. The ant-based approach does not deserve a predefined background volume, the scout ants determine surrounding background activity indirectly. Thus, the ant-based approach is less observer dependent than the thresholding approach using reference background activity.

6. Conclusions

In the present thesis, a new PET volume segmentation algorithm based on the behaviour of virtual ants was optimised and validated. The algorithm is specialised on low contrast PET data sets which appear in hypoxia imaging using FMISO PET. A detailed review of the research field of FMISO PET image analysis for the purpose of quantitative image analysis, for example for patient prognosis assessment, and radiotherapy treatment planning was given. Furthermore, methods for contour comparison and thus, the evaluation of segmentation algorithms were systematised to allow correct interpretation of validation results. The new algorithm was reviewed in detail as well. A number of algorithm parameters were reviewed, because they were analysed, presented and discussed in detail in previous publications. The applicability of the algorithm to clinical PET data sets was shown using four software experiments. Firstly, its principle was evaluated: The pheromone field the ants use for indirect communication allowed more accurate volume delineations by thresholding than the PET data sets from the phantom experiment which were processed to generate the pheromone fields. Thus, in principle further processing of the pheromone field to determine a volume definition is worthwhile. Secondly, the optimal threshold for segmenting the pheromone field after processing FMISO PET images from head-and-neck-cancer patients was determined and reviewed using cross-validation. The optimal threshold delivered contours which were not inferior to manually-created contours from experienced observers. The automatically-created contours were similar to the manually-created contours to a degree like the manually-created contours from different observers were similar to each other. The third experiment was performed to compare the results of the proposed segmentation algorithm with established methods in the field of PET image segmentation. During this experiment, another rather simple segmentation algorithm based on the threshold of 1.6 times background activity delivered reliable segmentation results as well. It stands to reason that the ant-based technique was superior if target volumes with activity concentration below 1.6 times background activity concentration had to be outlined. Thus, the ant-based algorithm is recommended for low contrast FMISO PET image analysis. Furthermore, it is less observer dependent, because the ants do not deserve a definition of a background volume, hence background signal. The virtual ants used in the algorithm indirectly determine background activity concentration on their own. Another advantage may be seen in the time needed for volume delineation, while an experienced observers needs minutes to inspect and outline a hypoxic target volume, the algorithm finishes segmentation in less than a minute. In the fourth and final experiment, contour reproducibility on subsequent FMISO PET imaging was studied. The study is limited in interpretability because during the time between the scans, the patients retrieved the first fractions of radiotherapy. Thus, changes in tracer distribution may be observed and thus lead to varying contours in these two compared FMISO PET data sets. On the one hand, ant-based contours

were not significantly different from those of the experienced observers. And on the other hand, contours drawn on different FMISO PET images of the same patient were significantly more reproducible using the ant-based technique. The contradiction of these two facts demonstrates an advantage of the algorithm: it allows mostly observer-independent volume delineations and target-to-muscle TMR measurements. Without such a reliable algorithm, accurate mean TMR estimation may be impossible. For this reason the proposed algorithm is further recommended for application in clinical research on FMISO PET imaging. The algorithm is specialised on low contrast FMISO PET images and thus, its contours should always be reviewed and probably corrected by experienced clinicians, especially if mean TMR is above 2.5. Furthermore, if the algorithms delivers contours with mean TMR below 1.4, the contour should be rejected, because the probability that experienced observers would define the corresponding data set as FMISO-negative is high. The present thesis outlines a way for optimising the ant-based algorithm to one specific image modality and rather, one specific PET tracer. The same methods could be used to optimise the algorithm to other tracers, such as FDG, and furthermore to other modalities. For example, diffusion weighted MRI also delivers image data with low contrast between target volume and background. Thus, applying and evaluating ant-based image segmentation in other fields of medical image processing appears promising.

7. Kurzfassung

Im Forschungsfeld moderner Radioonkologie ist heute Therapieindividualisierung ein Kernthema. Dabei werden verschiedene Biologie-basierte Ansätze zur Therapieanpassung diskutiert, von denen viele auf tomographischer Bildgebung der Tumorbiologie basieren. Oft genannt wird in diesem Kontext Positronenemissionstomographie (PET) unter Nutzung des Tracers $[^{18}\mathrm{F}]$-fluoromisonidazol (FMISO) für Hypoxiebildgebung. Hypoxie ist deswegen so wichtig für die Strahlentherapie, weil hypoxische Tumoren sich durch erhöhte Strahlenresistenz auszeichnen. Weiterhin haben Patienten mit hypoxischen Tumoren im Allgemeinen eine schlechtere Prognose als Patienten mit normoxischen Tumoren. Somit liegt Hypoxiebildgebung für Adaption der Strahlentherapie auf der Hand. Beispielsweise könnte FMISO-PET-Bildgebung dazu genutzt werden, sogenannte Hypoxiezielvolumendefinitionen zu erzeugen, die dann im Rahmen der Strahlentherapie mit höheren Strahlungsdosen behandelt werden, um den therapeutischen Effekt zu erhöhen. Dabei wird die Bildanalyse jedoch durch niedrigen Kontrast zwischen Zielvolumen im FMISO-PET und Bildhintergrund erschwert. Etablierte Methoden zur Zielvolumenabgrenzung in PET-Bilddaten versagen bei der Verarbeitung von FMISO-PET. Im Rahmen dieser Dissertationsschrift wurde sich auf diesen Aspekt konzentriert. Ein vorher für FMISO-PET entwickelter Schwarmintelligenz-basierter Segmentierungsalgorithmus wurde optimiert und validiert in einem klinisch relevanten Versuchsaufbau. Dazu wurden FMISO-PET-Bilddaten genutzt, die im Rahmen einer klinischen Studie an der Klinik und Poliklinik für Strahlentherapie und Radioonkologie des Universitätsklinikums Carl Gustav Carus Dresden aufgenommen wurden. Der Algorithmus wurde auf diese Bilddaten angewandt und mittels eines Kreuzvalidierungs-Ansatzes optimiert, bei dem manuell erzeugte Konturen von erfahrenen Observern als Referenzkonturen zum Einsatz kamen. Anschließend wurden die Eigenschaften des Algorithmus und die der resultierenden Konturen im Detail studiert. Dabei zeigte sich, dass die automatisch erzeugten Konturen denen der Observer zu einem Grad ähnlich sind, wie die manuell erzeugten Konturen einander ähneln. Somit ist der Einsatz des Algorithmus in klinischer Forschung empfehlenswert, um Observerabhängigkeiten zu eliminieren. Weiterhin konnte gezeigt werden, dass wiederholte FMISO-PET-Bildgebung vor Therapie beziehungsweise kurz nach Beginn der Therapie zu manuell erzeugten Konturen führte, die signifikant stärker voneinander abwichen als automatisch generierte Konturen unter Nutzung des vorgeschlagenen Algorithmus. Erhöhte Konturübereinstimmung in aufeinanderfolgend aufgenommenen Bilddaten verdeutlicht die Observerunabhängigkeit des Algorithmus. Während mehrere Observer unterschiedliche Konturen liefern, in identischen Datensätzen als auch in aufeinanderfolgender Bildgebung, erzeugt der Algorithmus stabiliere Konturen in beiden Fällen. Höhere Konturreproduzierbarkeit kann also durch den Einsatz des Algorithmus erreicht werden.

8. Summary

In the field of radiooncological research, individualised therapy is one of the hot topics at the moment. As a key aspect biologically-adapted therapy is discussed. Therapy adaption based on biological parameters may include tomographic imaging to determine biological properties of the tumour. One often invoked imaging modality is positron emission tomography (PET) using the tracer $[^{18}\mathrm{F}]$-fluoromisonidazole (FMISO) for hypoxia imaging. Hypoxia imaging is of interest, because hypoxic tumours are known to be radiorestistant. Even further, patients with hypoxic tumours have worse prognosis compared to patients with normoxic tumours. Thus, hypoxia imaging appears promising for radiotherapy treatment adaption. For example, volumetric analysis of FMISO PET could deliver additional hypoxia target volumes, which may be irradiated with higher radiation doses to improve the therapeutic effect. However, limited contrast between target volume and background in FMISO PET images interferes image analysis.Established methods for target volume delineation in PET do not allow determination of reliable contours in FMISO PET. To tackle this aspect, this thesis focusses on an earlier developed swarm intelligence based segmentation algorithm for FMISO PET and rather, its optimisation and validation in a clinically relevant setting. In this setting, clinical FMISO PET images were used which were acquired as part of a clinical trial performed at the Clinic and Policlinic for Radiation Therapy and Radiooncology of the University Hospital Carl Gustav Carus Dresden. The segmentation algorithm was applied to these imaging data sets and optimised using a cross-validation approach incorporating reference contours from experienced observers who outlined FMISO PET positive volumes manually. Afterwards, the performance of the algorithm and the properties of the resulting contours were studied in more detail. The algorithm was shown to deliver contours which were similar to manually-created contours to a degree like manually-created contours were similar to each other. Thus, the application of the algorithm in clinical research is recommended to eliminate inter-observer-variabilities. Finally, it was shown that repeated FMISO PET imaging before and shortly after the beginning of combined radiochemotherapy lead to manually-created contours with significantly higher variations than the variations of automatically-created contours using the proposed algorithm. Increased contour similarity in subsequently acquired imaging data highlights the observer-independence of the algorithm. While several observers outline different volumes, in identical data sets as well as in subsequent imaging data sets, the algorithm outlines more stable volumes in both cases. Thus, increased contour reproducibility is reached by automation of the delineation process by the proposed algorithm.

Bibliography

Abolmaali N, Zöphel K, Koch A, Haase R, Steinbach J, Baumann M. 2010. Wie weit bewegen sich die Schwerpunkte hypoxischer Tumorsubvolumina von HNO-Tumoren unter kombinierter Radiochemotherapie. Nuklearmedizin, 49:A80.

Abolmaali N, Haase R, Koch A, Zips D, Steinbach J, Baumann M, Kotzerke J, Zöphel K. 2012. Two or four hour [^{18}F]FMISO-PET in HNSCC. When is the contrast best? Nuklearmedizin, 50(1):22–27.

Abramyuk A, Tokalov S, Zöphel K, Koch A, Lazanyi K Szluha, Gillham C, Herrmann T, Abolmaali N. 2009. Is pre-therapeutical FDG-PET/CT capable to detect high risk tumor subvolumes responsible for local failure in non-small cell lung cancer? Radiotherapy and Oncology, 91:399–404.

Altman DG, Bland JM. 1983. Measurements in Medicine: The Analysis of Method Comparison Studies. Journal of the Royal Statistical Society. Series D (The Statistican), 32(3):307–317.

Baterina AV, Oppus C. 2010. Image Edge Detection Using Ant Colony Optimization. International Journal of Circuits, Systems and Signal Processing, 4(2):25–33.

Baumann M, Beuthien-Baumann B, Zips D. 2008. PET in der Strahlentherapie: Perspektiven. Der Nuklearmediziner, 31:8–11.

Bazañez-Borgert M, Bundschuh RA, Herz M, Martínez M-J, Schwaiger M, Ziegler SI. 2008. Radioactive spheres without inactive wall for lesion simulation in PET. Zeitschrift für Medizinische Physik, 18(1): 37–420.

Bentzen SM, Gregoire V. 2011. Molecular Imaging-Based Dose Painting: A Novel Paradigm for Radiation Therapy Prescription. Seminars in Radiation Oncology, 21(2):101–110.

Biehl K J, Kong F-M, Dehdashti F, Jin J-Y, Mutic S, El Naqa I, Siegel B A, Bradley J D. 2006. 18F-FDG PET Definition of Gross Tumor Volume for Radiotherapy of Non-Small Cell Lung Cancer: Is a Single Standardized Uptake Value Threshold Approach Appropriate? Journal of Nuclear Medicine, 47(11): 1808–1812.

Bittner M-I, Wiedemann N, Bucher S, Hentschel M, Mix M, Weber WA, Grosu A-L. 2013. Exploratory geographical analysis of hypoxic subvolumes using ^{18}F-MISO-PET imaging in patients with head and neck cancer in the course of primary chemoradiotherapy. Radiotherapy and Oncology, 108(3):511–516.

Black QC, Grills IS, Kestin LL, Wong C-YO, Wong JW, Martinez AA, Yan D. 2004. Defining a radiotherapy target with positron emission tomography. International Journal of Radiation Oncology Biology Physics, 60(4):1272–1282.

Bonabeau E, Dorigo M, Theraulaz G. 2000. Inspiration for optimization from social insect behaviour. Nature, 406(6791):39–42.

Brambilla M, Secco C, Dominietto M, Matheoud R, Sacchetti G, Inglese E. 2005. Performance characteristics obtained for a new 3-dimensional lutetium oxyorthosilicate-based whole-body PET/CT scanner with the National Electrical Manufacturers Association NU 2-2001 standard. Journal of Nuclear Medicine, 46(12):2083–2091.

Busk M, Horsman MR, Overgaard J. 2008. Resolution in PET hypoxia imaging: Voxel size matters. Acta Oncologica, 47(7):1201–1210.

Caldwell CB, Mah K, Ung YC, Danjoux CE, Balogh JM, Ganguli SN, Ehrlich LE. 2001. Observer variation in contouring gross tumor volume in patients with poorly defined non-small cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion. International Journal of Radiation Oncology Biology Physics, 51(4): 923–931.

Cao H, Huang P, Luo S. 2008. A Novel Image Segmentation Algorithm Based on Artificial Ant Colonies. Medical Imaging and Informatics, pages 63–71.

Cerello P, Cheran SC, Bagnasco S, Bellotti R, Bolanos L, Catanzariti E, De Nunzio G, Fantacci ME, Fiorina E, Gargano G, Gemme G, Torres EL, Masala GL, Peroni C, Santoro M. 2010. 3-D object segmentation using ant colonies. Pattern Recognition, 43:1476–1490.

Chang JH, Wada M, Anderson NJ, Joon DL, Lee ST, Gong SJ, Gunawardana DH, Sachinidis J, O'Keefe G, Gan HK, Khoo V, Scott AM. 2013. Hypoxia-targeted radiotherapy dose painting for head and neck cancer using ^{18}F-MISO-PET: A biological modeling study. Acta Oncologica, 52(8):1723–1729.

Cheebsumon P, van Velden FHP, Yaqub M, Frings V, de Langen AJ, Hoekstra OS, Lammertsma AA, Boellaard R. 2011a. Effects of Image Characteristics on Performance of Tumor Delineation Methods: A Test–Retest Assessment. Journal of Nuclear Medicine, 51:1550–1558.

Cheebsumon P, Yaqub M, van Velden FHP, Hoekstra OS, Lammertsma AA, Boellaard R. 2011b. Impact of [^{18}F]FDG PET imaging parameters on automatic tumour delineation: need for improved tumour delineation methodology. European Journal of Nuclear Medicine and Molecular Imaging, 38:2136–2144.

Chialvo D, Millonas MM. 1995. How Swarms Build Cognitive Maps. In Steels L, ed, The Biology and Technology of Intelligent Autonomous Agents, 144, NATO ASI Series, pages 439–450.

Colorni A, Dorigo M, Maniezzo V. 1992. Distributed Optimization by Ant Colonies. In European Conference on Artificial Life, pages 134–142. MIT Press, Cambridge.

Daisne J-F, Sibomana M, Bol A, Doumont T, Lonneux M, Grégoire V. 2003. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiotherapy and Oncology, 69(3):247–250.

Daisne J-F, Duprez T, Weynand B, Lonneux M, Harmoir M, Reychler H, Grégoire V. 2004. Tumor Volume in Pharyngolaryngeal Squamous Cell Carcinoma: Comparison at CT, MR Imaging, and FDG PET and Validation with Surgical Specimen. Radiology, 233(1):93–100.

Devic S, Tomic N, Faria S, Menard S, Lisbona R, Lehnert S. 2010. Defining a radiotherapy target using 18F-fluoro-deoxy-glucose positron emission tomography/computed tomography: still a Pandora's box? International Journal of Radiation Oncology Biology Physics, 78(5):1555–62.

Dewalle-Vignion AS, Yeni N, Petyt G, Verscheure L, Huglo D, Béron A, Adib S, Lion G, Vermandel M. 2012. Evaluation of PET volume segmentation methods: comparisons with expert manual delineations. Nuclear Medicine Communications, 33(1):34–42.

Dice LR. 1945. Measures of the Amount of Ecologic Association Between Species. Ecology, 26(3): 297–302.

Dirix P, Vandecaveye V, De Keyzer F, Stroobants S, Hermanns R, Nuyts S. 2009. Dose Painting in Radiotherapy for Head and Neck Squamous Cell Carcinoma: Value of Repeated Functional Imaging with 18F-FDG PET, 18F-Fluoromisonidazol PET, Diffusion-Weighted MRI and Dynamic Contrast-Enhanced MRI. Journal of Nuclear Medicine, 50:1020–1027.

Dorigo M. 2001. Ant Algorithms Solve Difficult Optimization Problems. In Kelemen J, Sosik P, eds, ECAL '01: Proceedings of the 6th European Conference on Advances in Artificial Life, pages 11–22, London, UK. Springer-Verlag.

Dorigo M, Birattari M, Stützle T. 2006. Ant colony optimization. IEEE Computational Intelligence Magazine.

Dubois L, Landuyt W, Cloetens L, Bol A, Bormans G, Haustermans K, Labar D, Nuyts J, Grégoire V, Mortelmans L. 2009. [^{18}F]EF3 is not superior to [^{18}F]FMISO for PET-based hypoxia evaluation as measured in a rat rhabdomyosarcoma tumour model. European Journal of Nuclear Medicine and Molecular Imaging, 36:209–218.

Eschmann S-M, Paulsen F, Reimold M, Dittmann H, Welz S, Reischl G, Machulla H-J, Bares R. 2005. Prognostic Impact of Hypoxia Imaging with 18F-Misonidazole PET in Non-Small Cell Lung Cancer and Head and Neck Cancer Before Radiotherapy. Journal of Nuclear Medicine, 46(2):253–260.

Eschmann S-M, Paulsen F, Bedeshem C, Machulla H-J, Hehr T, Bamberg M, Bares R. 2007. Hypoxia-imaging with 18F-Misonidazole and PET: Changes of kinetics during radiotherapy of head-and-neck cancer. Radiotherapy and Oncology, 83(3):406–410.

Fernandes C, Ramos V, Rosa AC. 2005a. Self-Regulated Artificial Ant Colonies on Digital Image Habitats. International Journal of Lateral Computing, 2:1–8.

Fernandes C, Ramos V, Rosa AC. 2005b. Varying the Population Size of Artificial Foraging Swarms on Time Varying Landscapes. In Duch W, Oja E, Zadrozny S, eds, ICANN´05: 15th International Conference of Artificial Neural Networks: Biological Inspirations, Warsaw, Poland, pages 311–316. Springer-Verlag.

Geets X, Lee JA, Bol A, Lonneux M, Grégoire V. 2007. A gradient-based method for segmenting FDG-PET images: methodology and validation. European Journal of Nuclear Medicine and Molecular Imaging, 34(9):1427–1438.

Goss S, Aron S, Deneubourg J-L, Pasteels JM. 1989. Self-organized shortcuts in the Argentine ant. Naturwissenschaften, 76:578–581.

Grgic A, Nestle U, Schaefer-Schuler A, Kremp S, Kirsch CM, Hellwig D. 2009. FDG-PET-based radiotherapy planning in lung cancer: optimum breathing protocol and patient positioning - an intraindividual comparison. International Journal of Radiation Oncology Biology Physics, 73(1):103–111.

Grosu AL, Souvatzoglou M, Röper B, Dobritz M, Wiedenmann N, Jacob V, Wester HJ, Reischl G, Machulla HJ, Schwaiger M, Molls M, Piert M. 2007. Hypoxia imaging with FAZA-PET and theoretical considerations with regard to dose painting for individualization of radiotherapy in patients with head and neck cancer. International Journal of Radiation Oncology Biology Physics, 69(2):541–551.

Gutjahr WJ. 2000. A Graph-based Ant system and its convergence. Future Generation Computer Systems, 16(9):873–888.

Haase R. 2010. Entwicklung eines Segmentierungsverfahrens für die Niedrigkontrast-Positronen-Emissions-Tomographie. University of Applied Sciences, HTW Dresden, Diploma thesis.

Haase R, Hietschold V, Andreeff M, Böhme H-J, Abolmaali N. 2010. A New Segmentation Algorithm for Low Contrast Positron Emission Tomography based on Ant Colony Optimization. In Crossing the Borders with ABC, Automation Biomedical Engineering and Computer Science, Proceeding of the 55th IWK, TU Ilmenau, pages 505–510.

Haase R, Böhme H-J, Zips D, Abolmaali N. 2011a. Swarm Intelligence for Medical Volume Segmentation: The Contribution of Self-reproduction. In Bach J, Edelkamp S, eds, KI2011: Advances in Artificial Intelligence, volume 7006 of Lecture Notes on Computer Science, pages 111–121. Springer Berlin Heidelberg.

Haase R, Hietschold V, Andreeff M, Böhme H-J, Abolmaali N. 2011b. Virtuelle Ameisen zur Segmentierung von Positronen-Emissions-Tomographie Daten: Auswirkungen auf die Receiver Operating Characteristic. In proceedings of 92. Deutscher Röntgenkongress und 6. Gemeinsame Jahrestagung DRG und ÖRG, Hamburg.

Haase R, Böhme H-J, Perrin R, Zöphel K, Abolmaali N. 2012a. Self-reproduction versus Transition Rules in Ant Colonies for Medical Volume Segmentation. In Dorigo M, Birattari M, Blum C, Christensen AL, Engelbrecht AP, GroßR, Stützle T, eds, ANTS2012, Eigth International Conference on Swarm Intelligence, Brussels, volume 7461 of Lecture Notes on Computer Science, pages 316–323. Springer Berlin Heidelberg.

Haase R, Zöphel K, Andreeff M, Steinbach J, Kotzerke J, Abolmaali N. 2012b. Usefulness of voxel-wise Pearson correlation coefficient for reproducibility and repeatability measurements in PET. Nuklearmedizin, 51:P27.

Haase R, Perrin R, Zöphel K, Böhme H-J, Abolmaali N. 2013a. Ant Based Segmentation of Positron Emission Tomography Imaging Data for Tumour Hypoxia Classification. In proceedings of Medical Imaging using Bio-Inspired and Soft-Computing (MIBISOC), Brussels.

Haase R, Zöphel K, Perrin R, Böhme H-J, Andreeff M, Kotzerke J, Abolmaali N. 2013b. Manual versus automatic delineation of [^{18}F]fluoromisonidazole PET data sets of head and neck cancer patients. Nuklearmedizin, 52:V54.

Haase R, Lohaus F, Rühaak J, van Straaten D, Zöphel K, Richter C. 2014. Einfluss deformierbarer Bildregistrierung von PET/CT-Verlaufsdaten auf die Signalwerte innerhalb einer Kontur: Deformation auf Konturen oder Bilddaten anwenden? In Proceedings of 20. Jahrestagung der Deutschen Gesellschaft für Radioonkologie (DEGRO), Düsseldorf.

Haberkorn U, Strauss LG, Dimitrakopoulou A, Seiffert E, Oberdorfer F, Ziegler S, Reisser C, Doll J, Helus F, van Kaick G. 1993. Fluorodeoxyglucose Imaging of Advanced Head and Neck Cancer After Chemotherapy. The Journal of Nuclear Medicine, 34(1):12–17.

Hamming RW. 1950. Error Detecting and Error Correcting Codes. The Bell System Technical Journal, 26(2):147–160.

Hamming RW. 1962. Numerical Methods for Scientists and Engineers. McGraw-Hill.

Han D, Bayouth J, Song Q, Taurani A, Sonka M, Buatti J, Wu X. 2011. Globally Optimal Segmentation in PET-CT Images: Graph-Based Co-Segmentation Method. Information Processing in Medical Imaging, 22:245–256.

Handl J, Knowles J, Dorigo M. 2006. Ant-based clustering and topographic mapping. Artificial Life, 12 (1):35–62.

Hatt M, Le Rest CC, Turzo A, Roux C, Visvikis D. 2009. A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET. IEEE Transations on Medical Imaging, 28(6):881–893.

Hatt M, Le Rest CC, Albarghach N, Pradier O, Visvikis D. 2011. PET functional volume delineation: a robustness and repeatability study. European Journal of Nuclear Medicine and Molecular Imaging, 38: 663–672.

Hausdorff F. 1914. Grundzüge der Mengenlehre. Veit and Co, Leipzig.

Höckel M, Knoop C, Schlenger K, Vorndran B, Baussmann E, Mitze M, Knapstein P G, Vaupel P. 1993. Intratumoral pO_2 predicts survival in advanced cancer of the uterine cervix. Radiotherapy and Oncology, 26(1):45–50.

Hofheinz F, Dittrich S, Pötzsch C, van den Hoff J. 2010. Effects of cold sphere walls in PET phantom measurements on the volume reproducing threshold. Physics in Medicine and Biology, 55(4):1099–1113.

Hofheinz F, Langner J, Beuthien-Baumann B, Oehme L, Steffen I, Apostolova I, Steinbach J, Zöphel K, Kotzerke J, van den Hoff J. 2012a. Automatisierte volumetrische Abgrenzung heterogener Tumoren in der PET. Nuklearmedizin, 51:V41.

Hofheinz F, Langner J, Petr J, Beuthien-Baumann B, Oehme L, Steinbach J, van den Hoff J. 2012b. A method for model-free partial volume correction in oncological PET. European Journal of Nuclear Medicine and Molecular Imaging Research, 2(16], pages=1-12).

Hofheinz F, Pötzsch C, Oehme L, Beuthien-Baumann B, Steinbach J, Kotzerke J, van den Hoff J. 2012c. Automatic volume delineation in oncological PET. Evaluation of a dedicated software tool and comparison with manual delineation in clinical data sets. Nuklearmedizin, 51(1):9–16.

Hölldobler B, Wilson E O. 2009. The superorganism: the beauty, elegance, and strangeness of insect societies. W. W. Norton.

Huang P, Cao H, Luo S. 2008. An artificial ant colonies approach to medical image segmentation. Computer Methods and Programs in Biomedicine, 92(3):267–273.

Jevtić A, Quintanilla-Domínguez J, Barrón-Adame J, Andina D. 2011. Image Segmentation Using Ant System-Based Clustering Algorithm. In Corchado E, Snásel V, Sedano J, Hassanien A, Calvo J, Slezak D, eds, Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011, volume 87 of Advances in Intelligent and Soft Computing, pages 35–45. Springer Berlin / Heidelberg.

Ji J, Huang Z, Wang Y, Liu C. 2008. A New Mechanism of Pheromone Increment and Diffusion for Solving Travelling Salesman Problems with Ant Colony Algorithm. International Conference on Natural Computation, 7:558–563.

Lee NY, Mechalakos JG, Nehmeh SA, Lin Z, Squire OD, Cai S, Chan K, Zanzonico PB, Greco C, Ling CC, Humm JL, Schröder H. 2008. Fluorine-18-Labeled Fluoromisonidazole Positron Emission and Computed Tomography-Guided Intensity-Modulated Radiotherapy for Head and Neck Cancer: A Feasibility Study. International Journal of Radiation Oncology Biology Physics, 70(1):2–13.

Li C, Wang X, Xia Y, Eberl S, Yin Y, Feng DD. 2012. Automated PET-guided liver segmentation from low-contrast CT volumes using probabilistic atlas. Computer Methods and Programs in Biomedicine, 107 (2):164–174.

Ma L, Wang K, Zhang D. 2009. A universal texture segmentation and representation scheme based on ant colony optimization for iris image processing. Computers and Mathematics with Applications, 57: 1862–1868.

Mahy P, de Bast M, de Groot T, Cheguillaume A, Gillart J, Haustermans K, Labar D, Grégoire V. 2008. Comparative pharmacokinetics, biodistribution, metabolism and hypoxia-dependent uptake of [18F]-EF3 and [18F]-MISO in rodent tumor models. Radiotherapy and Oncology, 89(3).

Mortensen LS, Johansen J, Kallehauge J, Primdahl H, Busk M, Lassen P, Alsner J, Sørensen BS, Toustrup K, Jakobsen S, Petersen J, Petersen H, Theil J, Nordsmark M, Overgaard J. 2012. FAZA PET/CT hypoxia imaging in patients with squamous cell carcinoma of the head and neck treated with radiotherapy: results from the DAHANCA 24 trial. Radiotherapy and Oncology, 105(1):14–20.

Mullen RJ, Monekosso D, Barman S, Remagnino P, Wilkin P. 2008. Artificial Ants to Extract Leaf Outlines and Primary Venation Patterns. In Dorigo M, Birattari M, Blum C, Clerc M, Stützle T, Winfield A F T, eds, ANTS 2008, LNCS 5217, page 251–258.

Nehmeh SA, Lee NY, Schröder H, Squire O, Zanzonico PB, Erdi YE, Greco C, Mageras G, Pham HS, Larson SM, Ling CC, Humm JL. 2008. Reproducibility of Intratumor Distribution of 18F-Fluoromisonidazole in Head and Neck Cancer. International Journal of Radiation Oncology Biology Physics, 70(1):235–242.

Nehmeh SA, El-Zeftawy H, Greco C, Schwartz J, Erdi YE, Kirov A, Schmidtlein CR, Gyau AB, Larson SM, Humm JL. 2009. An iterative technique to segment PET lesions using a Monte Carlo based mathematical model. Medical Physics, 36(10):4803–4809.

Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rübe C, Kirsch C-M. 2005. Comparison of Different Methods for Delineation of 18F-FDG PET–Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non–Small Cell Lung Cancer. Journal of Nuclear Medicine, 46(8):1342–1348.

Okamoto S, Shiga T, Yasuda K, Ito YM, Magota K, Kasai K, Kuge Y, Shirato H, Tamaki N. 2013. High Reproducibility of Tumor Hypoxia Evaluated by ^{18}F-Fluoromisonidazole PET for Head and Neck Cancer. Journal of Nuclear Medicine, 52(2):201–207.

Platzek I, Beuthien-Baumann B, Schneider M, Gudziol V, Langner J, Schramm G, Laniado M, Kotzerke J, van den Hoff J. 2013. PET/MRI in head and neck cancer: initial experience. European Journal of Nuclear Medicine and Molecular Imaging, 40(1):6–11.

Poncelet AJ, Lonneux M, Coche E, Weynand B, Noirhomme P, d'Oncologie Thoracique des Cliniques Saint-Luc Groupe. 2001. Pet-fdg scan enhances but does not replace preoperative surgical staging in non-small cell lung carcinoma. European Journal Cardiothorac Surgery, 20(3):468–475.

Rajendran JG, Schwartz DL, O'Sullivan J, Peterson LM, Ng P, Scharnhorst J, Grierson JR, Krohn KA. 2006. Tumor Hypoxia Imaging with [F-18] Fluoromisonidazole Positron Emission Tomography in Head and Neck Cancer. Clinical Cancer Research, 12(18):5435–5441.

Ramos V, Almeida F. 2000. Artificial Ant Colonies in Digital Image Habitats - A Mass Behaviour Effect Study on Pattern Recognition. In Dorigo M, Middendorf M, Stüzle T, eds, From Ant Colonies to Artificial Ants - 2nd International Workshop on Ant Algorithms, pages 113–116.

Reischl G, Dorow DS, Cullinane C, Katsifis A, Roselt P, Binns D, Hicks RJ. 2007. Imaging of tumor hypoxia with [124I]IAZA in comparison with [18F]FMISO and [18F]FAZA - first small animal PET results. Journal of Pharmaceutical Sciences, 10(2):203–211.

Schaefer A, Kremp S, Hellwig D, Rübe C, Kirsch CM, Nestle U. 2008. A contrast-oriented algorithm for FDG-PET-based delineation of tumour volumes for the radiotherapy of lung cancer: derivation from phantom measurements and validation in patient data. European Journal of Nuclear Medicine and Molecular Imaging, 35(11):1989–1999.

Schaefer A, Nestle U, Kremp S, Hellwig D, Grgic A, Buchholz HG, Mischke W, Gromoll C, Dennert P, Plotkin M, Senftleben S, Thorwarth D, Tosch M, Wahl A, Wengenmair H, Rübe C, Kirsch CM. 2012. Multi-centre calibration of an adaptive thresholding method for PET-based delineation of tumour volumes in radiotherapy planning of lung cancer. Nuklearmedizin, 51(3):101–110.

Schwartz J, Humm JL, Gonen M, Klaigian H, Schroder H, Larson SM, Nehmeh SA. 2011. Repeatability of SUV measurements in serial PET. Medical Physics, 38(5):2629–2638.

Shannon CE. 1998. Communication in the Presence of Noise. In PROCEEDINGS OF THE IEEE, volume 86, pages 447–457.

Shepherd T, Teräs M, Beichel RR, Boellaard R, Bruynoooghe M, Dicken V, Gooding MJ, Julyan PJ, Lee JA, Lefèfre S, Mix M, Naranjo V, Wu X, Zaidi H, Zeng Z, Minn H. 2012. Comparative Study With New Accuracy Metrics for Target Volume Contouring in PET Image Guided Radiation Therapy. IEEE Transactions on Medical Imaging, 31(11):2006–2024.

Sørensen TJ. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. Biologiske Skrifter, 5:1–34.

Souvatzoglou M, Grosu AL, Röper B, Krause BJ, Beck R, Reischl G, Picchio M, Machulla HJ, Wester HJ, Piert M. 2007. Tumour hypoxia imaging with [18F]FAZA PET in head and neck cancer patients: a pilot study. European Journal of Nuclear Medicine and Molecular Imaging, 34(10):1566–1575.

Spratt DE, Diaz R, McElmurray J, Csiki I, Duggan D, Lu B, Delbeke D. 2010. Impact of FDG PET/CT on delineation of the gross tumor volume for radiation planning in non-small-cell lung cancer. Clinical Nuclear Medicine, 35(4):237–243.

Steenbakkers RJHM, Duppen JC, Fitton I, Deurloo KEI, Zijp L, Uitterhoeve ALJ, Rodrigus PTR, Kramer GWP, Bussink J, De Jaeger K, Belderbos JSA, Hart AAM, Novak PJCM, van Herk M, Rasch CRN. 2005. Observer variation in target volume definition of lung cancer related to radiation oncologist-computer interaction: A 'Big Brother' evaluation. Radiotherapy and Oncology, 77:182–190.

Stützle T, Dorigo M. 2002. A short convergence proof for a class of Ant Colony Optimization algorithms. IEEE Transactions on Evolutionary Computation, 6(4):358–365.

Swanson KR, Chakraborty G, Wang CH, Rockne R, Harpold HLP, Muzi M, Adamsen TCH, Krohn KA, Spence AM. 2009. Complementary but Distinct Roles for MRI and ^{18}F-Fluoromisonidazole PET in the Assessment of Human Glioblastomas. Journal of Nuclear Medicine, 50:36–44.

Tachibana I, Nishimura Y, Shibata T, Kanamori S, Nakamatsu K, Koike R, Nishikawa T, Ishikawa K, Tamura M, Hosono M. 2013. A prospective clinical trial of tumour hypoxia imaging with ^{18}F-fluoromisonidazole positron emission tomography and computed tomography (F-MISO-PET/CT) before and during radiation therapy. Journal of Radiation Research, 54(6):1078–1084.

Thie JA. 2004. Understanding the Standardized Uptake Value, Its Methods, and Implications for Usage. Journal of Nuclear Medicine, 45(9):1431–1434.

Thorwarth D, Alber M. 2008. Funktionelle Hypoxie-Bildgebung mit FMISO PET zur Individualisierung der Strahlentherapie. Zeitschrift für Medizinische Physik, 18(1):43–50.

Thorwarth D, Alber M. 2010. Implementation of hypoxia imaging into treatment planning and delivery. Radiotherapy and Oncology, 97:172–175.

Thorwarth D, Eschmann S-M, Scheiderbauer J, Paulsen F, Alber M. 2005. Kinetic analysis of dynamic 18F-fluoromisonidazole PET correlates with radiation treatment outcome in head-and-neck cancer. BMC Cancer, 5(1):152.

Thorwarth D, Eschmann S-M, Holzner F, Paulsen F, Alber M. 2006. Combined uptake of [18F]FDG and [18F]FMISO correlates with radiation therapy outcome in head-and-neck cancer patients. Radiotherapy and Oncology, 80(2):151–156.

Thorwarth D, Eschmann S-M, Paulsen F, Alber M. 2007. Hypoxia Dose Painting by Numbers: A Planning Study. International Journal of Radiation Oncology Biology Physics, 68:291–300.

Tylski P, Stute S, Grotus N, Doyeux K, Hapdey S, Gardin I, Vanderlinden B, Buvat I. 2010. Comparative Assessment of Methods for Estimating Tumor Volume and Standardized Uptake Value in 18F-FDG PET. Journal of Nuclear Medicine, 51(2):268–276.

Valk PE, Bailey DL, Townsend DW, Maisey MN. 2004. Positron Emission Tomography, Basic Science and Clinical Practice. Springer.

van den Hoff J, Hofheinz F. 2012. Phantom measurements with glass inserts in a hot background are not suitable for performance assessment of volume delineation algorithms in PET. IEEE Transactions on Medical Imaging, 32(6):1146–1148.

van den Hoff J, Hofheinz F, Oehme L, Schramm G, Langner J, Beuthien-Baumann B, Steinbach J, Kotzerke J. 2013. Dual time point based quantification of metabolic uptake rates in 18F-FDG PET. European Journal of Nuclear Medicine and Molecular Imaging Research, 3(16).

Vanderhoek M, Perlman SB, Jeraj R. 2012. Impact of the Definition of Peak Standardized Uptake Value on Quantification of Treatment Response. Journal of Nuclear Medicine, 53(1):4–11.

Weichert S. 2011. Bewertung von Vergleichsgrößen für Segmentierungsverfahren der Positronen-Emissions-Tomografie. University of Applied Sciences, HTW Dresden, Diploma thesis.

Westgard JO. 1998. Points of Care in Using Statistics in Method Comparison Studies. Clinical Chemistry, 44(11):2240–2242.

Westgard JO. 2008. Use and Interpretation of Common Statistical Tests in Method Comparison Studies. Clinical Chemistry, 54(3):612.

Westgard JO, Hunt MR. 1973. Use and Interpretation of Common Statistical Tests in Method-Comparison Studies. Clinical Chemistry, 19(1):49–57.

Xia Y, Eberl S, Wen L, Fulham M, Feng DD. 2012. Dual-modality brain PET-CT image segmentation based on adaptive use of functional and anatomical information. Computerized Medical Imaging and Graphics, 36(1):47–53.

Yasuda K, Onimaru R, Okamoto S, Shiga T, Katoh N, Tsuchiya K, Suzuki R, Takeuchi W, Kuge Y, Tamaki N, Shirato H. 2012. [^{18}]F-fluoromisonidazole and a New PET System With Semiconductor Detectors and a Depth of Interaction System for Intensity Modulated Radiation Therapy for Nasopharyngeal Cancer. International Journal of Radiation Oncology Biology Physics, 85(1):142–147.

Youden WJ. 1950. Index for rating diagnostic tests. Cancer, 3(1):32–35.

Zips D, Zöphel K, Abolmaali N, Perrin R, Abramyuk A, Haase R, Appold S, Steinbach J, Kotzerke J, Baumann M. 2012. Exploratory prospective trial of hypoxia-specific PET imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer. Radiotherapy and Oncology, 105(1):21–28.

Zöphel K, Appold S, Abolmaali N. 2008. PET/CT-basierte Bestrahlungsplanung bei Patienten mit Kopf-Hals-Tumoren. Der Nuklearmediziner, 31(1):55–59.

A. Acknowledgements

Sometimes long lasting research leads to a text document which finally reaches the state of a ready-to-hand-in thesis. But it wouldn't come to this, if not many people have helped the one who did the research and wrote the document. Thus, I need to thank all of these people a lot. First of all, I would like to thank Prof. Dr. Nasreddin Abolmaali for his supervision during the years. When I started working as student assistant at the Biological and Molecular Imaging Group headed by him almost six years ago, I was not able to imagine all the things I would learn in this environment. He taught me not just knowledge about human anatomy and physiology, molecular and classical imaging techniques, research practice and how to write in a scientific style. Even more importantly, he showed me how to sustain in a research environment. While I was more and more leaving the classical working-field of a computer scientist, Prof. Dr. Hans-Joachim Böhme fetched me back a bit whenever I met him. I would like to thank him for keeping qualifying me as professional and academic computer scientist. When talking to clinicians and medical physicists the whole day, it was always a pleasure to come back to my former university the University of Applied Sciences HTW Dresden to talk with him about the algorithmic parts of my work. I also need to thank Prof. Dr. Walter Pätzold, because he was the one who told us students of computer science about the abilities of ant-based algorithms. Thus, he is partly responsible for the idea which years later became topic of this thesis. The list of people who further pushed the development of the proposed algorithm must be completed with Dr. Rosalind Perrin. She was the first scientist coming from outside to the OncoRay who suggested using the new algorithm in her own research projects. At this point the algorithm left the stage of development and entered clinical research. I would like to thank her for the trust in my source code.

This thesis is about PET image segmentation. It wouldn't have been possible without the support of the staff of the Clinic and Policlinic for Nuclear Medicine headed by Prof. Dr. Jörg Kotzerke. I would like to thank the medical technical assistants Gabriele Kotzerke and Kathrin Gericke who were in practice responsible for the image quality of the acquired data sets. Without their carefullness and patience, even on busy days, the acquired images wouldn't have been so proper. Furthermore, I would like to thank Dr. Michael Andreeff, who supported me in planning and performing the phantom experiments. Last but not least from the colleques of the Nuclear Medicine, I would like to thank PD Dr. Klaus Zöphel for his talent of pushing my thoughts into clinical applicability. Imaging of patients is not a purely technical process. The patients had to wait four hours from tracer injection to the final image acquisition and were lying in the scanner an hour alltogether. Thus, when I raised the idea of scanning longer or later after tracer injection, he taught me to think about the patients perspective to ideas like that. At this point, I would

also like to thank the patients who were part of the clinical study I was in charge to support with image analysis. As stated above, they needed a lot of patience when being scanned at the PET/CT. And this couldn't be harder, than if you are suffering from a serious disease.

I would like to thank some people from other research groups even though they were not directly involved in my project, such as Prof. Dr. Wolfgang Enghardt, Dr. Jörg Pawelke, Dr. Thomas Kormoll and Stephan Helmbrecht, who gave me scientific advice increasing the professionality of my work. Last but not least I would like to thank my colleques and roommates Anna Bandurska-Luque, Agnieszka Bos-Liedke, Julia Buckwar, Ellen Dickreuter, Claudia Förster, Christian Golnik, Annika Jakobi, Dr. Uwe Just, Dr. Daniela Kunath, Stefan Leger, Dr. Steffen Löck, Dr. Fabian Lohaus, Andreas Müller, Dr. Christian Richter, Matthias Stein, Dr. Katja Storch, Dr. Kristin Stützer, Prasad Thute, Falk Tillner, Kerstin Wetzig, Sebastian Zschaeck and Dr. Katja Zscheppang. I always had a pleasant time with all of you.

The author was supported by a grant from the Sächsische Landesexzellenzinitiative (project number 100066308) and the German Federal Ministry of Education and Research (BMBF-03ZIK042).

B. Erklärungen zur Eröffnung des Promotionsverfahrens

1. Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

2. Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten: Prof. Dr. Nasreddin Abolmaali, Dr. Michael Andreeff, Prof. Dr. Hans-Joachim Böhme, Julia Buckwar, Stephan Helmbrecht, Dr. Rosalind Perrin und PD Dr. Klaus Zöphel.

3. Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

4. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

5. Die Inhalte dieser Dissertation wurden in folgender Form veröffentlicht:

Haase R. 2010. Entwicklung eines Segmentierungsverfahrens für die Niedrigkontrast-Positronen-Emissions-Tomographie. University of Applied Sciences, HTW Dresden, Diploma thesis.

Haase R, Hietschold V, Andreeff M, Böhme H-J, Abolmaali N. 2010. A New Segmentation Algorithm for Low Contrast Positron Emission Tomography based on Ant Colony Optimization. In Crossing the Borders with ABC, Automation Biomedical Engineering and Computer Science, Proceeding of the 55th IWK, TU Ilmenau, pages 505-510.

Haase R, Böhme H-J, Zips D, Abolmaali N. 2011. Swarm Intelligence for Medical Volume Segmentation: The Contribution of Self-reproduction. In KI2011: Advances in Artificial Intelligence, volume 7006 of Lecture Notes on Computer Science, pages 111-121. Springer Berlin Heidelberg.

Haase R, Hietschold V, Andreeff M, Böhme H-J, Abolmaali N. 2011. Virtuelle Ameisen zur Segmentierung von Positronen-Emissions-Tomographie Daten: Auswirkungen auf die Receiver Operating Characteristic. In proceedings of 92. Deutscher Röntgenkongress und 6. Gemeinsame Jahrestagung DRG und ÖRG, Hamburg.

Haase R, Böhme H-J, Perrin R, Zöphel Z, Abolmaali N. 2012. Self-reproduction versus Transition Rules in Ant Colonies for Medical Volume Segmentation. In Dorigo M, Birattari M, Blum C, Christensen A L, Engelbrecht A P, GroßR, Stützle T, eds, ANTS2012, Eigth International Conference on Swarm Intelligence, Brussels, volume 7461 of Lecture Notes on Computer Science, pages 316-323. Springer Berlin Heidelberg.

Haase R, Zöphel Z, Andreeff M, Steinbach J, Kotzerke J, Abolmaali N. 2012. Usefulness of voxel-wise Pearson correlation coeffcient for reproducibility and repeatability measurements in PET. Nuklearmedizin, 51:P27.

Haase R, Perrin R, Zöphel Z, Böhme H-J, Abolmaali N. 2013. Ant Based Segmentation of Positron Emission Tomography Imaging Data for Tumour Hypoxia Classification. In proceedings of Medical Imaging using Bio-Inspired and Soft-Computing (MIBISOC), Brussels.

Haase R, Zöphel Z, Perrin R, Böhme H-J, Andreeff M, Kotzerke J, Abolmaali N. 2013. Manual versus automatic delineation of [^{18}F]fluoromisonidazole PET data sets of head and neck cancer patients. Nuklearmedizin, 52:V54.

Haase R, Perrin R, Zöphel Z, Böhme H-J, Abolmaali N. 2013. Schwarmintelligenz zur Segmentierung von FMISO-PET-Daten: Den Experten überlegen? 44. Jahrestagung der Deutschen Gesellschaft für Medizinische Physik (DGMP), Köln.

Haase R, Perrin R, Zöphel Z, Böhme H-J, Abolmaali N. 2013. Schwarmintelligenz zur Beurteilung der Tumorhypoxie in FMISO-PET-Aufnahmen. 58. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS), Lübeck.

6. Ich bestätige, dass es keine zurückliegenden erfolglosen Promotionsverfahren gab.

7. Ich bestätige, dass ich die Promotionsordnung der Medizinischen Fakultät der Technischen Universität Dresden anerkenne.

8. Ich habe die Zitierrichtlinien für Dissertationen an der Medizinischen Fakultät der Technischen Universität Dresden zur Kenntnis genommen und befolgt.

Ort, Datum

Unterschrift des Doktoranden

C. Erklärung zur Einhaltung rechtlicher Vorschriften

Hiermit bestätige ich die Einhaltung der folgenden aktuellen gesetzlichen Vorgaben im Rahmen meiner Dissertation

☒ das zustimmende Votum der Ethikkommission bei Klinischen Studien, epidemiologischen Untersuchungen mit Personenbezug oder Sachverhalten, die das Medizinproduktegesetz betreffen

Aktenzeichen der zuständigen Ethikkommission: EK166082004

☐ die Einhaltung der Bestimmungen des Tierschutzgesetzes

Aktenzeichen der Genehmigungsbehörde zum Vorhaben/zur Mitwirkung

.....

☐ die Einhaltung des Gentechnikgesetzes

Projektnummer

.....

☒ die Einhaltung von Datenschutzbestimmungen der Medizinischen Fakultät und des Universitätsklinikums Carl Gustav Carus.

Ort, Datum

Unterschrift des Doktoranden