UNIVERSITY OF WEST BOHEMIA
Faculty of Applied Sciences
Department of Cybernetics

AUTOMATIC LOCALIZATION AND CLASSIFICATION OF LIVER LESIONS

Doctoral Thesis

Author:
Ing. Tomáš Ryba

Supervisor:
Doc., Ing. Miloš Železný, Ph.D.

ZÁPADOČESKÁ UNIVERZITA V PLZNI
Fakulta aplikovaný věd
Katedra kybernetiky

AUTOMATICKÁ LOKALIZACE A KLASIFIKACE JATERNÍCH LÉZÍ

DISERTAČNÍ PRÁCE

Autor:
Ing. Tomáš Ryba

Školitel:
Doc. Ing. Miloš Železný, Ph.D.

Declaration

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Pilsen, October 5, 2016

. . . .
Tomáš Ryba

Acknowledgements

I would like to gratefully acknowledge my supervisor Doc. Ing. Miloš Železný, Ph.D. for his advice and mentoring and the Department of Cybernetics at the Faculty of Applied Sciences, University of West Bohemia for making this work possible. Special thanks belong to the Teaching Hospital Pilsen, which has provided me with anonymous data. I am very grateful to MUDr. Hynek Mírka, who has shown me the doctors' point of view and taught me the medical background and motivation.

The biggest thanks, of course, belong to my entire family for their support the whole long way: my beloved wife for her patience and my small daughters for taking my mind of the work. Thanks to them, the sleepless nights were much easier to bear. They always believed I could finish this work, even at times when I did not believe it myself.

Pilsen, October 5, 2016

....
Tomáš Ryba

Abstract

Computer-aided diagnostic (CAD) systems are widely used in technical and medical fields. Using the CAD systems in medicine allows the application of image processing methods as well as the methods of artificial intelligence. The purpose of the systems is to assists doctors. A radiologist needs to diagnose a great amount of image data, which is very focus-demanding work. Using a CAD system can support doctor's effectiveness in the sense of processing speed and/or accuracy.

The goal of the thesis is to develop a CAD system for the automatic localization and subsequent classification of liver lesions. Liver cancer is mostly diagnosed from a differential diagnosis that consists of analysis of two serial CT screening. Because the screenings are taken at the time interval of several seconds, data registration needs to be performed. In both series the liver region is found using a fully autonomous method based on the Grow Cut algorithm and the results obtained are further refined by a localized active contour method. The liver region is then analyzed and searched for lesions. The localization of the lesion is performed by Markov Random Fields initialized with a combination of saliency maps. The lesions found are then paired-up and classified by a decision tree.

The results prove that the system presented in this work can deal with the high diversity of medical data and is able to achieve interesting accuracy.

Keywords: Computer-aided diagnostic, liver, lesion, tumor, image segmentation, image registration, localization, classification.

Abstrakt

Počítačem asistované diagnostické (CAD) systémy jsou ve velké míře používány v technických a lékařských oborech. Jejich nasazení v lékařství umožňuje aplikaci metod zpracování obrazu, umělé inteligence apod. Cílem těchto systémů je asistovat lékařům při jejich každodení práci. Práce radiologa zahrnuje diagnostikování velkého množství obrazových dat, což je úloha velice náročná na pozornost. Nasazení CAD systému může lékaři pomoci dosáhnout vyšší efektivity ve smyslu rychlosti zpracování dat i ve smyslu rychlosti.

Cílem disertační práce je navržení CAD systému pro automatickou lokalizaci a následnou klasifikaci jaterních lézí. Rakovina jater bývá nejčastěji diagnostikována pomocí diferenciální diagnostiky, která zahrnuje analýsu dvou CT sérií. Jelikož jsou jednotlivé série snímany s časovým odstupem několik vteřin, je zapotřebí provést zpětnou registraci dat. V obou seriích jsou segmentována játra pomocí plně automatizovaného přístupu založeném na metodě Grow cut. Získané výsledky jsou dále zpřesněny pomocí lokalizovaných aktivních kontur. Výsledná oblast korespondující s jaterním parenchymem je následně analyzována pro výskyt lézí. Samotná lokalizace lézí je založena na Markovských náhodných polích, které jsou inicializovány pomocí map salience. Nalezunutá ložiska jsou následně spárována mezi jednotlivými sériemi a klasifikována pomocí rozhodovacího stromu.

Získané výsledky prokazují schopnost navrženého přístupu vypořádat se s obrovskou rozmanitostí lékařských dat, přičemž je dozaženo zajímavé přesnosti.

Klíčová slova: Počítačem asistovaná diagnostika, játra, léze, tumor, segmentace, registrace, lokalizace, klasifikace.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Image processing is a branch of computer science that is becoming more and more popular these days. Its area of application is very wide and can be used for output control in industry, in security systems for evaluating people's biometric, etc. A significant part of the field is the medical imaging where the methods of image processing and artificial intelligence are used to solve medical tasks such as liver segmentation or the localization of tumors. This work presents a computer-aided diagnostic system that fits the field of the medical imaging and its goal is to localize and identify liver tumors.

1.1 Medical motivation

Based on the statistics published by [American Cancer Society, 2016] and [World Cancer Research Fund International - American Institute for Cancer Research, 2015], liver cancer is the sixth most common cancer worldwide. In 2012, 782 000 new cases were diagnosed in the world; in 2016 it is estimated that almost 40 000 new cases will be diagnosed in the USA. Liver cancer is more common in men than women. The highest incidence is in Asia and Africa, the lowest in Europe and in Latin America. The incidence and the mortality are visualized in Figs. 1.1 and 1.2, respectively. The risk of the cancer increases with age, the most cases are diagnosed over the age of 75 as stated in [Ferlay et al., 2015]. However, in people from less developed countries in Asia and Africa, the disease can develop at a younger age - typically around 40 ([Ferlay et al., 2015], [Llovet et al., 2003]).

An insidious problem of liver cancer is that it does not produce any symptoms when at an early stage. This means that the disease is usually diagnosed at an advanced stage. This yields generally poor survival rates. For example, an European adult diagnosed with liver cancer between 2000 and 2007 has a 12 percent chance that he will survive the next five years (so called five-year relative survival rate).

The liver is the largest parenchymatose organ in the human body that performs multiple functions. This yields a high volume of blood that flows through the liver which makes the liver prone to the incidence of secondary tumors. The vast majority of the liver tumors originate in different organs as stated in [Válek et al., 2006]. For successful treatment, the tumors needs to be localized and identified.
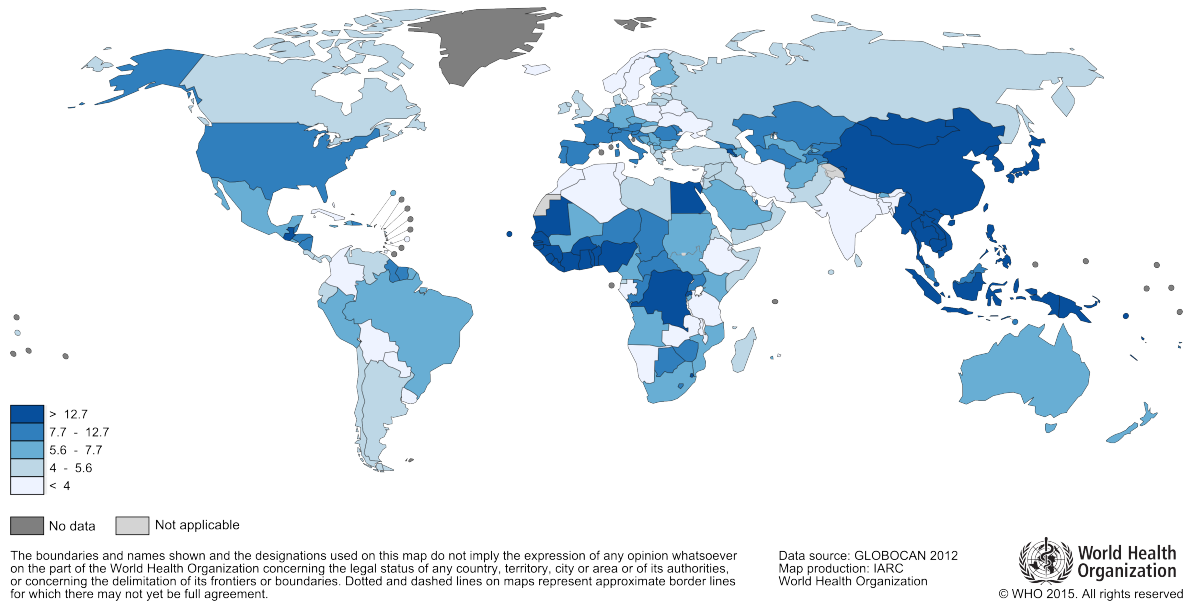
(a)
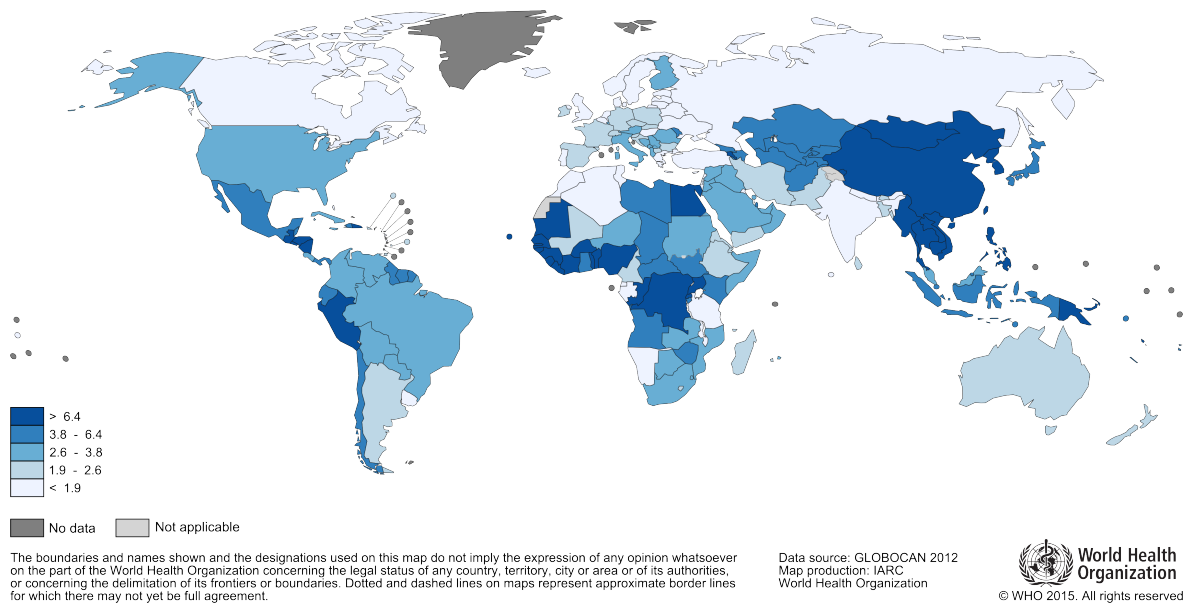
(b)

Figure 1.1: Liver cancer incidence in men (a) and women (b).

Data source: GLOBOCAN 2012
Map production: IARC
World Health Organization

World Health
Organization

(a)

Data source: GLOBOCAN 2012
Map production: IARC
World Health Organization

World Health
Organization
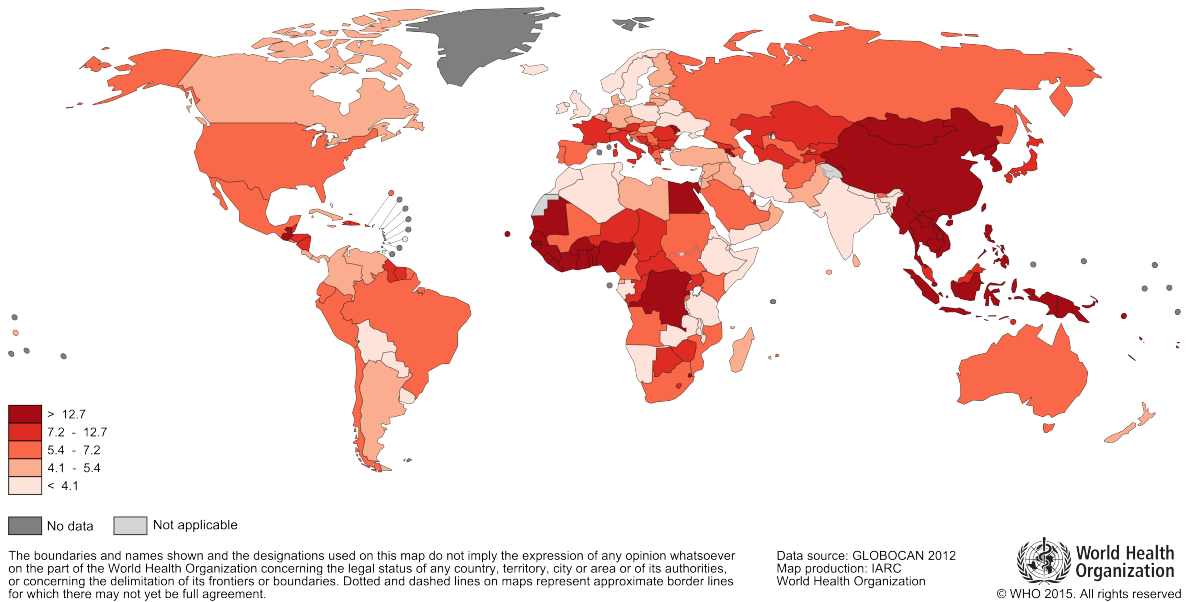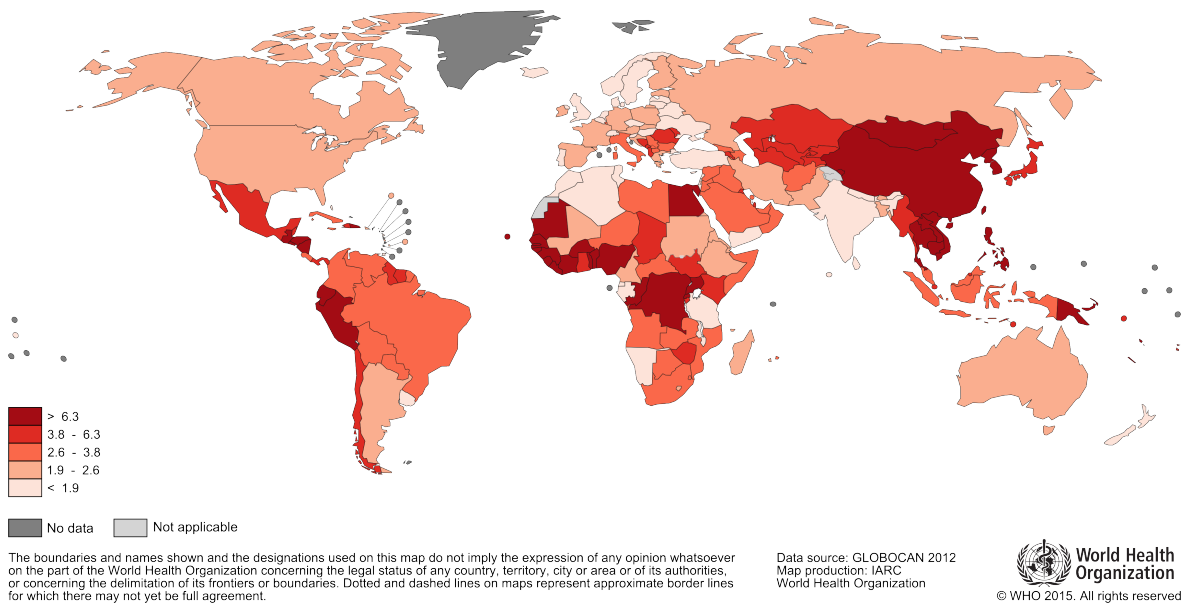
(b)

Figure 1.2: Liver cancer mortality in men (a) and women (b).

1.2   CAD Systems

A common task of a radiologist is to create a measurement that characterizes an object in the data, e.g. an organ or a tumor. Such a measurement could be a definition of the object's size, its mean density or relative position. Most of the commonly used softwares at hospitals include tools for making such measurements but these tools are often not user friendly or require vast amount of user interaction. For example, to segment a tumor the doctor needs to manually localize it by a mouse click. Despite the inefficiency, the process is highly subjective - the same doctor often marks the same object differently, yielding slightly different results.

These are just a few examples that created the impulse for developing a new field in the medical imaging, called computer-aided diagnostic (CAD). The CAD allows developing complex systems that help and support the doctors in the interpretation of the data or directly assists them when determining a diagnosis. The CAD systems make use of methods from image processing and artificial intelligence and together with the doctors' expert knowledge achieve very interesting results. The role of the CAD systems is definitely not to replace the work of the doctors; they are only supporting them and the final decision is always on a human doctor. More information on the CAD systems can be found for example in [Doi, 2007].

1.3   Goals of the thesis

The main goal of this thesis is to develop a system for localization and classification of liver tumors. This is a complex task and solving it means solving more subtasks. Cancer is diagnosed mostly from a CT and using the so called differential diagnosis. It is a serial screening of a patient with a time interval of a few seconds. To be able to spatially analyse both types of data, the system needs to register the data first. The registration task is described in Chapter 5. The next step is to segment the liver parenchyma in both data. Only then the segmented parenchyma can be analysed and searched for tumors. The final step is to pair up the localized tumors and identify them.

The main contributions of this work can be summarised into the following points:

1. Automatic liver segmentation . . . a fully autonomous system that does not require any user interaction.

2. Lesion localization . . . a fully autonomous approach that is able to deal with the nature of the medical images.

3. Lesion classification . . . definition of discriminative features whose application yields high classification accuracy.

The developed algorithm for the liver segmentation needs to be robust enough to deal with the nature and originality of each data. The liver has only approximately the same shape; on the other hand they can drastically differ. Moreover, the density of the liver parenchyma is different from patient to patient and depends on the patient's health condition and the

screening intervals. The approach for the liver segmentation presented in this work is area-based. The cellular automaton called grow cut is used and due to the problems described, it is initialized by analysing the gray-level co-occurrence matrix (GLCM) that combines spatial and color features. This way, the data are separated into individual blobs and the blob that corresponds to the liver is identified. To increase the accuracy of liver segmentation the results are refined by localized active contours. The liver segmentation is described in Chapter 4.

The tumors are often poorly distinguishable from their background. The main problem that must be solved is therefore to design an approach that would be sensitive enough to find such tumors. Simultaneously, the approach cannot produce many false positive detections. Methods based only on areas or edges often fail to deal with this task. Therefore, an algorithm using the markov random fields (MRF) that combines both of the approaches was defined. The goal is to define a fully autonomous algorithm; therefore, the MRF are initialized by saliency maps. The total number of seven different saliency maps was defined that extract different image features. To achieve the best accuracy, a combination of these saliency maps was used and the resulting mask was again further refined by the localized active contours. The lesion localization is described in Section 6.1.

Probably the biggest problem associated with the task is the diversity of medical data that is most evident in the last step. The same type of lesions can appear absolutely different depending on the health condition of the patient, the screening protocol and the progress of the disease. Despite all the effort it is hard to classify an atypical tumor that looks exactly like a representative of a different tumor type. Unfortunately, this situation occurs relatively frequently. Another problem which affects the classification accuracy is the lack of data. The analysed dataset is not big enough to train a modern classifier like (convolutional) neural networks. Therefore, a solution in the form of building a decision tree was used. Each level of the tree compares the tumor's intensity or the presence of a discriminative feature inside the tumorous tissue. The lesion classification is described in Section 6.2

The conclusion of the work is stated in Chapter 7. There, the results of the individual subtasks are described and discussed. Moreover, the future work that could yield better results is described here as well.

Chapter 2

Liver Cancer

2.1 Introduction

A tumorous liver disease is a very serious condition, whose elimination requires the cooperation of a number of different disciplines. Naturally, the largest proportion is represented by medical fields, e.g. radiology, surgery, gastroenterology, endoscopy, hepatology and more. However, non-medical fields applies here its potential as well; they can contribute to more accurate and faster diagnostics, participate in research for new treatments e.t.c.

Image processing is one representative of this group. Its field of application is the analysis of captured image data.

2.2 Liver

The liver is the second largest organ (the first one is the skin) that is located in the abdomen. It has a complex role in the human body including detoxification (blood filtration), metabolism, digestion and protein synthesis, just to name a few. During the intrauterine life the liver participates in the formation of blood. The liver processes are relatively intense, therefore approximately 12% of oxygen from the blood is consumed here and blood flowing from the liver reaches temperatures of up to 40 °C ([Válek et al., 2006]). To be able to run these processes, there is a rich vascular system through which flows up to 2 l of blood every minute ([Čihák, 2011]). Liver weight is in the range from 1 to 2.5 kg, corresponding to 2.5 % of body weight (on average).

Liver parenchyma is a relatively soft tissue and therefore its shape is partly defined by the enclosing organs. These organs are making so called imprints on the parenchyma. On the left lobe, there comes the most significant imprint from the stomach. The right lobe is affected by the right kidney and the right adrenal. From the image analysis point of view, an indirect contact between the liver and the heart is also very important. This could cause problems, because in the venous phase the density of the heart (35-75 HU) is very close to the density of the liver (40-60 HU). These organs are very often close together, thus making their separation difficult.

2.3   Vascular system

The liver contains a rich vascular system. Therefore, one can imagine the liver parenchyma as a blood-soaked sponge. The vascular system of the liver is composed of two components: nutritive and functional. This dual blood supply is formed by the the portal vein and hepatic arteries.

The **portal vein** (venea portae) ensures the functional role of the liver. It supplies the parenchyma with blood from unpaired organs of the abdominal cavity (stomach, intestines, pancreas etc.). The portal vein brings about 75% of the the overall blood volume that is supplied to the liver. This blood contains important nutrients obtained from the digestion process. These nutrients are further filtered by the liver parenchyma. One interesting thing about the portal vein is that it does not lead into the heart. Therefore, it is not a true vein in the classic conception. It enters the liver as one branch and soon divides into the right (ramus dexter) and the left branch (ramus sinister). The right branch then splits further into the front (superior) and the back (inferior) branch. The left branch consists of two parts - pars transversa and pars umbilicalis. The division to the left and the right branches of the portal vein often occurs prior to entering the liver parenchyma. Sometimes the right branch may be missing. In this case, two or more branches emerge from the portal vein and supply the right lobe.

The nutritive function, i.e. supplying oxygenated blood to the liver, is ensured mainly by the **hepatic artery** (arteria hepatica communis). This artery has three main branches: one branch for nourishing the gallbladder and the right and the left branches for nourishing the liver lobes. The blood in the portal vein contains a great amount of oxygen, the main task of the hepatic arteries is the nourishing of the biliary tract. Therefore, the hepatic arteries are vessels with a relatively small diameter.

The drainage of the liver blood is taken care of by the liver veins that have three main branches: the right (vena hepatica dextra), the left (v. h. sinistra) and the middle branch (v. h. media). The left and the middle branches merge together before leaving the parenchyma thus making a common short segment. All branches of the hepatic veins lead to the inferior vena cava. As in the case of the portal vein morphological variations can occur in the hepatic arteries as well. Most of them occur in different branching of the arteries. For example, a branch of the middle vein can lead directly to the inferior vena cava.

Besides the blood vessel system, there is also the biliary tract. This system, together with the hepatic artery, follows the portal vein system. On the other hand, the liver veins are not followed by any biological systems [Čihák, 2011].

2.4   Segments of liver parenchyma

The liver parenchyma can be divided into a certain number of different segments. This division is based on a functional anatomy. There are several approaches for the division; among them, the Couinaud's segmentation is the most used approach in Europe. Each segment in the Couinaud's segmentation is defined as a part of the liver parenchyma into

which enters one branch of the portal vein, one branch of the liver artery and one branch of the biliary tract. The output of the segment is ensured by one branch of the liver vein.

The middle liver vein splits the parenchyma into the right and left lobes. These lobes are further split vertically by the right and the left liver vein to individual sectors. The final horizontal split of the sectors is made by a horizontal plane located on the level of the first split of the portal vein.

The **left lobe** is significantly smaller and is divided by the left liver vein into the lateral and medial sectors. The lateral sector is further divided into the segments II and III. The medial sector forms the segment IV, which is further divided into the upper (IVa) and lower (IVb) segments.

The **right lobe** is divided by the right liver vein into paramedian and posterolateral sectors. These are further divided by the transversal plane into upper and lower segments. The right lobe is therefore divided into four segments: V, VI, VII, VIII, that are indexed clockwise.

The remaining segment I is a special segment, a so called lobus caudatus, which has its own vascularization.

Couinaud's segmentation is shown in Fig. 2.1. The identification of these segments in the CT of the liver is much more complicated and involves a great deal of expert knowledge. The radiologist is looking mainly at the morphology of the portal vein and based on its splitting defines the segment's borders. The visualization of individual segments in a CT is shown in Fig. 2.2.
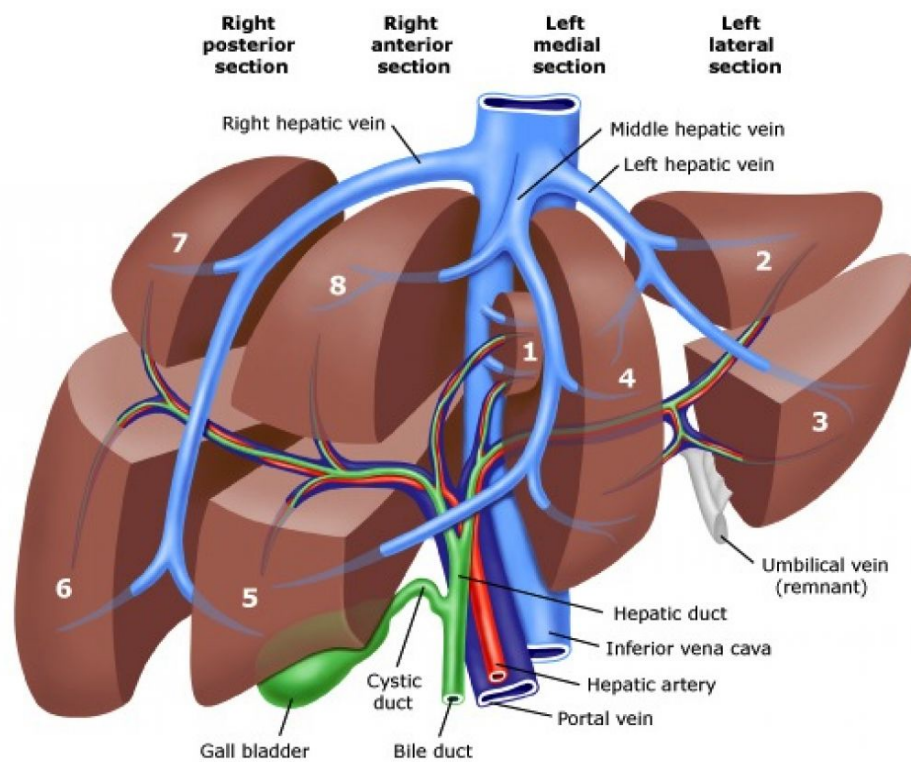
Figure 2.1: Couinaud's liver segmentation, [Siriwardena et al., 2014].

(a) I   (b) II   (c) III

(d) IVa   (e) IVb   (f) 5

(g) 6   (h) 7   (i) 8
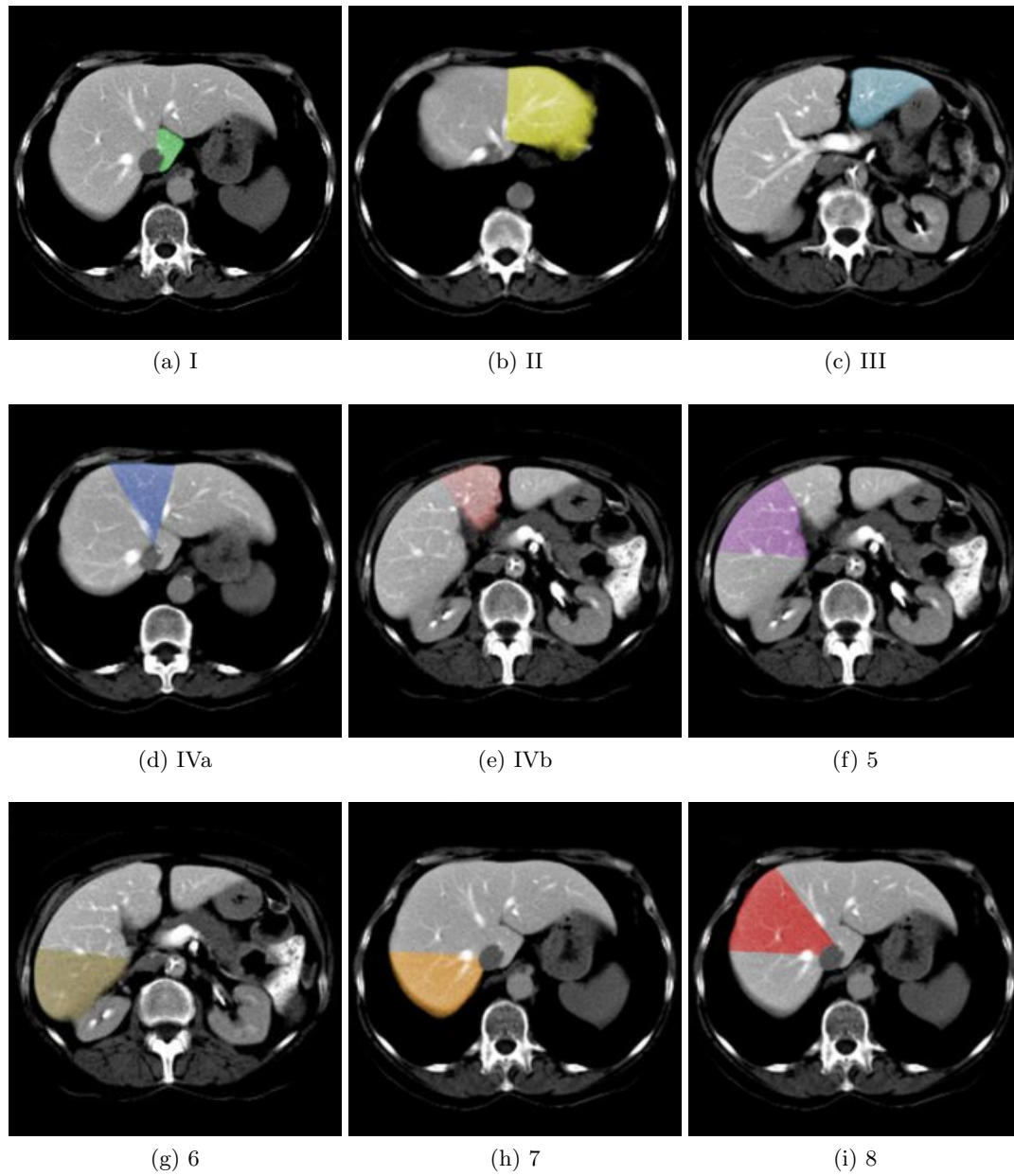
Figure 2.2: Visualization of individual liver segments, [Bohatá et al., 2009].

2.5 Diagnosis

In most cases, cancer has no symptoms at its early stage. As the cancer grows, the following symptoms may arise:

- pain in the upper abdomen,

- unexplained fever,

- unexplained weight loss,

- nausea,

- jaundice (yellow coloring of the eyes and skin),

- ...

There are several ways of diagnosing the presence of the cancer. These methods could be either invasive (e.g. biopsy) or non-invasive (e.g. imaging techniques). In the following text, some of the most used approaches are briefly mentioned.

2.5.1 Physical exam

The first thing to do is to undergo a physical examination. An affected liver can have a different shape and/or size. There can be also lumps in the abdomen. A blood sample will be sent for analysis to check the presence of certain proteins and antigens.

2.5.2 Ultrasound

The first test the patient must undergo if cancer is suspected is the ultrasound. This examination can reveal tumorous mass. To be able to localize and describe bearings different methods has to be used, e.g. a tomographic one. An example of US image is shown in Fig. 2.3a.

2.5.3 Computed Tomography (CT)

The CT provides better spatial precision as well as visualization of other structures such as vessels. This allows the doctor to examine the bearing's location w.r.t. the vascular system. Screening the whole abdomen also provides the possibility to examine the spread of the cancer (in the form of metastasis) into other organs. This and the relatively low cost of an examination makes it the most frequently used diagnostic tool. An example of US image is shown in Fig. 2.3b.

The use of the native CT has its limitation, and for the purpose of tumor imaging the differential diagnosis is more suitable. Here, a contrast agent is injected into a patient. This agent enhances the contrast of the human parts to which it travels in the blood. This way, it is possible to visualize concrete vascular system. Three phases are used in the differential diagnosis, see [Terrier et al., 2000]:

1. the arterial phase ... 20-28 sec after the initiation of contrast infusion,

2. the portal/venous phase ... 60-70 sec after the initiation of c.i.,

3. the delayed phase ... 10 min after the initiation of c.i.

2.5.3.1   Magnetic Resonance Imaging (MRI)

The cost of the MRI is much higher than the cost of the CT, but it could provide additional information regarding tumors. It is more suited to cases where the bile duct needs to be visualized. The main drawbacks of this method is the higher cost of acquiring the machine as well as the higher cost of individual examinations. The use of the MRI is also problematic in patients with joint replacements, etc. An example of the MRI image is shown in Fig. 2.3c.

(a)   (b)   (c)
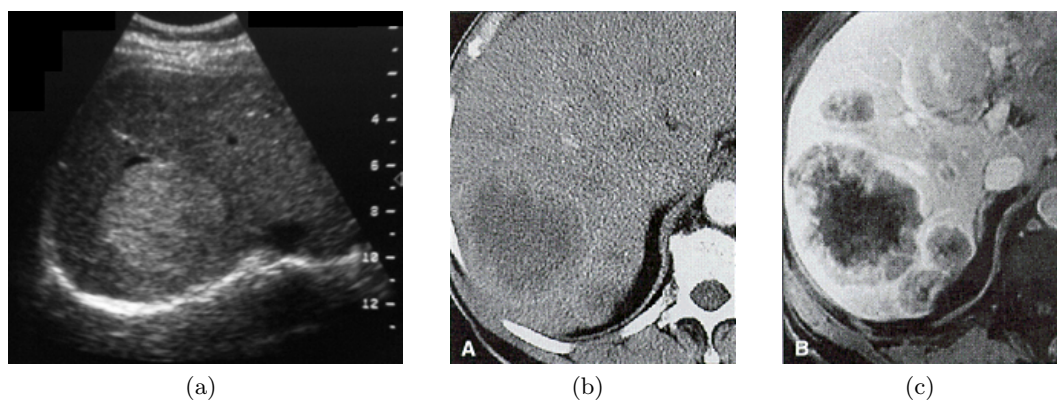
Figure 2.3: Comparison of CT (a), MRI (a) and US (c) images of the liver parenchyma.

2.6   Treatment

The next step after the liver cancer was diagnosed is to determine its extent. This process is called staging. The goal of staging is to find out whether the cancer has spread to other parts of the body. Therefore, we differentiate between primary and secondary liver tumors. The former are located in the liver parenchyma and are not metastasis from other tumors. The latter are metastasis of liver tumors that are located in different parts of the body.

There are several options for cancer treatment. Proper treatment depends on several factors such as the type, number, size, and location of tumors, health condition of the liver and/or the patient, the stage of the cancer etc. Liver cancer can then be classified in one of the three main groups:

- localized resectable,

- localized unresectable,

- advanced.

The first group describes one or more tumors located only in the liver parenchyma that can be surgically removed. The second group also describes tumors located only in the liver but which are not operable (due to the condition of the liver, the location of the tumor etc.). The last group describes metastatic cancer that has spread to both lobes of the liver or to other parts of the body.

The most used options for cancer treatment are briefly described in the following text.

2.6.1 Surgery

The liver has an outstanding ability to regenerate and grow back if a part of it has been removed. This allows the surgeons to safely remove up to 70% of the liver parenchyma if the parenchyma is otherwise healthy [Válek et al., 2006]. The process of removing a part of the liver is called liver resection. In most cases, it is not possible to cut off just the tumor. This is because of the complex vessel system and the intense blood flow through the parenchyma. Instead, the surgeon cuts off vessel sub-tree after the first or second branching of the portal vein. Then all liver segments that are supplied with blood by this branch are removed.

If the cancer has not spread to other parts of the body and a suitable donor can be found then liver transplant is another option.

2.6.2 Ablation

Ablation is often used when a patient cannot have surgery or waits for liver transplant. Radiofrequency ablation is the most used method. A special probe with tiny electrodes is guided directly to the tumorous tissue. The tumor is then destroyed by heat generated by electrodes.

2.6.3 Radiation Therapy

In this method, the tumorous tissue is killed by high-energy rays. The radiation comes either from a large machine aiming beams to the right position or from tiny radioactive spheres that are injected into the hepatic artery.

2.6.4 Embolization

Embolization is a process of blocking blood flow that supplies a part of the liver with the tumor. This is done by a catheter that is inserted into an artery in patient's leg and guided to the desired position - the hepatic artery. The doctor then injects small particles which stop the blood flow. This way, the tumor is not nourished any more and dies. The healthy liver parenchyma is nourished by the portal vein.

Another method of embolization is chemoembolization. The doctor injects an anticancer drug (chemotherapy, see below) just before blocking the blood flow. This way, the drug stays in the liver longer and therefore is more effective.

2.6.5   Targeted Therapy

Targeted therapy is a drug that slows the growth of the tumors and reduce their blood supply. This is again an option for patients with non-operable cancer.

2.6.6   Chemotherapy

Chemotherapy is a process well-known to the general public. It is a process of using drugs that kill the cancer cells. These drugs are given intravenously and then travel throughout the patient's body. In most cases, not only the cancer cells are killed by the drug. Therefore this treatment has high number of side effects. If the chemotherapy had not killed the tumor it could have reduced it enough so the patient can have a surgery. This approach for reducing the size of the tumour is often used intentionally.

2.7   Liver Tumors

The liver is relatively susceptible to cancer. The reason for this is the liver size and especially the amount of the blood flow and the metabolic processes. Given the amount of blood that flows through the liver, the majority (up to 90%) of hepatic tumors constitute metastases from other organs, [Válek et al., 2006]. The most common sources of these metastases are pancreas (50%), colon (25%) and stomach (20%), [Třeška et al., 2010].

The tumors can be classified with respect to several criterion:

- benign, malignant,

- primary, secondary (i.e. metastasis),

- hypervascular, hypovascular,

- epithelial, mesenchymal,

- frequent, rare,

- . . .

The following text describes the most common liver tumors with their radiological findings in differential CT.

2.7.1   Hepatocellular Carcinoma (HCC)

The HCC is the most frequent, primary, malignant liver tumor - it represents up to 80% of all primary liver tumors occurrences as stated in [Collier and Sherman, 1998]. Its yearly mortality is around 1.2 million patients, of which 83% are men, [Klener, 2011]. Spread of the HCC varies by geographic location. Hyperendemic[1] areas are the east and the southeast Asia (China, Korea, etc.) and the sub-saharan Africa. Yearly incidence[2] is reported 120 new

---

[1] High and continued incidence
[2] The number of new cases.

cases in 100 000 persons, [World Cancer Research Fund International - American Institute for Cancer Research, 2015]. Areas with intermedia incidence are for example the Japan, the south Europe and the Middle East. Low incidence is reported in the USA, the north Europe and the south America (2-5 / 100 000).

Significantly higher incidence is in patients with a diffuse liver disease (described in Section 2.8). In these cases the tumor is located relatively soon as an incidental finding in screening due to the diffuse disease. In patients without diffuse diseases, the tumors are located much later based on clinic signs (pain and/or pressure in the abdomen, weight loss e.t.c.). The worst case is that the tumor is localized based on its metastasis.

The tumor is often hypervascularized. Therefore, it is better visualized using a contrast agent. Basically, three main types of HCC can be derived:

- encapsulated unilocular foci,

- nodular type - the parenchym contains multiple small deposits,

- diffuse type.

In the native[3] CT the HCC is visualized as a homogeneous or a heterogeneous hypodense tumor that is well defined. The key is a two-phase CT with a contrast agent. In the arterial phase, the tumor saturates homogeneously and therefore is shown as a hyperdense blob. A so called wash-out is a process that is characteristic for the HCC. During the wash-out in the portal phase, the contrast agent is leaving the tumorous tissue faster then the surrounding parenchyma. Hence, the foci is shown in the portal phase as a hypodense blob. However, at the beginning of the portal phase the foci can be isodense[4] or even slightly hyperdense. Therefore, the delayed portal phase is recommended. Unfortunately, in most cases only the arterial and the portal phase are taken.

There can also be a capsule visible in the images that natively appears as hypodense tissue, see Fig. 2.6a. In the arterial phase the capsule is shown still as hypodense (Fig. 2.6b). The capsule is best visible in the delayed phase, where it is visualized as hyperdense, see Fig. 2.6c. The capsule is reported approximately in 58% cases , [Válek et al., 2006].

The [American Association for Cancer Research,] stated that the HCC can be diagnosed in the cases when the tumor is grater then 2cm, it is hyperdense in the arterial phase and hypodense in the portal phase.

2.7.2   Fibrolamellar Carcinoma (FLC)

The FLC is a rare form of the HCC with incidence 1 in 5,000,000 persons, [Válek et al., 2006]. It has a low age of onset - approximately 20-30 years with median of 27 years as stated in [Stipa et al., 2006]. Neither diffuse disease nor the markers of the HCC are presented in patients with diagnosed FLC. Therefore, the diagosis alone frequently comes at the advanced stage.

The native CT visualizes the FLC as a slightly non-homogeneous hypodense mass, which is relatively extensive due to the late diagnosis. The foci is well saturated in the arterial

---

[3]Without a contrast agent.
[4]Having the same density as surrounding tissue.

Figure 2.4: An advanced HCC in the native CT (a), the arterial (b), the portal (c) and the delayed phase (d), [Válek et al., 2006].

as well as in the portal phase, the saturation is also non-homogeneous. The delayed phase often visualizes a hypodense scar, the rest of the foci is isodense. In some cases, a so called pseudo-capsule can be visualized. This capsule forms the boundary between the expanding tumorous mass and the healthy tissue. The tumor contains calcification in up to 60% of diagnosed cases, [Válek et al., 2006]. These calcifications are most often located in the central scar. The typical FLC is shown in Fig. 2.7.
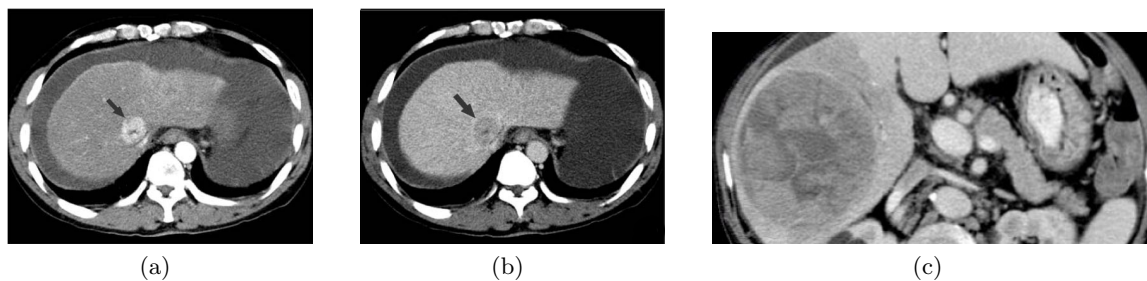
Figure 2.5: A HCC in an cirrhotic patient in the arterial (a) and the portal phase (b). Image (c) shows an advanced HCC with typical mosaic structure, [Radiology Assistant,].
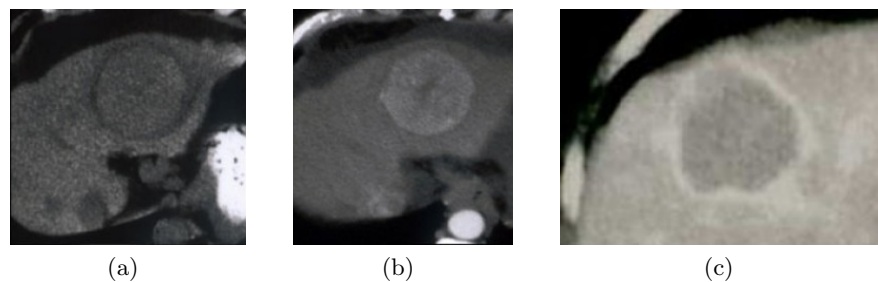
Figure 2.6: A HCC in the native CT (a), the arterial (b) and the delayed phase (c), [Radiology Assistant,].
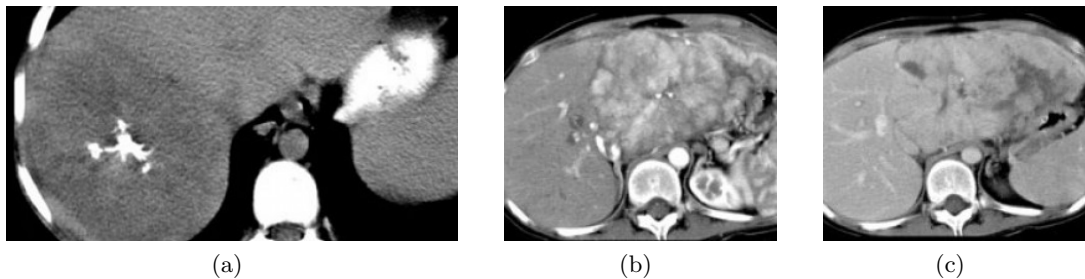
Figure 2.7: The native CT image of a FLC (a), the arterial (b) and the delayed phase (c), [Radiology Assistant,].

2.7.3 Cholangiocellular Carcinoma (ICC)

The ICC is another primary liver tumor that originates in the bile duct. The tumor can be further classified into three categories, from which only the intrahepatic case (therefore the abbreviation ICC) is related to this thesis. Although it is close to the rare type, it is the second most common primary malignant liver tumor that is more often diagnosed in older patients (60 years).

In the native CT the tumor is visualized as hypodense with possible calcifications. There is a significant saturation in the arterial phase that forms a hyperdense rim at its border. The portal phase is characterized by the non-homogeneous central saturation of the foci. The whole foci gradually saturates in the delayed phase, thus making itself hyperdense. However, there are cases in which this central saturation did not occur. A characteristic is the periferal washout in the delayed phase, see [Lee et al., 2001]. Individual phases of an ICC are shown in Fig. 2.8.

2.7.4 Angiosarcoma (HAS)

The HAS (the H is for hepatic) is a malignant tumor that, in addition to liver, can be localised in other parts of the body. According to [Válek et al., 2006], HAS is in up to 80% cases diagnosed in men, most often at 60-70 years of age, and the average survival is about 6
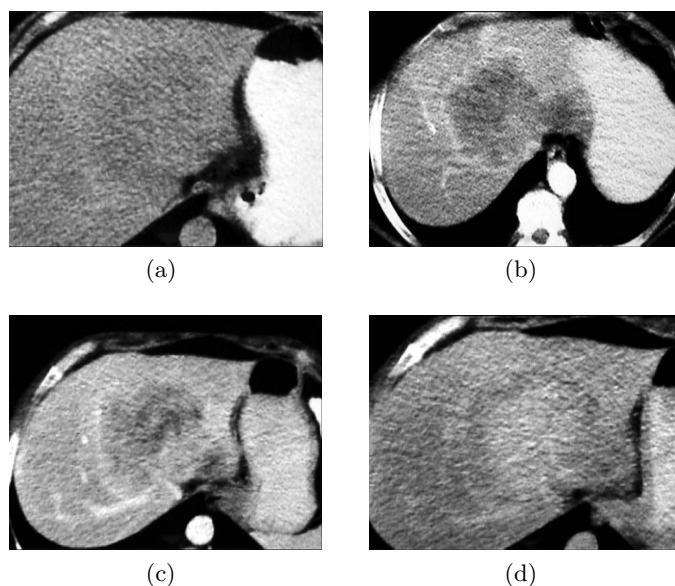
Figure 2.8: An ICC visualized in the native CT (a), the arterial (b), the portal (c) and the delayed phase (d), [Válek et al., 2006].

months. Based on the way of growth, the HAS may occur in the form of extensive tumorous tissue or as a combination of a dominant foci with multiple small nodes.

The native CT shows HAS as hypodense tumors. Hyperdense areas can be visualized in the tumorous tissue in some cases. These areas arise due to possible bleeding or calcifications. Saturation by a contrast agent can be varied. Both the arterial and portal phase visualize the tumor as hypodense in most cases. It could be also slightly hyperdense, but the density is lower than the density of the abdominal part of the aorta. The worst thing for automated diagnosis is that each foci can saturate in different manner. Figure 2.9 illustrates multiple tumors.
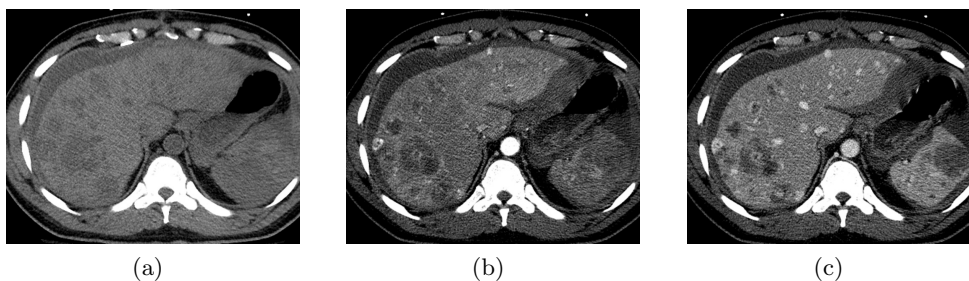
Figure 2.9: Multiple hypodense HAS are shown in the native CT (a), the arterial (b) and the portal phase (c), [The Liver Imaging Atlas,].

2.7.5   Secondary tumors (meta)

As stated in [Válek et al., 2006], the vast majority of liver tumors are metastases from tumors in different organs. This is because of the high amount of blood that flows through the liver parenchyma. Besides the vascular system, the metastasis can invade the liver parenchyma through lymphatic vessels or by direct ingrowth. Only 10% of metastasis are solitary[5] and occupy both liver lobes in 77% , [Radiology Assistant,]. Most liver metastases arises from the colorectal carcinoma.

In the native CT the metastasis are shown as hypodense tumors that can be non-homogeneous and poorly differentiable. The arterial phase visualizes a characteristic hyperdense rim that is washed-out in the delayed phase, see Fig. 2.10. The density of the rim is higher than the density of cysts and abscesses. Moreover, most metastasis can contain a necrotic center. To make the situation a little more complicated, smaller metastasis can saturate whole. In this case they remain saturated even in the delayed phase, see Fig. 2.11.

| (a) | (b) | (c) |
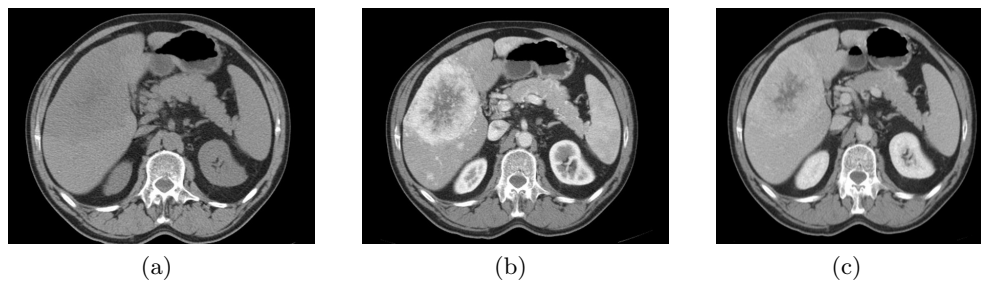
Figure 2.10: A big metastasis shown in the native CT (a), the portal (b) and the delayed phase (c), [Válek et al., 2006].

| (a) | (b) |
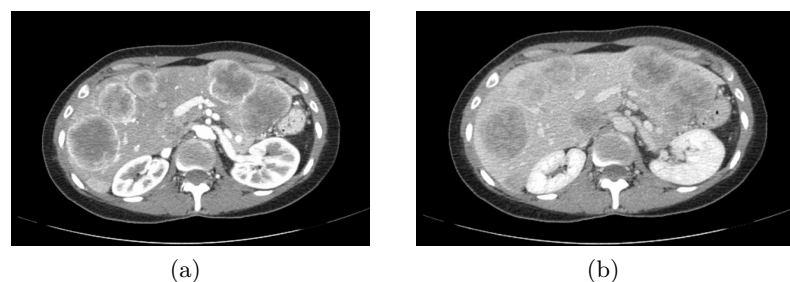
Figure 2.11: Visualization of multiple metastases in the portal (a) and the delayed phase (b), [Válek et al., 2006].

---

[5]Having only one foci.

2.7.6   Hemangioma (HG)

The Hemangioma is the most frequent benign liver tumor, which is diagnosed most often in adulthood, more in women than in men. The long-term screening shows no change in size in most cases. However, tumors that grew or changed their structure are documented as well.

Natively, the hemangioma is visualized as hypodense tumor. The saturation takes place in a nodular way from the periphery. The intensity of the saturation corresponds in all phases with the saturation of the corresponding vessel. As the density decreases in the delayed phase, the tumor becomes isodense. A typical hemangioma is shown in Fig. 2.12.

(a)   (b)

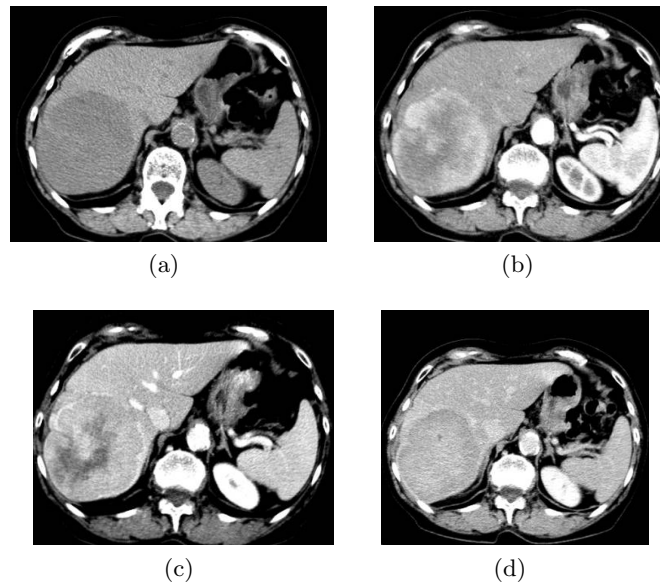(c)   (d)
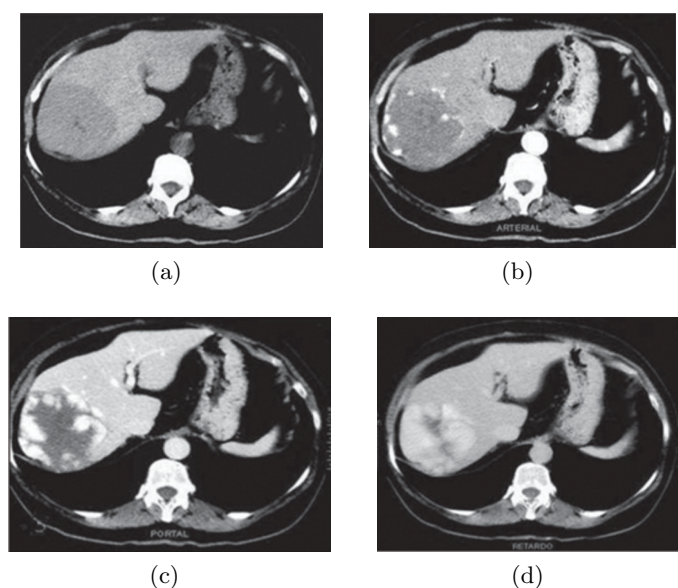
Figure 2.12: A hemangioma visualized in the native CT (a), the arterial (b), the portal (c) and the delayed phase (d), , [Válek et al., 2006].

2.7.7   Focal Nodular Hyperplasia (FNH)

The second most common benign liver tumor is the FNH and occurs more often in women (80-95% as stated in [Válek et al., 2006]. It is accompanied by the occurrence of hemangiomas, vessel malformations can occur as well. The size of the tumor is approximately 5cm and in 20% of cases the FNH consists of multiple bearings.

A typical sign of the FNH is a hypodense scar in the center of the loci. The native CT visualizes the tumor as a hypodense tissue with a strongly hypodense center (the aforementioned scar). In the arterial phase the tumor is shown as hyperdense, which persists until the portal phase. In the delayed phase the tumor is hardly localizable due to its isodense appearence. The only sign of the FNH in this phase may be the scar, which can be visualized as hypodense, isodense or also hyperdense. The tumor is of homogeneous character in all of the phases. Figure 2.13 shows the visualization of a FNH.

(a)   (b)

(c)   (d)
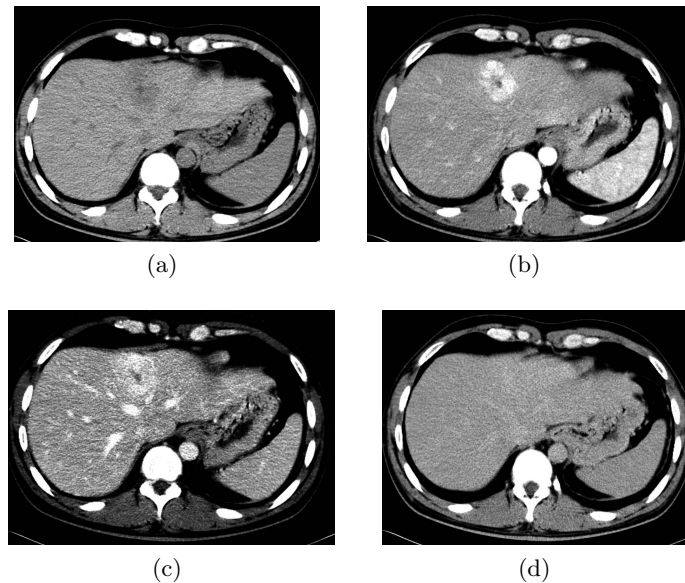
Figure 2.13: A FNH visualized in the native CT (a), in the arterial (b), the portal (c) and the delayed phase (d), [Válek et al., 2006].

2.7.8   Hepatocellular Adenoma (HA)

The HA is a relatively rare benign liver tumor, which, however, can turn into a malignant HCC. The most frequent risk factors are orally administered contraception and usage of anabolic steroids, just to name a few. Upon the discontinuation in intake of such a drug, the tumor may disappear. The size of the tumor is 5-10 cm and occurs most often as a solitary tumor.

In the native CT, the HA appears as a well differentiable hypodense lesion that can occur as homogeneous or non-homogeneous. Bleedings are often visible with density corresponding to the blood density (60-80 HU). The tumor quickly saturates in the arterial phase and is therefore visualized as a hyperdense foci. This holds through the portal phase and in the delayed phase the tumor is shown as isodense or hypodense as shown in Fig. 2.14.

2.7.9   Cysts

Cysts are relatively common objects that occur approximately in 4.5% of the population. They can have solitary or multiple form. Also, the size can vary a great deal - from milimeters up to tens of centimeters. Bigger cysts can cause pressure difficulties. Findings of the cysts are often incidental.

A cyst is essentially a cavity filled with fluid. For this reason, they are visualized in all phases as hypodense and well-differentiable tumor with the density 0-20 HU. In some cases, a cyst can contain blood. Even in such cases the tumor did not saturate after administration
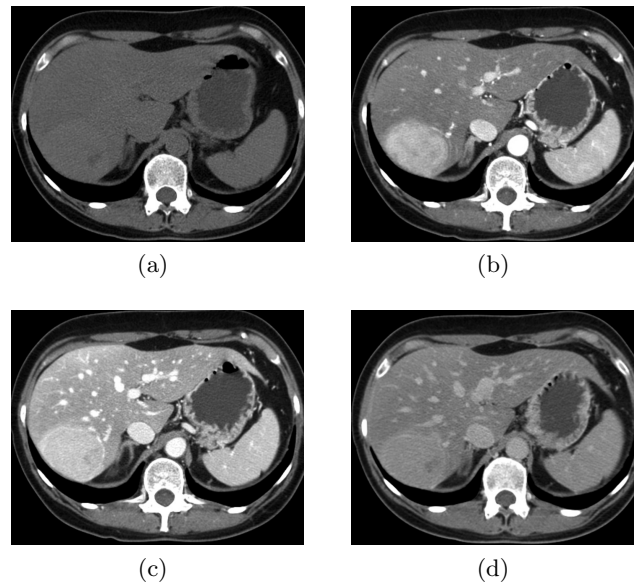
Figure 2.14: A HA visualized in the native CT (a), the arterial (b), the portal (c) and the delayed phase (d). The liver of the patient is fatty (hypodense), therefore the density of the tumor seems to be slightly higher, [The Liver Imaging Atlas,].

of a contrast agent. The reason for this is that the blood is not connected to the vascular system. A typical cyst is visualized in Fig. 2.15.
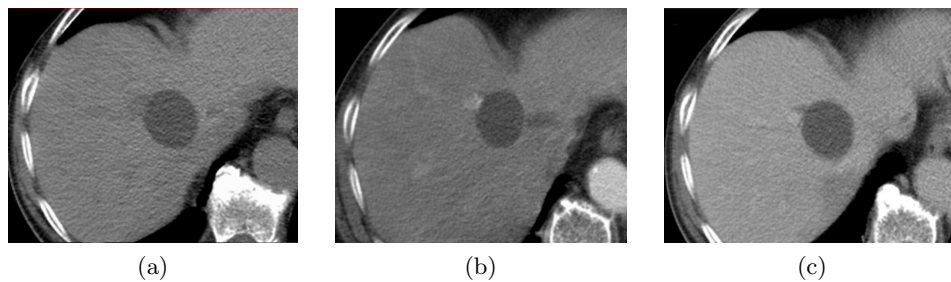
Figure 2.15: A cyst visualized in the native CT (a) the arterial (b) and the venous phase (c), [Ressurreição, 1970].

2.8   Diffuse procceses in liver

Processes which take place in a diffuse manner affect the whole liver parenchyma or its major part. Generally, these processes make the localization and classification of liver tumors difficult, because they are changing the native density of the parenchyma. In some tumors, this leads to a reduced contrast between the tumorous and the healthy tissue. This situation
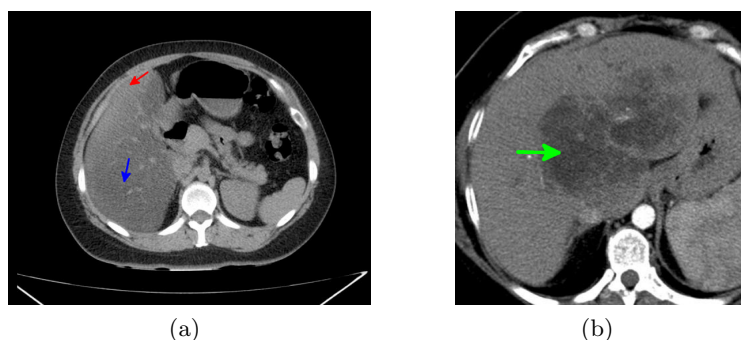
Figure 2.16: A diffuse steatosis (a) is manifested by hypodense parenchyma (blue arrow). Healthy tissue then seems to be hyperdense (red arrow). In the case of focal steatosis (b), hypodense areas (green arrow) can be found, [Radiopaedia,].

is much worse if the processes did not affect the whole parenchyma. If they take place locally, they create an impression of pseudo-tumors. This section describes the most frequent diffuse processes in greater detail.

2.8.1   Steatosis

Steatosis is the most frequent metabolic disorder of the liver, during which changes in the normal biological processes occur. The risk factors of this disorder are alcohol abuse, obesity, toxic damage, diabetes or even the pregnancy. The steatosis causes abnormal retention of lipids within a cell. Therefore, the doctors sometimes refer to it as fatty liver.

In the CT, it leads to slightly hypodense parenchyma (under 30 HU); negative values are reported as well, see [Válek et al., 2006]. The only segment that is spared this process is the segment I, i.e. lobus caudatus. Therefore, it may seem to be hyperdense with respect to the rest of the parenchyma. This can also be used as a feature for diagnosing steatosis. It can be of a diffuse, mosaic or even tumorous character.

Due to the hypodense parenchyma it is difficult to localize tumors that are natively visualized as hypodense (e.g. the HCC) as can be seen in Fig. 2.16a. Another complication involves cases of tumorous (i.e. focal) character, where individual tumor can be confused with a real tumor. This is shown in Fig. 2.16b.

2.8.2   Hemochromatosis

Hemochromatosis is caused by the abnormal accumulation of iron; therefore, it is also called iron overload. It can be innate or acquired by repeated transfusions, for example. Besides liver, iron is accumulated in other organs as well, especially in the heart and the pancreas. Therefore, these organs are at risk of failure as well.

Unlike the steatosis, the hemochromatosis presents itself by hyperdense parenchyma (75-130 HU), as shown in Fig. 2.17. It is often a consequence of cirrhosis, which is described in

greater detail in the following section. The increase of density is of a diffuse character; the focal character was not reported.

Both the portal and hepatic vein are shown as hypodense objects; it is also harder to localize hyperdense tumors. The hemochromatosis also increases the risk of HCC.

(a)   (b)
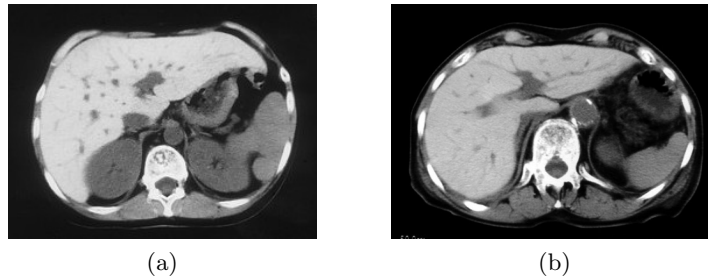
Figure 2.17: CT images of liver affected by the hemochromatosis, [Radiopaedia,].

2.8.3   Cirrhosis

Cirrhosis is a terminal irreversible damage of the liver parenchyma that is combined by an extensive fibrosis[6]. The most common causes of the cirrhosis are alcohol, hepatitis C, autoimmune diseases, metabolic defects etc.

From the radiological point of view, cirrhosis reflects in the CT images through deformations of the parenchyma - changes in size as well as in shape. The density of the parenchyma is also changed. The liver becomes mosaic, where hypodense and hyperdense areas alternate. The images can show fibrotic septa, which form hypodense stripes.

All of the above described changes need to be considered when analysing a cirrhotic data. Especially the presented hypodense areas make the localization of hypodense tumors difficult. Generally, one can expect higher false positive areas being detected. The cirrhosis is also often accompanied by hemochromatosis, making the analysis even more problematic. Cirrhotic data are shown in Fig. 2.18.

---

[6]Thickening/compression of connective tissue.
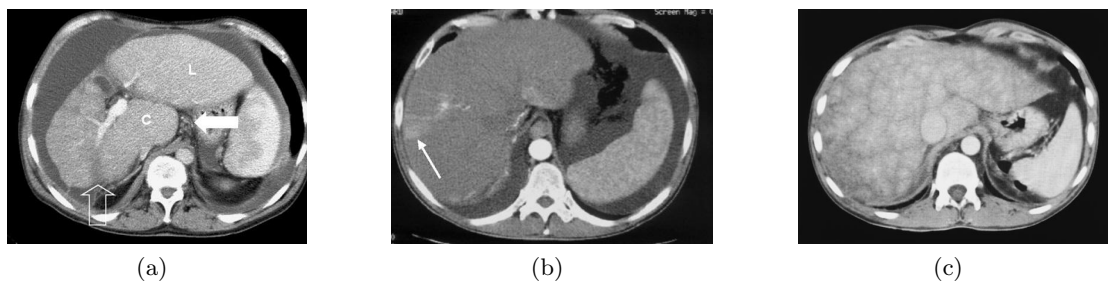
Figure 2.18: Image (a) shows advanced cirrhosis with enlarged left lobe (L) and lobulus caudatus (C). The empty arrow points to fibrotic tissue (not a tumor). The arrow in image (b) points to an arterial shunt that looks like a HCC. A typical cirrhotic mosaic structure is shown in image (c). Images taken from [Taylor, 2011].

Chapter 3

Image Segmentation

The topic of this thesis involves several tasks that need to be solved in order to achieve the desired goals. One such problem is finding liver parenchyma and being able to extract tumorous regions from it. Therefore, a method for image segmentation needs to be applied. Of course, there are numerous methods suitable for solving this task. This chapter describes methods that are used for solving the thesis topic.

3.1 The Labeling Problem

A significant part of the tasks of machine vision can be defined as a labeling problem, i.e. assignment of a label to each image element. To specify a labeling problem, one has to define a set of sites and a set of labels first as in [Li, 2009]. Let $\boldsymbol{\mathcal{S}}$ index a discrete set of $\boldsymbol{m}$ sites

$$\boldsymbol{\mathcal{S}} = \{\boldsymbol{1}, \ldots, \boldsymbol{m}\}, \tag{3.1}$$

where $\boldsymbol{1}, \ldots, \boldsymbol{m}$ are indices. In vision problems, a site does not necessarily have to be an image pixel. A site can represent any kind of a spel[1] or an image feature such as a line, a corner point, an edge, a surface patch, or a volume. A set of sites may be categorized in terms of their regularity to spatially regular, e.g. a rectangular 2D image lattice, or spatially irregular, e.g. line segments and other similar image features extracted in more abstract level. In MRF models the sites are treated as unordered, i.e. a pixel $(\boldsymbol{i}, \boldsymbol{j})$ can be reindexed by a single number $\boldsymbol{k} \in \{\boldsymbol{1}, \ldots, \boldsymbol{m}\}$. The interrelationship between two sites is characterized by a neighborhood system.

Let $\boldsymbol{\mathcal{L}}$ be a set of labels. In contrast to the definition of the set of sites the set of labels does not have to be only a discrete set. For example, $\boldsymbol{L}$ can represent the dynamic range for an analog pixel intensity:

$$\boldsymbol{\mathcal{L}_c} = [\boldsymbol{x_l}, \boldsymbol{x_u}] \subset \mathbb{R} \tag{3.2}$$

where $\boldsymbol{x_l}$ and $\boldsymbol{x_u}$ represents the lower and the upper value of the dynamic range. In the case of image segmentation the set $\boldsymbol{L}$ is defined on a discrete space

$$\boldsymbol{\mathcal{L}_d} = \{\boldsymbol{1}, \ldots, \boldsymbol{M}\} \tag{3.3}$$

---

[1]Spel (spatial element) can represent one pixel or an arbitrary image region.

in which $\boldsymbol{M}$ denotes the number of labels that corresponds to the number of image segments.

In many cases (image restoration, etc.) the label set is ordered, which means that a distance measure can be defined to describe the qualitative difference between two labels, e.g. the Euclidean distance. In contrast, the segmentation problem often allows only the "equal" and "nonequal" relationship.

Once both the set of sites and the set of labels are defined one can proceed to assigning a label to each of the sites. This is called the labeling problem. The labeling of the sites in $\boldsymbol{S}$ in terms of the labels in $\boldsymbol{\mathcal{L}}$ is the set

$$f = \{f_1, \ldots, f_m\}. \tag{3.4}$$

One can look at the labeling as a mapping from $\boldsymbol{\mathcal{S}}$ to $\boldsymbol{\mathcal{L}}$:

$$f : \boldsymbol{\mathcal{S}} \to \boldsymbol{\mathcal{L}}. \tag{3.5}$$

where each site is assigned a unique label. In the MRF theory, labeling is often called a *configuration* of the field. The total number of possible configurations of a markov field over $\boldsymbol{m}$ sites with $\boldsymbol{M}$ possible labels is computed as $\boldsymbol{M^m}$:

$$\mathbb{F} = \underbrace{\boldsymbol{\mathcal{L}} \times \boldsymbol{\mathcal{L}} \ldots \times \boldsymbol{\mathcal{L}}}_{\boldsymbol{m} \text{ times}} = \boldsymbol{\mathcal{L}^m}. \tag{3.6}$$

This number is astronomic even for a moderate number of labels and small images, thus making it impossible to optimize by brute force.

In some cases, possible labels can differ from site to site, which gives the configuration space

$$\mathbb{F} = \boldsymbol{\mathcal{L}_1} \times \boldsymbol{\mathcal{L}_2} \ldots \times \boldsymbol{\mathcal{L}_m} = \boldsymbol{\mathcal{L}^m}. \tag{3.7}$$

3.1.1 Types of Labeling Problems

In accordance with the work [Li, 2009], labeling problems can be classified w.r.t. the regularity of the sites and the continuity of the labels into the following four categories:

- LP1: Regular sites with continuous labels.

- LP2: Regular sites with discrete labels.

- LP3: Irregular sites with discrete labels.

- LP4: Irregular sites with continuous labels.

The first category involves problems such as real (non binary) image restoration and smoothing. A typical representative of LP2 problems is image segmentation, where the set of labels is given by the number of image segments. Edge detection also belongs to this category. Object matching based on image features like corner points or line segments belongs to the category LP3 and a typical example of LP4 problem is pose estimation from a set of correspondences.

3.1.2   Involving the Contextual Constraints

The use of contextual information is ultimately necessary for proper image understanding as shown in [Pavlidis, 1986]. The first use of contextual information for solving an image analysis problem is published in [Chow, 1962].

From the probability point of view, the contextual constraints may be expressed either locally or globally. For local expression, conditional probabilities $P(f_i|\{f_{i'}\})$ take place. Here $f_{i'}$ denotes the set of labels at the other sites $i \neq i'$. To express contextual constraints globally, the joint probability $P(f)$ is used. Due to the more frequent cases where local information is more directly observed, a global inference is made based on those local properties.

In situations where no contextual constraints are made, the labels are independent of one another:

$$P(f_i|\{f_{i'}\}) = P(f_i),\ i \neq i'. \tag{3.8}$$

Therefore the joint probability is the simple product of local probabilities:

$$P(f) = \prod_{i \in S} P(f_i). \tag{3.9}$$

While this is advantageous for problem solving, in most cases the contextual constraints should be involved to achieve better image understanding. Equation 3.8 does not hold anymore and making a global inference based on local information becomes a nontrivial task.

3.2   The Role of Optimization Principles

The exact solution to a vision problem hardly exist in many cases. This is due to various uncertainties in the processes involved, such as noise and other degradation factors. Using the optimization principle leads to a solution that optimizes a predefined objective function either explicitly or implicitly.

From optimization point of view, three basic issues need to be figured out ([Li, 2009]): problem representation, formulation of objective function and deriving an optimizing algorithm. To solve the first issue the site and label sets have to be defined. Probably the most critical part is the formulation of an objective function. It is a function that maps the solution to a real number and thus reflects the quality of that solution. The formulation also describes how various image features and contextual constraints are involved in the optimization procedure. The last issue describes the procedure for optimizing the objective function. A perfect algorithm should be able to find a local or even a global extrema of the objective function in an efficient way. The efficiency here reflects not only the computational time but also the size of the searched solution space.

Assume without loss of generality that our goal is to minimize the objective function. Then the objective function is often called an energy function and should be formulated so that a solution to a given problem corresponds to its minimum. Another role of the energy function is to guide the search through the solution space. From this point of view, the

minimum should be located in a convex area. The problem that often arises is that a low-quality local minimum is found instead of the global one or a superior local one. To overcome the problem of finding a local minimum, a stochastic method that can temporarily accept a worse solution can be used. A typical representative of this kind of methods is simulated annealing. Naturally, methods able to find a global minimum are preferred but not always usable.

The formal approaches to optimizing an energy function provides a convenient way for the evaluation of different solutions. To improve the search efficiency by pruning of the solution space, different heuristics can be used. Combining both approaches leads to a good strategy for solving a general optimization task.

An energy function $\boldsymbol{E}$ is defined by its form and the parameters involved. The minimal solution is then expressed as follows:

$$\boldsymbol{f}^* = \arg\min_{\boldsymbol{f}} \boldsymbol{E}(\boldsymbol{f}|\boldsymbol{d}, \boldsymbol{\theta}) \qquad (3.10)$$

where $\boldsymbol{f}$ denotes a solution, $\boldsymbol{d}$ are the observed data and $\boldsymbol{\theta}$ corresponds to the set o parameters. Different $\boldsymbol{f}$ and/or $\boldsymbol{d}$ defines a different energy function, which leads to a different optimal solution $\boldsymbol{f}^*$. The parameters $\boldsymbol{\theta}$ are either known a priori or are estimated from the data.

As already stated above, an energy function in the MRF theory is formulated using established criteria. Probably the most popular one is the Bayes criteria and the maximal a priori (MAP) principle that leads to the MAP-MRF framework described in greater detail in the next section.

3.3   Markov Random Fields

As time goes by, optimization principles become more and more popular for solving image analysis problems. Finding an optimal solution to a vision task is in most real cases impossible due to various uncertainties. Hence, using optimizing principles is a natural way to solve these issues. The solution can be sought not only explicitly but implicitly as well. For the proper solving of a computer vision task, contextual constraints have to be used. The context has a critical influence for understanding visual information in images. From this point of view, the solution of a problem can be divided to two subproblems:

1. objective function definition involving contextual constraints

2. optimizing this objective to find an optimal solution.

The Markov random field forms a branch of probability theory and provides tools suitable for the characterization of contextual constraints. A certain MRF model favors its own class of patterns and assigns it a higher probability. This is a classical probability approach used in many classifiers. The added value of the MRF is the possibility of involving contextual constraints. Together with the decision and estimation theory, the MRF can be interpreted as a systematic way for developing algorithms and thus avoiding the use of ad hoc heuristics. One such an approach is the maximum a posteriori (MAP) concept that leads to the MAP-MRF framework, which is one of the most used approaches. The MAP-MRF framework was introduced in [Geman and Geman, 1984a].

The issue of involving the needed contextual constraints while defining a objective function is solvable by MRF theory, which is both convenient and consistent. For this purpose, the conditional MRF distributions are used, which describe mutual influences between each other. From these conditional distributions the joint distribution needs to be derived in most applications. But in the case of the MRF this task turns out to be very difficult. A breakthrough in the area of practical use of the MRF is represented by the Hammersley-Clifford theorem stating the equivalence between the MRF and the Gibbs distribution. It was established in [Hammersley and Clifford, 1971] and further extended in [Besag, 1974]. Using this theorem, the joint distribution of the MRF can be replaced by the Gibbs distribution, which is of a simple form.

3.3.1   Spatial Relation of Sites

To specify contextual constraints between sites, one has to define their spatial relation first. This section describes the connection between neighborhood systems used when working with images and cliques when working with graphs.

3.3.1.1   Neighborhood System

The spatial relation between sites from a set $\mathcal{S}$ can be specified by a neighborhood system:

$$\mathcal{N} = \{\mathcal{N}_p | \forall p \in \mathcal{S}\} \tag{3.11}$$

where $\mathcal{N}_p$ denotes a set of all sites that neighbors with given site p. One property of neighborhood system is that a site is not neighboring to itself, i.e. $p \notin \mathcal{N}_p$. Another important property is the mutuality, i.e. if a site p is neighbor of another site $q \neq p$, than q is also neighbor of the site p: $p \in \mathcal{N}_q \iff q \in \mathcal{N}_p$.

A typical definition of a neighborhood system is via a distance measure.

$$\mathcal{N}_p = \{q \in \mathcal{S} \mid [\mathrm{dist}(s_q, s_p)]^2 \leq r^2, \; q \neq p\} \tag{3.12}$$

where r denotes the radius and $dist$ is an arbitrary distance measure. Defining the system this way is suitable for both regular and irregular set of sites.

In most cases the *Euclidean distance* is used, especially for the irregular sites. The Euclidean distance between pixels (p, q) and (r, s) in a 2D image is as follows:

$$D_E\left((p,q),(r,s)\right) = \sqrt{(p-r)^2 + (q-s)^2}. \tag{3.13}$$

The main advantage of the Euclidean distance is its intuitive meaning. On the other hand, due to the square root its calculation is costly. In the case where sites are placed on a regular lattice, other distances can be used to overcome the problems with time efficiency. The *Manhattan distance* (or city block or D_4 distance) is described as a minimal number of elementary steps in the digital grid which are needed to move from one point to another, [Sonka et al., 2007]:

$$D_4\left((p,q),(r,s)\right) = |p-r| + |q-s|. \tag{3.14}$$
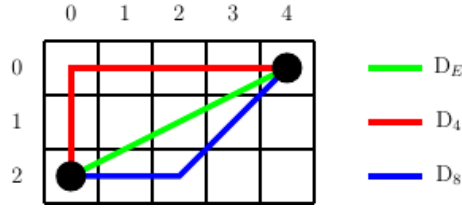
Figure 3.1: Different distance measures
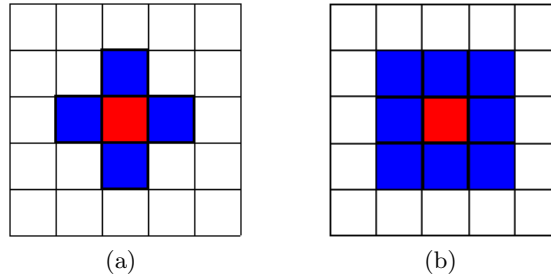
Figure 3.2: Most common types of neighborhood systems are the first (a) and the second (b) order systems.

Another representative that is often used in discrete cases is the *chessboard distance* (or D_8). In comparison with the Manhattan distance the diagonal moves are allowed int his case:

$$D_8\left((p, q), (r, s)\right) = \max\{|p - r|, \ |q - s|\}. \tag{3.15}$$

The most frequently used types of neighborhood system are the first (von Neumann) and the second (Moore) order system. Its 2D version are shown in Figure 3.2.

3.3.1.2 Cliques in a Graph

For further use, the term clique in a graph needs to be defined. A graph can be represented by a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$ between those nodes, i.e. $\mathcal{G} \triangleq (\mathcal{V}, \mathcal{E})$. Next, let the edges be derived based on a neighborhood system $\mathcal{N}$. Then, a subset $c \subset S$ is called a clique if every pair of sites in this subset consists of neighbors, i.e. $\forall s_p, s_q \in c, p \neq q : \ s_p \in \mathcal{N}_q$. In other words, for each pair of sites in a clique there exists an edge in $\mathcal{E}$ between corresponding nodes: $\forall s_p, s_q \in c, p \neq q : \ (s_p, s_q) \in \mathcal{E}$. The cliques can be categorized w.r.t. the number of sites/nodes of which they consists. The cliques can be of the order of one - $c_1 = \{p\}$, two - $c_2 = \{p, q\}$ and so on. Based on their order they are called singletons, doubletons etc. The set of all cliques of the first order is denoted $\mathcal{C}_1$, of the second order $\mathcal{C}_2$, and so on. The set of all cliques is then denoted $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \ldots \mathcal{C}_n$ where n is the maximal clique order.

For the first order neighborhood system shown in Fig. 3.2a, there is the total number of three cliques coresponding to the singleton and the horizontal and vertical doubletons shown in Fig. 3.3a and 3.3b, respectively. In the case of the second order neighborhood system ,
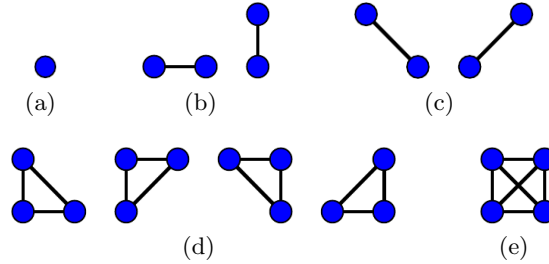
Figure 3.3: Cliques on a regular lattice: singleton (a), doubletons (b) and (c), tripletons (d) and a quadrupleton (e).

see Fig. 3.2b, there are also the diagonal doubletons (Fig. 3.3c), the tripletons (Fig. 3.3d) and the quadrupleton (Fig. 3.3e).

The total number of cliques grows rapidly with the increasing order of the neighborhood system. To overcome computational expenses only the first or second order neighborhood system is used in most cases. More information about neighborhood systems and cliques in a graph can be found, for example, in [Li, 2009] and [Blake et al., 2011].

3.3.2   The MAP-MRF Framework

To describe the motivation for using probability theory and the MAP principle specifically, let us define a set of image features $d = \{\overrightarrow{d}_s : s \in \mathcal{S}\}$ (a so-called observation) and a set of labels $f = \{f_s : s \in \mathcal{S}, f_s \in \mathcal{L}\}$ as earlier. For an image of size $N \times M$, there are the total number of $|\mathcal{L}|^{NM}$ possible labelings. That is an intractable set of possibilities even for a moderate size of images. To decide which of them is the right one, the probabilistic approach can be used. This leads to the definition of a probability measure over the set of all possible labelings and choosing the most probable one. In other words, one needs to define a measure $P(f|d)$ that corresponds to the probability of labeling f given observation d. Then, the goal is to find an optimal labeling $\hat{f}$ that maximizes this probability $P(f|d)$. This procedure is called the Maximum a Posterior (MAP) estimate, [Li, 2009]:

$$\hat{f}^{\mathrm{MAP}} = \arg \max_{f \in \mathbb{F}} P(f|d) \qquad (3.16)$$

where $\mathbb{F}$ denotes the set of all possible labelings.

Sometimes this is called an inverse problem, where one tries to recover a large number of unknown or hidden variables - f in our case. This procedure is based on another set of variables called observation - d in our case. The observation corresponds to image features derived in advance. Both of the sets can be of the same nature, e.g. two images in restoration problems, or of completely different nature, e.g. an input image and labeling in segmentation task. In some cases, more than one observation needs to be used for inference of the hidden variables, e.g. two input images in stereo vision.

For computing the posterior probability the Bayes rule is used:

$$P(f|d) = \frac{p(d|f)P(f)}{p(d)}.$$ (3.17)

Here $P(f)$ denotes the prior probability of labelings (often called simply prior) f and $p(d)$ is the density function of d. The last term $p(d|f)$ is the conditional probability density function (p.d.f.) of the observation d for given labeling f. This is called the likelihood function of f for fixed d also known as the evidence for the labeling f from the observation d and it describes how the labeling f fits given data d. For given data d the density $p(d)$ is constant, thus simplifying Equation 3.17 to

$$P(f|d) \propto p(d|f)P(f).$$ (3.18)

To solve this equation, one needs to define both of the terms $p(d|f)$ and $P(f)$ and to do that, the MRF theory is used.

3.3.3 Markov Random Fields

To be consistent with previous definitions the MRF will be defined in terms of graph theory. Having this in mind, the MRF can be defined as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ denotes the set of nodes. Each node is associated with a random variable F_p for $p = 1, 2, \ldots, N$, each random variable can take a value $f_p \in \mathcal{L}$. To define the set of edges $\mathcal{E}$ the neighborhood system $\mathcal{N}$ must be known. The set of nodes to which the node p is adjacent defines its neighborhood $\mathcal{N}_p$, i.e. $q \in \mathcal{N}_p \iff (p, q) \in \mathcal{E}$. The set $\mathcal{N}_p$ is called the Markov blanket of node p.

A realization of the MRF f, i.e. an event where each node takes a value $f = \{f_1, f_2, \ldots, f_m\}$, is called a configuration of that field. The probability that a random variable F_p takes the value f_p is denoted $P(F_p = f_p) = P(f_p)$ and the probability of a configuration f is denoted $P(F = f) = P(f)$. In the case of continuous set of possible labels $\mathcal{L}$ the probability density functions $p(F_p = f_p)$ and $p(F = f)$ are used instead.

If a random field (RF) is to be a Markov random field two following conditions must be satisfied:

1. positivity:

$$P(f) > 0, \ \forall f \in \mathbb{F}$$ (3.19)

2. Markovianity:

$$P(f_p|\{f_q\}_{j \in \mathcal{V} \setminus i}) = P(f_p|q \in \mathcal{N}_p) = P(f_p|f_{\mathcal{N}_p})$$ (3.20)

The positivity guarantees that the joint probability $P(f)$ of any random field is uniquely determined by its local conditional probabilities as stated in [Besag, 1974]. The Markovianity describes a *knock-on effect*, where explicit short-range linkages give rise to implied long-range correlations ([Blake et al., 2011]). This feature is the great attraction for using MRFs.

To specify a MRF, two approaches can be used. The first one is based on conditional probabilities $P(f_i|f_{\mathcal{N}_i})$ and the second one is based on joint probability $P(f)$. The former approach is preferred still there were some problems using it. Fortunately, the already mentioned Hammersley-Clifford theorem defines the equivalence between MRFs and Gibbs distributions that provides mathematical tools for specifying the joint probability of a MRF.

A MRF can have two important properties - homogenity and isotropy. A MRF is said to be *homogeneous* if the conditional probability $P(f_p|f_{\mathcal{N}_q})$ is independent of the relative location of the node/site p in $\mathcal{S}$. The homogenity is assumed in most MRF vision models making the computations more convenient. Moreover, to define an RF that is not homogeneous is much more difficult. A MRF is said to be *isotropic* if the clique potentials V_c are independent of the orientation of the clique c. This property is met if undirected graphs are used for describing a vision problem, e.g. image segmentation in most cases. For more information see [Li, 2009].

3.3.4   Gibbs Random Fields

A set of random variables F forms a *Gibbs random field* (GRF) on the set of sites $\mathcal{S}$ w.r.t the neighobrhood system $\mathcal{N}$ if and only if its configurations obey a Gibbs distribution of the following form:

$$P(f) = \frac{1}{Z} \times e^{-\frac{1}{T}E(f)} \tag{3.21}$$

where T is a constant called temperature that controls the sharpness of the distribution, $E(f)$ denotes the energy function of a configuration f and Z is a normalizing coefficient often called the partition function. It ensures that that the distribution sums to 1 and takes the following form:

$$Z = \sum_{f \in \mathbb{F}} e^{-\frac{1}{T}E(f)} \tag{3.22}$$

The advantage of the Gibbs distribution is that the energy function can be easily expressed as a sum of *clique potentials* $V_c(f)$:

$$E(f) = \sum_{c \in \mathcal{C}} V_c(f) \tag{3.23}$$

where $\mathcal{C}$ is the set of all possible cliques in a graph.

Equation 3.23 can be rewritten in a form that couples cliques based on their order:

$$E(f) = \sum_{\{p\} \in \mathcal{C}_1} V_1(f_p) + \sum_{\{p,q\} \in \mathcal{C}_2} V_2(f_p, f_q) + \ldots + \sum_{\{p,q\ldots\} \in \mathcal{C}_n} V_n(f_p, f_q \ldots) \tag{3.24}$$

where n denotes the maximal clique order. If the first order neighborhood system is used, only the cliques of order up to two are considered. This is an important special case very often used for solving vision problems. Equation 3.24 then simplifies to the following form:

$$E(f) = \sum_{\{p\} \in \mathcal{C}_1} V_1(f_p) + \sum_{\{p,q\} \in \mathcal{C}_2} V_2(f_p, f_q) = E_{\text{data}}(f) + E_{\text{smoothness}}(f) \tag{3.25}$$

The first term E_{data} of Equation 3.25 is often called simply the data term and evaluates the fit of the labeling to the observation. The second term $E_{\mathrm{smoothness}}$ of Equation 3.25 is often called simply the smoothness term and encourages homogeneous regions. In literature, the data term of site p is often denoted as $D_p(f_p)$ and the smoothness term between two sites p and q as $V_{p,q}(p, q)$. V_i with i been a numeric index denotes a clique potential of the clique of i-th order. Thus, Equation 3.25 can be rewritten as follows:

$$E(f) = E_{\mathrm{data}}(f) + E_{\mathrm{smoothness}}(f) = \sum_{p \in \mathcal{S}} D_p(f_p) + \sum_{\{p,q\} \in \mathcal{N},\, p,q \in \mathcal{S}} V_{p,q}(f_p, f_q) \quad (3.26)$$

This notation will be used further in this text.

One problem that must be solved to calculate a Gibbs distribution is evaluation of the partition function Z. As already stated, the set of all possible labelings $\mathbb{F}$ is intractable even for moderate number of random variables. Therefore one of approximation methods is often used to deal with this problem, see Section 3.3.6.

Equation 3.21 makes it clear that more probable configurations are those with smaller energy. In other words, the lower the energy of a configuration, the higher its probability. Therefore, finding the most probable configuration of an GRF is another energy minimizing task.

The main difference between a MRF and a GRF lies in their characterization. A MRF is characterized by the Markovianity, i.e. its local property, whereas a GRF is characterized by the Gibbs distribution, i.e. its global property. As already said, the motivation for using GRF is the possibility of defining the energy function in terms of clique potentials. Fortunately, the Hammersley-Clifford theorem stated in [Hammersley and Clifford, 1971] establishes the equivalence between MRFs and GRFs. The theorem says that the set of random variables F is a MRF on the set of sites $\mathcal{S}$ w.r.t. the neighborhood system $\mathcal{N}$ if and only if F is a GRF on the same set of sites $\mathcal{S}$ w.r.t. the same neighborhood system $\mathcal{N}$.

For more information on the GRF see [Li, 2009].

3.3.5 Models of Markov Random Fields

The Hammersley-Clifford theorem allows a very general basis for the specification of MRF joint distribution function based on clique potentials. One particular, important and frequently used group of MRF models are called *auto-models* by [Besag, 1974]. The form of auto-models can be derived from the dual energy equation (Eq. 3.25) when $D_p = f_p G_p(f_p)$ and $V_{p,q} = \beta_{p,p} f_p f_q$. The $G_p(f_p)$ are unspecified functions and $\beta_{p,q}$ denotes pre-defined model parameters reflecting the pair-site interactions. The energy function of auto-models is then characterized as follows:

$$E(x) = \sum_{\{p\} \in \mathcal{C}_1} f_p G_p(f_p) + \sum_{\{p,q\} \in \mathcal{C}_2} \beta_{p,q} f_p f_q \quad (3.27)$$

The auto-models can be further classified w.r.t. the specifications of individual terms in energy function (Eq. 3.27). Probably the most used are the *Ising* and the *Potts* model that are described in greater detail in following text. More information about individual models can be found in [Li, 2009].

3.3.5.1   Ising Model

One of the basic features of our world is smoothness. The world preserves smoothness not only in time but also in space assuming that physical properties in close neighborhood or in a short interval of time do not change abruptly. Transferring this assumption to the domain of image processing can be expressed as an assumption that site labels in close neighborhood should be the same. In other words, changing labels should be penalized in some way. And exactly this assumptions is encoded in Ising and Potts models.

The Ising model originated in statistical physics and was first introduced in [Ising, 1925]. Its use for image analysis is described for example in [Geman and Geman, 1984b] and [Winkler, 1995]. It is a binary model which means that each site can be assigned with one out of two labels, i.e. $\mathcal{L} = \{0, 1\}$ or sometimes $\mathcal{L} = \{-1, +1\}$. Another assumption is that only the first order neighborhood system (Fig. 3.2a) is used. The energy function can be then written as follows:

$$E(f) = \sum_{\{p\} \in \mathcal{C}_1} \alpha_p f_p + \sum_{\{p,q\} \in \mathcal{C}_2} \beta_{p,q} f_p f_q, \qquad (3.28)$$

where $\beta_{p,q}$ are called the interaction coefficients. In the case where homogeneity is preserved the coefficients α_p and $\beta_{p,q}$ are independent of the position of site p, i.e. $\alpha_p = \alpha$ and $\beta_{p,q} = \beta$.

The Ising model allows defining the interaction potentials in the form of label discontinuities. The parameter β can be interpreted as a penalty for discontinuity and the interaction potentials can be defined as follows:

$$V_{p,q}(f_p, f_q) = \beta |f_p - f_q| = \beta \delta(f_p, f_q) \qquad (3.29)$$

where $\delta(f_p, f_q)$ is the Kronecker delta:

$$\delta(f_p, f_q) = \begin{cases} 1 & \text{if } f_p \neq f_q \\ 0 & \text{if } f_p = f_q \end{cases} \qquad (3.30)$$

This model prefers configurations with large-scale agreement between the labels of adjacent sites, i.e. it prefers few large areas in the image over many smaller ones. The interaction term can also be interpreted as a penalty for boundary - the total penalty of the interaction term is the total perimeter of boundaries separating individual regions times β. Having this in mind, the interaction potential can be rewritten in the following form:

$$V_{p,q}(f_p, f_q) = \beta \cdot (\text{boundary length}) \qquad (3.31)$$

thus making the longer boundaries less probable. The boundary length is given by Manhattan (D_4) distance.

The influence of the parameter β is shown in Figure 3.4, where a few samples from the Gibbs distribution that corresponds to the Ising model are shown.

3.3.5.2   Potts Model

The Ising model described in the previous section is usable where there are only two possible labels, e.g. binary image segmentation. However, many vision task need to solve problems

(a) $\beta = 0.7$ (b) $\beta = 1.1$ (c) $\beta = 2$
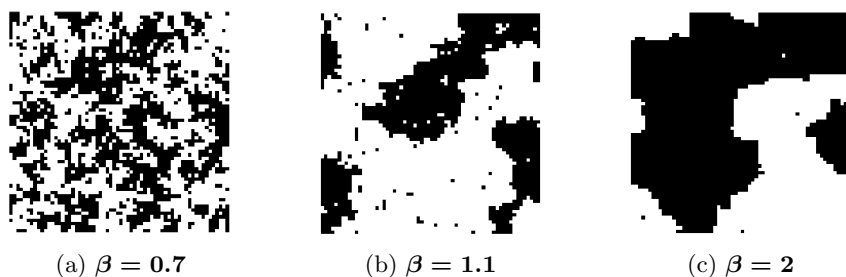
Figure 3.4: Samples from Gibbs distribution of an Ising model for different values of the parameter β, [Perez, 1998].

with more than two discrete labels. For this purpose, the Ising model can be generalized to the Potts model, where there are $M > 2$ discrete labels in the label set $\mathcal{L} = \{1, 2, \ldots, M\}$. The Potts model was introduced in [Potts, 1952] and further extended in [Boykov et al., 1998]. The interaction potentials can be defined as follows:

$$V_{p,q}(f_p, f_q) = \left\{ \begin{array}{ll} \beta_c & \text{if } f_p \neq f_q \\ -\beta_c & \text{if } f_p = f_q \end{array} \right. \tag{3.32}$$

The parameter β_c denotes the cost of an edge connecting nodes with different labels and depends on the type c of the clique. In the case of isotropic model, the clique potentials are independent of orientation, which means that $\beta_c = \beta, \ \forall c \in \mathcal{C}_2$. As β increases, regions become more homogeneous.

3.3.6 Optimization techniques

Currently, a large number of methods of finding a local minimum are available but finding a global minimum of a function is still a considerable challenge. Methods for global minimization can be based on finding the exact global minimum or an approximation. The use of the former methods usually yields an exhaustive search and inefficient algorithms. Approximation methods are efficient in both the computational load and the size of search space but the result is still only an approximation that could be far away from the exact solution.

The following text describes optimization techniques related to this work.

3.3.6.1 Genetic Algorithms

Genetic algorithms are methods based on heuristic approaches introduced for example in [Holland, 1975] and [Goldberg, 1989]. They are inspired by the principle of natural evolution in the biological world, where the rule of the survival of the fittest takes place. Natural selection is guided by the quality of individuals ensuring that better individuals will survive into the next generation.

Genetic algorithm is parallel in its nature - it stores several solutions called *individuals*; the set of all solutions in a given time is called a *population*. One individual is determined by its *chromozome*, i.e. a set of *genes*. A value of a gene is often called an *allele*. In the labeling
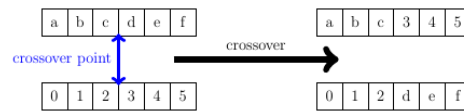
Figure 3.5: Illustration of the crossover operator based on a point.

process one gene can represent a certain site and its allele value corresponds to one of the class labels. The fitness of the individual then corresponds to the energy of the given labeling, which can be calculated in the sense of MRF. When most (typically over 90%) of individuals share the same value of one particular gene, then we say that this gene has converged. When all the genes have converged, then we say that the population has converged.

Recombination

To produce new individuals, i.e. possible solutions, a process called recombination is applied. Recombination could be seen as an unary or a binary operator. The latter one is called *crossover*. The input of the crossover are two individuals that are combined in a certain way to create two offspring. The parents are chosen w.r.t. their fitness, i.e. the better the individual, the higher the probability that it will be chosen; for more information see [Bäck, 1996]. This simulates the natural selection. With this in mind, the offspring bear the information of a possible good solution. Furthermore, the combination of two such a individuals can lead to a new, even better one.

The unary recombination process is *mutation*. Its role is to slightly change one or more genes of an individual to produce a different one. Thus, the mutation has a similar role as the temperature parameter in simulated annealing, i.e. to prevent getting stuck in a local minimum.

A scheme of a standard genetic algorithm is shown in Algorithm 1.

The recombination processes are not always invoked. The probability of crossover is somewhere between 0.6 and 1 [Li, 2009] and the probability of mutation is typically 0.0001. To define where the crossover takes place, several approaches exists. One such an approach is to randomly derive a crossover point. To this point, all genes take value from the first parent and after this point, from the second parent as shown in Fig. 3.5. In the case of images an approach based on a random mask can be used. In this case, a binary mask is derived that specifies from what parent a gene takes value as shown in Fig. 3.6. The disadvantage of this approach is that it could break homogeneous areas with no obvious progress. The mutation operator is shown in Fig. 3.7.

To prevent situations where a good solution can be lost due to recombination, one can use a process called *elitism*. Here, a portion (typically 10% of the total number of individuals) is saved and then the same portion of the worst individuals is replaced. One thing that should be highlighted is that having few bad individuals is not necessarily a disadvantage. These individuals bear different information that could produce, in combination with a good individual, a solution that can outperform all others. Moreover, bad individuals are also helpful for leaving a local minimum and for searching a different part of the solution space.
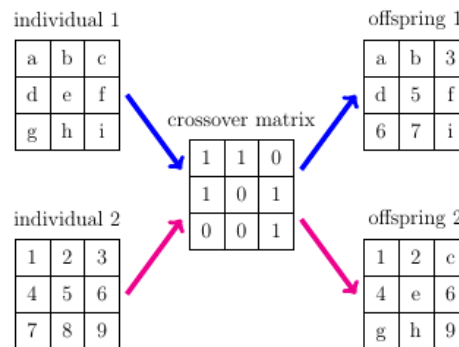
Figure 3.6: Illustration of the crossover operator based on a matrix.
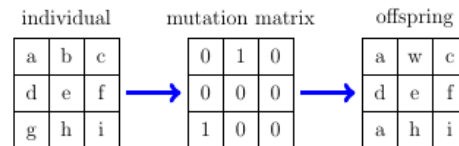
Figure 3.7: Illustration of the mutation operator.

Locust Swarm

Though the main principle remains the same, many different genetic algorithms exists that differs for example in the chromozome representation or in the way recombination is done. A well known example of such a different approach is locust-based methods first introduced in [Kennedy and Eberhart, 1995] and further used, for example, in [Park and Song, 1998] and [Amelio and Pizzuti, 2012]. This principle was used for image segmentation in [Jedlicka and Ryba, 2016].

IN this approach, the allele values do not represent class labels. Instead, they represent other genes, thus making links between each other. The more similar two neighboring genes are, the higher the probability that a link will be created between them. One gene can have more than one allele value and recombination is done by switching an combining those alleles. The final image segments are represented by individual connected components that are created as the algorithm evolves.

The inspiration of this algorithm comes from the behavior of a swarm. A swarm is formed by many individuals of the same species, i.e. bees or grasshoppers, where each individual follows only a few of his neighbors. The swarm behavior is then given by combination of the behavior of several smaller groups.

3.3.6.2 Min Cut / Max Flow

The main goal of this method is to find a minimum cut in a graph that represents an image. Finding such a minimum cut is not a trivial task. Fortunately, the minimum cut problem is equivalent to the maximum flow problem, as stated in the Ford-Fulkerson theorem introduced

```
begin
    generate initial population P = {f¹, ..., fⁿ}
    repeat
        compute E(f) ∀f ∈ P
        // creating new population
        P_new = {}
        repeat
            fᵃ, fᵇ = individuals_selection(P)
            // crossover
            if rand[0,1] < τ_crossover then
            |   f¹, f² = crossover(fᵃ, fᵇ)
            else
            |   f¹ = fᵃ
            |   f² = fᵇ
            end
            // mutation
            for fⁱ ∈ {f¹, f²} do
            |   if rand[0,1] < τ_mutation then
            |   |   fⁱ = mutation(fⁱ)
            |   end
            end
            P_new = P_new ∪ {f¹, f²}
        until new population created
        P = P_new
    until termination condition satisfied
    return(arg min_{f∈P} E(f))
end
```

Algorithm 1: A standard genetic algorithm

in [Ford and Fulkerson, 1956]. Finding a maximum flow belongs to the classical optimizing problems.

The first step is to obtain a weighted graph $\mathbf{G}(\mathcal{V}, \mathcal{E})$ from the image, where $\mathcal{V}$ represents a set of nodes and $\mathcal{E}$ represents a set of edges. Each node corresponds to an image site, i.e. $\mathcal{V} = \{p | p \in \mathcal{S}\}$ where $\mathcal{S}$ is traditionally a set of all image sites. Each edge between these nodes corresponds to a neighboring relationship between these nodes, i.e. $\mathcal{E} = \{e_{p,q} | p, q \in \mathcal{V}, p \in \mathcal{N}_q\}$. These edges are often called *n-links* (neighborhood links). The graph contains two additional nodes, the source $\boldsymbol{s}$ and the sink $\boldsymbol{t}$, respectively. The former one represents an object while the latter one the background. These special nodes are called terminals. Each node from the set $\mathcal{V}$ is connected to the terminals with a link called *t-link* (terminal link). An illustration of a graph constructed this way is shown in Figure 3.9(c).

We can imagine the graph like a pipeline system (in the basic version of Graph cut method). Each pipe has its own capacity given by the weight of the corresponding edge.
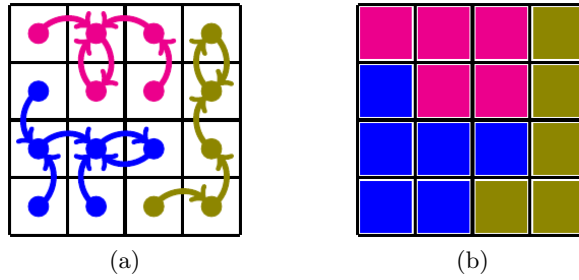
Figure 3.8: In lucust swarm methods allele values represent links to neighbors (a). The given individual represents segmentation shown in (b).
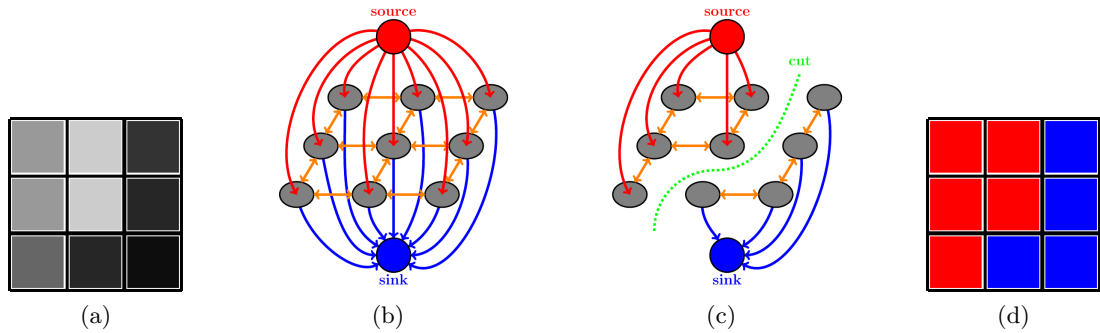
Figure 3.9: Illustration of image segmentation by graph cut. The input image (a) and its corresponding graph (b) is segmented by a cut shown in (b) which yields the final segmentation (d).

Segmentation consists in gradual filling this pipeline with a medium through the source and the leakage of this medium through the sink. Some of the pipes become saturated in this process, which means that through this pipe flows the maximum value of the medium given by its capacity. After a certain amount of the medium was sent to the system, there can occur a state in which there is no path from the source to the sink without any saturated edge. In this case, the segmentation ends and the object is separated from the background by the set of saturated edges. This set of saturated edges represents a cut of the graph.

In other words, a cut on a graph is a set of edges $\mathcal{C} \subset \mathcal{E}$ such that the terminal nodes are separated in the induced graph $\mathcal{G}(\mathcal{C}) = \{\mathcal{V}, \mathcal{E} - \mathcal{C}\}$. Moreover, there is no subset $\mathcal{C}'$ of $\mathcal{C}$ that will separate the terminals in $\mathcal{G}(\mathcal{C})$. The cost of a cut is denoted as $|\mathcal{C}|$ and corresponds to the sum of all edges in this cut. This value is equal to the value of maximum flow that can be sent to the graph as said in the already mentioned Ford-Fulkerson theorem ([Ford and Fulkerson, 1956]). Figure 3.9 illustrates the process of image segmentation, where the input image and its corresponding graph is segmented by a cut, which yields the final segmentation.

The energy minimized by this process takes the typical form given by Equation 3.25. One common example how to define the data term (Equation 3.33) and the smoothness term

(Equations 3.34 and 3.35) that yields neither the Ising nor the Potts model is formulated in [Boykov and Jolly, 2001]:

$$V_{p,q}(f_p, f_q) = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\frac{1}{||p,q||}\right),\qquad(3.33)$$

$$D_p(p|obj) = -\ln\left(P(I_p|Obj)\right),\qquad(3.34)$$

$$D_p(p|bgd) = -\ln\left(P(I_p|Bgd)\right),\qquad(3.35)$$

where p, q are two different image sites that are $||p,q||$ far apart and have intensities I_p and I_p, σ can be interpreted as expected intensity variation within the object and/or background. The probabilities $P(I_p|Obj)$ and $P(I_p|Bgd)$ describe how the intensity value I_p fits to object or background, respectively. Following these equations, one can assign weights to individual edges as shown in following table:

edge	weight	for
(p,q)	$V_{p,q}(f_p, f_q)$	$(p,q) \in \mathcal{N}$
(s,p)	$\lambda D_p(p\|bgd)$	$p \in \mathcal{S}, p \notin (O \cup B)$
	K	$p \in O$
	0	$p \in B$
(p,t)	$\lambda D_p(p\|obj)$	$p \in \mathcal{S}, p \notin (O \cup B)$
	0	$p \in O$
	K	$p \in B$

Table 3.1: Edge weights for graph cut image segmentation. Constant K is big enough so that the given edge never gets saturated.

There are two basic algorithms for finding the maximum flow in a graph. The first one is the so called *push-relabel* algorithm introduced by [Goldberg and Tarjan, 1988], the second one is based on *augmenting paths* and was first presented in [Ford and Fulkerson, 1956] and [Ford and Fulkerson, 1962]. A relatively new approach which outperforms the former ones is formulated in [Boykov and Kolmogorov, 2004]. This algorithm is based on two competitive trees that "grow" in image domain and try to occupy image sites.

More information about optimization of the MRF by the graph cut methods can be found, for example, in [Boykov and Jolly, 2000], [Boykov and Jolly, 2001], [Kolmogorov and Boykov, 2005], [Boykov and Veksler, 2006] and [Boykov and Funka-Lea, 2006].

3.3.6.3 Graph Cut With Large Moves

In paper [Boykov et al., 2001] the problems of finding a global optimum are discussed. A labeling f is said to be a local minimum of the energy E if

$$E(f) \leq E(f') \text{ for any } f' \text{ "near to" } f. \tag{3.36}$$

The nearness of two labelings is defined in the sense of a single optimization move. Many optimization algorithms can change label only of one site at a time - these moves are called *standard moves* in [Boykov et al., 2001]. This often yields poor results as shown in Figure 3.10. In contrary, algorithms described in this section use large moves that allows changing labels of arbitrarily large set of image sites simultaneously. Thanks to that, the algorithms are very efficient in approximating energy minimization.

| (a) | (b) | (c) | (d) |
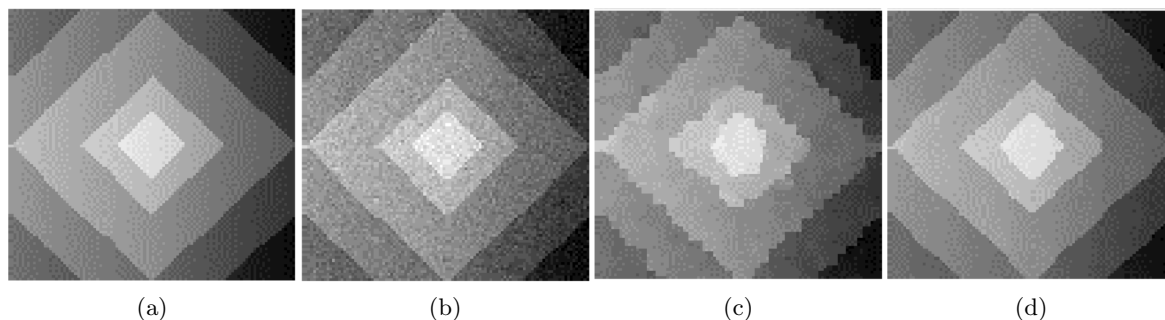
Figure 3.10: Comparison of image restoration. The input image is shown in (a) and its noisy observation in (b). Using standard moves yields to low quality solution (c). The solution (d) is obtained using expansion moves. Images were taken from [Boykov et al., 2001].

Expansion Moves

The alpha expansion algorithm is based on successive labeling of all α and non-α sites with graph cuts. In each iteration, an optimal expansion move is determined that expands the α label in an optimal way. The algorithm then continues with another label. The pseudocode of α expansion algorithm is shown in alg. 2.

In each iteration, a graph $\mathcal{G}_\alpha$ is constructed that reflects current labeling. The set of vertices contains not only image sites $p \in \mathcal{S}$ but two terminals α and $\bar{\alpha}$ as well. In addition, an *auxiliary* node is created between each pair of neighboring sites that have different label. Each site is connected to both terminals with links t_p^α and $t_p^{\bar{\alpha}}$. Between each pair of neighboring sites $\{p, q\} \in \mathcal{N}$ with different labels $f_p \neq f_q$ a triplet of edges $\mathcal{E}_{\{p,q\}} = \{e_{\{p,a\}}, e_{\{a,q\}}, t_a^{\bar{\alpha}}\}$ is created, where a is the corresponding auxiliary node. An example of a graph constructed this way is shown in Fig. 3.11 and the weights of individual edges are shown in Table 3.2.

Start with an arbitrary labeling f
begin
    repeat
        for each label $\alpha \in \mathcal{L}$ do
            $\hat{f} = \arg\min E(f')$ **among f' within one α expansion move of f**
            if $E(f') < E(f)$ then
                $f = f'$
                improved = True
            else
                improved = False
            end
        end
    until not improved
end

Algorithm 2: Alpha expansion

To use this algorithm, the interaction term $V(\alpha, \beta)$ have to be a metric. $V(\alpha, \beta)$ is called a metric if it satisfies the following three conditions:

$$V(\alpha, \beta) = 0 \quad \Leftrightarrow \quad \alpha = \beta \tag{3.37}$$

$$V(\alpha, \beta) \quad = \quad V(\beta, \alpha) \tag{3.38}$$

$$V(\alpha, \beta) \quad \leq \quad V(\alpha, \gamma) + V_2(\gamma, \beta) \tag{3.39}$$

for any labels $\alpha, \beta, \gamma \in \mathcal{L}$. The use of expansion moves in case where the interaction term is a metric brings a very strong optimality property. This property says that a local minimum found by the expansion moves is within a known factor c of the global minimum - described in [Boykov et al., 2001] as the following theorem:

Theorem 1. *Let $\hat{f}$ be a local minimum when the expansion moves are allowed and f^* be the globally optimal solution. Then $E(\hat{f}) \leq 2cE(f^*)$.*

The factor c is defined in the next equation:

$$c = \frac{\max_{\alpha \neq \beta \in \mathcal{L}} V_2(\alpha, \beta)}{\min_{\alpha \neq \beta \in \mathcal{L}} V_2(\alpha, \beta)} \tag{3.40}$$

If the interaction term is dependent on a neighboring pairs (p, q), then the previous equation is changed to the following:

$$c = \max_{\{p,q\} \in \mathcal{N}} \left(\frac{\max_{\alpha \neq \beta \in \mathcal{L}} V_2(\alpha, \beta)}{\min_{\alpha \neq \beta \in \mathcal{L}} V_2(\alpha, \beta)} \right) \tag{3.41}$$

In a special case the interaction term is given by the Potts model (described in Section 3.3.5.2). In this case, discontinuities between any pair of labels are penalized equally, thus the factor c is given by Equation 3.40. Moreover, as shown in [Boykov et al., 2001] the factor $c = 1$ and thus the local minimun is within a factor of two of the global minimum.
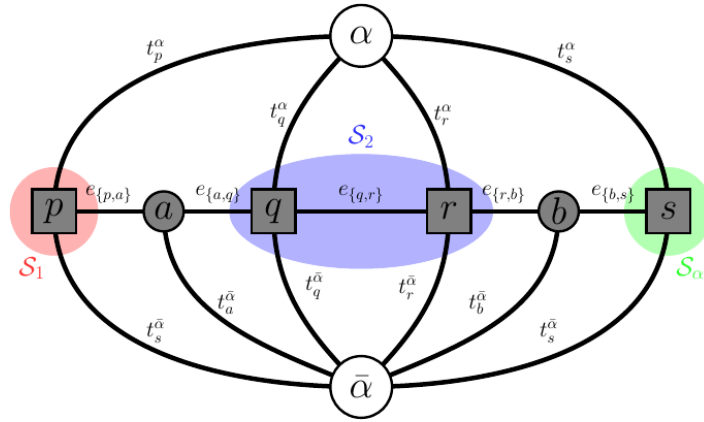
Figure 3.11: An example of $\mathcal{G}_\alpha$ with sites p, q, r and s, terminals α and $\bar{\alpha}$ and two auxiliary nodes a and b. Three different partitions are shown in this graph: $\mathcal{S}_1 = \{p\}$, $\mathcal{S}_2 = \{q, r\}$ and $\mathcal{S}_\alpha = \{s\}$. The original image can be found in [Boykov et al., 2001].

Swap Moves

The alpha-beta swap algorithm is based on successive labeling of all α and β sites with graph cuts. In each iteration, an optimal swap move is determined that changes labels between segments α and β. This is done successively for each pair of labels $\{\alpha, \beta\} \in \mathcal{L}$. The pseudocode of α-β swap algorithm is shown in alg. 3.

Start with an arbitrary labeling f
begin
  repeat
    for each pair $\{\alpha, \beta\} : \alpha, \beta \in \mathcal{L}$ do
      $\hat{f} = \arg\min E(f')$ among f' within one α-β swap move of f
      if $E(f') < E(f)$ then
        $f = f'$
        improved = True
      else
        improved = False
      end
    end
  until not improved
end

Algorithm 3: Alpha-beta swap

Similarly as in the expansion algorithm, a graph $\mathcal{G}_{\alpha\beta} = \{\mathcal{V}_{\alpha\beta}, \mathcal{E}_{\alpha\beta}\}$ is constructed in each iteration. Let the sites with label α and β be denoted $\mathcal{S}_\alpha$ and $\mathcal{S}_\beta$, respectively. Next, let the set of sites with label α or β be denoted $\mathcal{S}_{\alpha\beta} = \mathcal{S}_\alpha \cup \mathcal{S}_\beta$. Then the set of vertices contains two terminals α and β and sites with label α or β. Each vertex is connected to the terminals

edge	weight	for
$t_p^{\bar{\alpha}}$	∞	$p \in \mathcal{S}_\alpha$
$t_p^{\bar{\alpha}}$	$D_p(f_p)$	$p \notin \mathcal{S}_\alpha$
t_p^{α}	$D_p(\alpha)$	$p \in \mathcal{S}$
$e_{\{p,a\}}$	$V_{p,a}(f_p, \alpha)$	
$e_{\{a,q\}}$	$V_{a,q}(\alpha, f_q)$	$\{p,q\} \in \mathcal{N},\ f_p \neq f_q$
$t_a^{\bar{\alpha}}$	$V_{p,q}(f_p, f_q)$	
$e_{\{p,q\}}$	$V_{p,q}(f_p, \alpha)$	$\{p,q\} \in \mathcal{N},\ f_p = f_q$

Table 3.2: Edge weights for a $\mathcal{G}_\alpha$ used in α expansion algorithm.

by edges t_p^{α} and t_p^{β}, respectively. Between each pair of neighboring sites $\{p, q\} \in \mathcal{N}$ an edge $e_{\{p,q\}}$ is created. An example of a graph constructed this way is shown in Fig. 3.12 and the weights of individual edges are shown in Table 3.3.
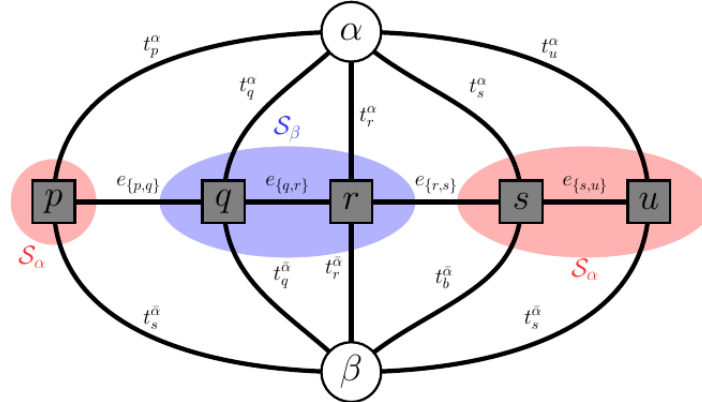
Figure 3.12: An example of $\mathcal{G}_{\alpha\beta}$ with sites p, q, r, s and u and terminals α and β. Two label segments are shown in this graph: $\mathcal{S}_\alpha = \{p, s, u\}$ and $\mathcal{S}_\beta = \{q, r\}$. The original image can be found in [Boykov et al., 2001].

While in the α expansion algorithm has a guaranteed optimality properties, a solution found by α-β swap algorithm can be arbitrarily far from the global minimum. On the other hand, it can be used for optimization of wider class of energies because the interaction term could be "only" a semimetric. An interaction term is said to be a semimetric when Equations 3.37 and 3.38 are satisfied, i.e. the triangular inequality does not have to be met.

In the paper [Boykov et al., 2001] the possibility of achieving a solution within a known factor from the global optimum even when the interaction term is a semimetric is discussed. The approach suggests to approximate the semimetric and use α expansion algorithm to get a

edge	weight	for
t_p^α	$D_p(\alpha) + \sum_{q \in \mathcal{N}_p,\, q \notin \mathcal{S}_{\alpha\beta}} V_{p,q}(\alpha, f_q)$	$p \in \mathcal{S}_{\alpha\beta}$
t_p^β	$D_p(\beta) + \sum_{q \in \mathcal{N}_p,\, q \notin \mathcal{S}_{\alpha\beta}} V_{p,q}(\beta, f_q)$	$p \in \mathcal{S}_{\alpha\beta}$
$e_{\{p,q\}}$	$V_{p,q}(\alpha, \beta)$	$\{p, q\} \in \mathcal{N},\ p, q \in \mathcal{S}_{\alpha\beta}$

Table 3.3: Edge weights for a $\mathcal{G}_{\alpha\beta}$ used in α-β swap algorithm.

(a)   (b)   (c)   (d)
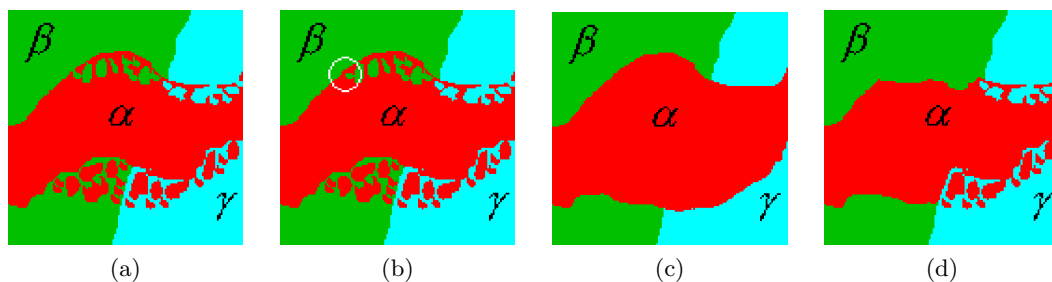
Figure 3.13: Comparison of standard and large moves from given labeling (a). The α expansion (c) and the α-β swap (d) change many site labels simultaneously where one standard move of simulated annealing (b) changes the label of a single site (in the circled area). The images were taken from [Boykov et al., 2001].

local minimum. This minimum is then used as a starting point for the successive application of α-β swap algorithm.

3.4   Geometric active contours - Level Sets

Active contours became one of the most popular algorithms that can be applied to a wide range of problems. This approach can be classified into two main classes: parametric (PAC) and geometric / geodesic (GAC). The former one is based on evolving a curve and its drawback is the possibility that a sharp corner might arise, for example. The GAC, that were first presented by [Caselles et al., 1993] and [Malladi et al., 1993], are immune to these kind of problems. The algorithm is based on an evolution of a curve as well, but unlike the parametric case, the calculations are based on geometric properties only. The curve that represents the borders of an object of interest can be interpreted as a zero level set of a higher-dimensional function. Therefore, these methods are frequently referred to as level sets.

The evolution does not depend on curve's parameters; therefore, the GAC are often referred to as implicit ([Sonka et al., 2007], [Xu et al., 2000]). The evolution of the curve is driven by a criterion that can be based on an image gradient (edges) or on image regions.

3.4.1   Stopping criterion based on edges

Let a moving curve be parameterized by time t and a parameter s: $X(s,t) = [x(s,t), y(s,t)]$, where $x(s,t)$ and $y(s,t)$ are its coordinates. Let N be an inward normal and c the curve's curvature. Then the evolution of the curve along its normal direction is given by the *speed function* $V(c)$ and can be described by the following partial equation ([Sonka et al., 2007]):

$$\frac{\delta X}{\delta t} = V \cdot N. \tag{3.42}$$

In the case where the speed function V is constant, see Fig. 3.14b, singularities like sharp corners may arise. To overcome this problem the speed function is defined as a function of curvature c, i.e. $V = V(c)$, see Fig. 3.14c.

| (a) | (b) | (c) |
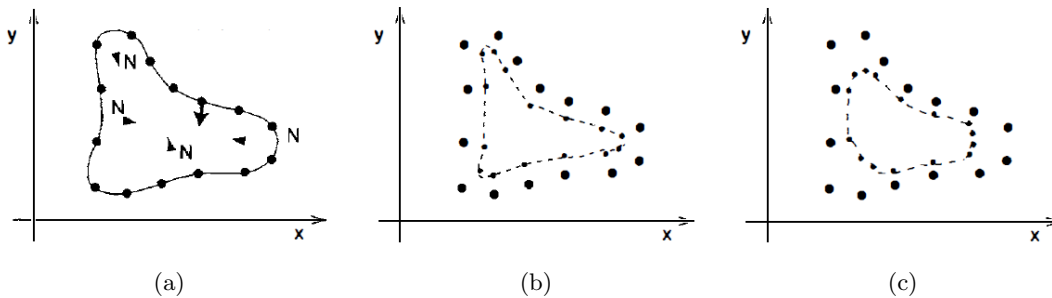
Figure 3.14: The evolution of a curve (a) with a constant speed function (b) and a speed function based on curvature (c). Images from [Sonka et al., 2007].

The speed function drives the curve's evolution, therefore we can say that it is the core of this method. In publications [Caselles et al., 1993] and [Malladi et al., 1993], a speed function

combining constant deformation with curvature is described:

$$\frac{\delta\phi}{\delta t} \quad = \quad k(c + V_0)|\nabla\phi|, \tag{3.43}$$

$$k \quad = \quad \frac{1}{1 + |\nabla(G_\sigma * I)|}, \tag{3.44}$$

where **phi** represents the moving curve and the term $\nabla(G_\sigma * I)$ denotes image I smoothed by a gaussian filter G with smoothness parameter σ. The k is a parameter that slows down the speed of the curve's evolution in areas with high image gradient. This way, the method tends to stop its evolution on object borders.

A significant drawback of this approach are situations in which the curve is not slowed down enough and cross areas with high gradient. In such situations, the speed of the curve's evolution is speeded up again and the curve starts to move away from the desired boundary. There is no force that would pull the curve back to the boundary. A possible solution is presented in [Caselles et al., 1997] and [Yezzi et al., 1997] is to enhance the speed function by another term. The role of this term is to pull the curve back in case of crossing over a boundary. After the enhancing, Equation 3.43 takes the following form:

$$\frac{\delta\phi}{\delta t} = k(c + V_0)|\nabla\phi| + \nabla k \nabla\phi \tag{3.45}$$

To be able to effectively drive the curve's evolution, the curve is represented as a zero level set of higher-dimensional function $\phi(t)$, as already noted. The curve can then be specified by solving Equation 3.46.

$$\phi(X(s,t), t) = 0 \tag{3.46}$$

The function ϕ is defined in such a way that inside the curve (i.e. inside an object of interest) it takes negative values and positive values outside the curve. The initialization of this function is frequently done by taking the distance transformation of the initial curve. The inward normal is then defined by Equation 3.47.

$$N = -\frac{\nabla\phi}{|\nabla\phi|} \tag{3.47}$$

By taking derivative in time of the Equation 3.46 we obtain the following form:

$$\frac{\delta\phi}{\delta t} + \nabla\phi \frac{\delta X}{\delta t} = 0 \tag{3.48}$$

The final equation for the curve's evolution is given by Equations 3.43, 3.48, 3.47:

$$\frac{\delta\phi}{\delta t} = V(c)|\nabla\phi|, \tag{3.49}$$

where the curvature c can be defined by Equation 3.50.

$$c = \nabla\frac{\nabla\phi}{|\nabla\phi|} = \frac{\phi_{xx}\phi_y^2 - 2\phi_x\phi_y\phi_{xy} + \phi_{yy}\phi_x^2}{(\phi_x^2 + \phi_y^2)^{3/2}} \tag{3.50}$$

3.4.2   Stopping criterion based on regions

Another approach for driving the evolution of the curve is to use a functional that takes image regions into account. Probably the most used approach of this kind was presented in [Chan and Vese, 2001]. This method is based on the minimization of a so-called *Chan-Vese energy functional* defined by Equation 3.51.

$$
\begin{aligned}
C(\phi, a_1, a_2) &= C_1(\phi, a_1, a_2) + C_2(\phi, a_1, a_2) = \\
&= \int_{\text{inside}(\phi)} (I(x,y) - a_1)^2 \mathrm{dxdy} + \int_{\text{outside}(\phi)} (I(x,y) - a_2)^2 \mathrm{dxdy},
\end{aligned}
\tag{3.51}
$$

where the constants a_1 and a_2 represent the mean intensity value of the object and the background. It is easy to see that the functional is minimized when the curve accurately follows the boundary of the object, see Fig. 3.15. In this case $C_1(\phi, a_1, a_2) \approx 0$ and $C_2(\phi, a_1, a_2) \approx 0$.

(a) $C_1(\phi) > 0$, $C_2(\phi) \approx 0$ (b) $C_1(\phi) \approx 0$, $C_2(\phi) > 0$

(c) $C_1(\phi) > 0$, $C_2(\phi) > 0$ (d) $C_1(\phi) \approx 0$, $C_2(\phi) \approx 0$
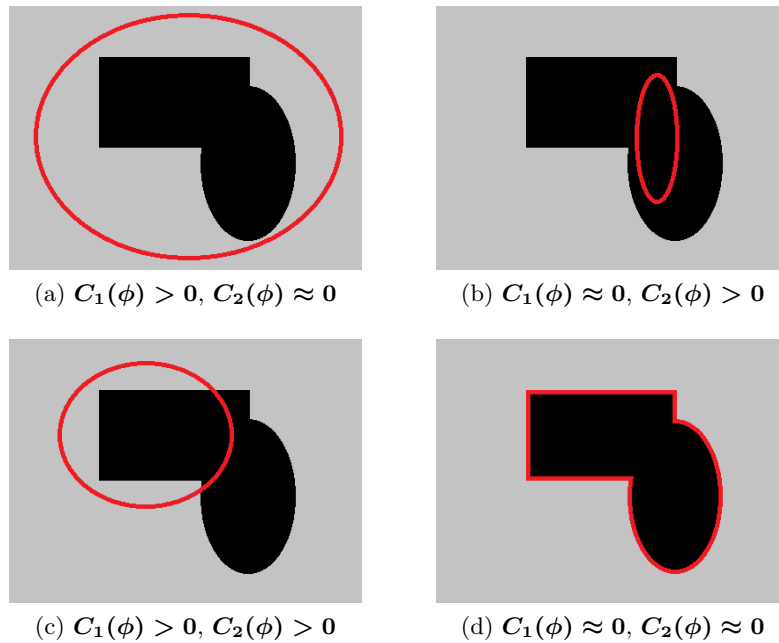
Figure 3.15: An example of Chan-Vese functional as shown in [Sonka et al., 2007].

This functional can be further extended by adding other terms that could penalize the length of the boundary, the areas inside the curve or directly the terms C_1, C_2. In that case, Equation 3.51 takes the following form:

$$
C(\phi, a_1, a_2) = \mu(\text{length}\phi) + \nu(\text{area inside}\phi) + \lambda_1 C_1(\phi, a_1, a_2) + \lambda_2 C_2(\phi, a_1, a_2),
\tag{3.52}
$$

where $\mu, \nu, \lambda_1, \lambda_2 \geq 0$.

An example showing the use of GAC can be found in Fig. 3.16.

(a) initialization   (b) after 20 iterations   (c) after 40 iterations
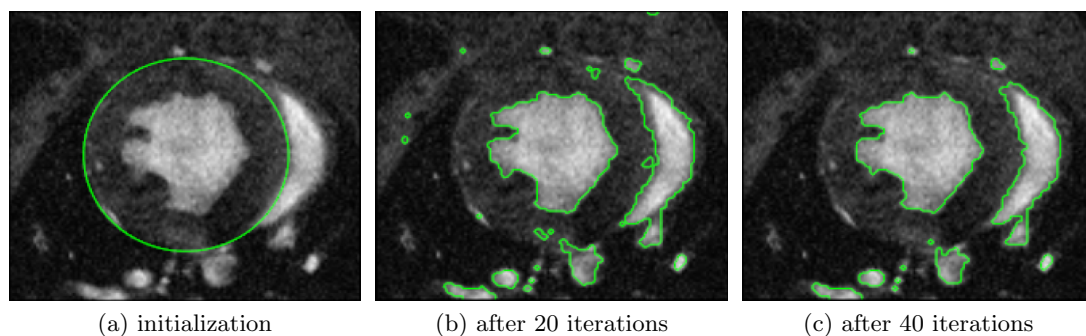
Figure 3.16: Segmentation by GAC as shown in [Mosaliganti et al., 2009].

3.4.3   Localization of calculations

To speed up the calculations it is possible to use some kind of localization. The best-known approach is probably the *narrow band*, which was first presented in [Malladi et al., 1995]. Here, the level set function is calculated only on small surroundings of the zero level set and not on the whole image domain. See Fig.3.17 for illustration.

(a)   (b)
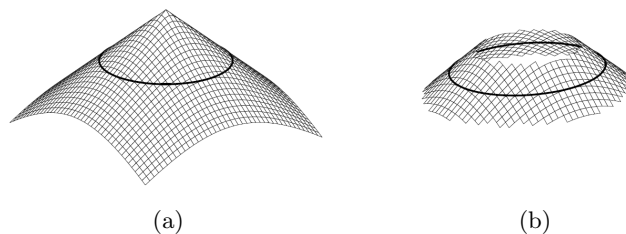
Figure 3.17: Localization of GAC (a) using the narrow band (b). Images from [Adalsteinsson and Sethian, 1995].

The publication [Lankton and Tannenbaum, 2008] presents another approach based on image regions. The localization is applied not only on the small area around the zero level set but on a small area around of currently updating curve's point as well. Therefore, the method can also be applied for segmentation tasks, where object and/or background are non-homogeneous. In this situation, the Chan-Vese functional produces the wrong results due to the inadequate global mean intensity values. This is ilustrated in Fig. 3.18.

In [Lankton and Tannenbaum, 2008] the level set function is initialized by an approximation of smoothed Heaviside function, see Equation 3.53.

(a)   (b)   (c)
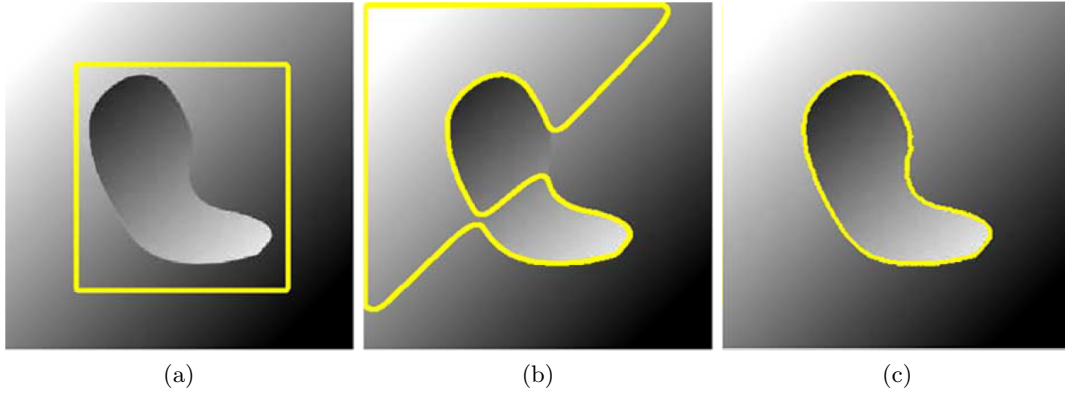
Figure 3.18: Comparison between traditional (b) and localized (c) GAC that were initialized by the curve (a). Images come from [Lankton and Tannenbaum, 2008].

Inside the curve:
$$H\phi(x)_{\text{in}} = \begin{cases} 1, & \Phi(x) < -\epsilon \\ 0, & \Phi(x) > \epsilon \\ \frac{1}{2}\left\{1 + \frac{\Phi}{\epsilon} + \frac{1}{\pi}\sin\left(\frac{\pi\Phi(x)}{\epsilon}\right)\right\}, & \text{otherwise} \end{cases} \tag{3.53}$$
Outside the curve:
$$H\phi(x)_{\text{out}} = (1 - H\phi(x)_{\text{in}})$$

The narrow band is defined as a derivative of above mentioned Heaviside function:

$$\delta\phi(x) = \begin{cases} 1, & \Phi(x) = 0 \\ 0, & |\Phi(x)| > \epsilon \\ \frac{1}{2\epsilon}\left\{1 + \cos\left(\frac{\pi\Phi(x)}{\epsilon}\right)\right\}, & \text{otherwise} \end{cases} \tag{3.54}$$

The localization on a curve point is achieved by a circular neighborhood $B(x,y)$ as shown in Fig. 3.19:

$$B(x,y) = \begin{cases} 1, & ||x - y|| < r \\ 0, & \text{jinak} \end{cases} \tag{3.55}$$

where x and y denote two image points (not coordinates of a point) and r denotes the radius of the circular neighborhood used.

Now it's possible to define an energy that is used for minimization:

$$E(\Phi) = \lambda \int_{\Omega_x} \delta\phi(x)||\nabla\Phi(x)||\mathrm{dx} + \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x,y)F(I(y), \Phi(y))\mathrm{dxdy}. \tag{3.56}$$

The first term of Equation 3.56 penalizes the length of the curve. The second term is used for masking out the neighborhood for each point x of the curve and penalizes it w.r.t. the function $F(.)$. This function can be defined in several ways - as the already stated Chan-Vese

(a)   (b)
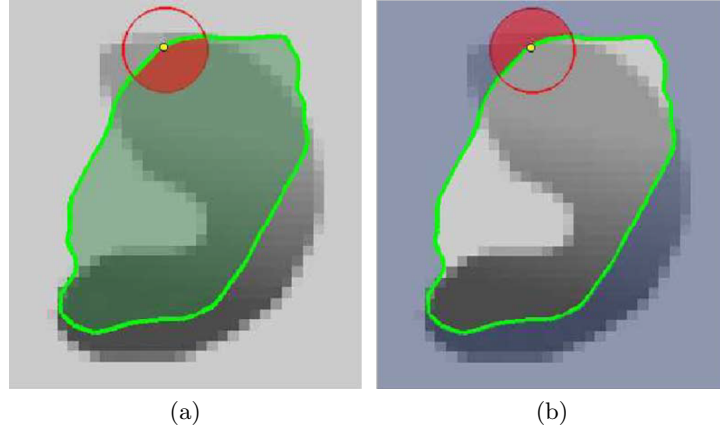
Figure 3.19: Illustration of point (yellow) localization. The red area in (a) defines the inner part, whereas in (b) the outer part of neighborhood. Images come from [Lankton and Tannenbaum, 2008].

functional [Chan and Vese, 2001] or as an energy separating mean values presented in [Yezzi et al., 2002].

In the latter case, the function $\boldsymbol{F(.)}$ takes the following form:

$$\boldsymbol{F}_{\mathrm{MS}} = (\boldsymbol{u_x} - \boldsymbol{v_x})^2, \tag{3.57}$$

where $\boldsymbol{u_x}$ and $\boldsymbol{v_x}$ is the localized mean value inside/outside the neighborhood of the current point as shown in Fig. 3.19:

$$\boldsymbol{u_x} = \frac{\int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)} \boldsymbol{H}\phi(\boldsymbol{y}) \boldsymbol{I(y)} \mathrm{dy}}{\int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)} \boldsymbol{H}\phi(\boldsymbol{y}) \mathrm{dy}} \tag{3.58}$$

$$\boldsymbol{v_x} = \frac{\int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)}(1 - \boldsymbol{H}\phi(\boldsymbol{y})) \boldsymbol{I(y)} \mathrm{dy}}{\int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)}(1 - \boldsymbol{H}\phi(\boldsymbol{y})) \mathrm{dy}} \tag{3.59}$$

By taking the derivative of Equation 3.57 and substituting it in Equation 3.56, we obtain the final form of energy of the localized GAC is achieved :

$$\begin{aligned} \frac{\delta\phi(\boldsymbol{x})}{\delta \boldsymbol{t}} &= \delta\phi(\boldsymbol{x}) \int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)}\delta\phi(\boldsymbol{x}) \left(\frac{(\boldsymbol{I(y)} - \boldsymbol{u_x})^2}{\boldsymbol{A_u}} - \frac{(\boldsymbol{I(y)} - \boldsymbol{v_x})^2}{\boldsymbol{A_v}} \right) \mathrm{dy} + \\ &+ \boldsymbol{\lambda}\delta\phi(\boldsymbol{x})\mathrm{div}\left(\frac{\boldsymbol{\nabla}\phi(\boldsymbol{x})}{|\boldsymbol{\nabla}\phi(\boldsymbol{x})|} \right), \end{aligned} \tag{3.60}$$

where $\boldsymbol{A_u}$ and $\boldsymbol{A_v}$ represents areas of circular neighborhood inside/outside the curve, see Equations 3.61 and 3.62.

$$\boldsymbol{A_u} = \int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)} \boldsymbol{H}\phi(\boldsymbol{y}) \mathrm{dy} \tag{3.61}$$

$$\boldsymbol{A_v} = \int_{\boldsymbol{\Omega_y}} \boldsymbol{B(x,y)}(1 - \boldsymbol{H}\phi(\boldsymbol{y})) \mathrm{dy} \tag{3.62}$$
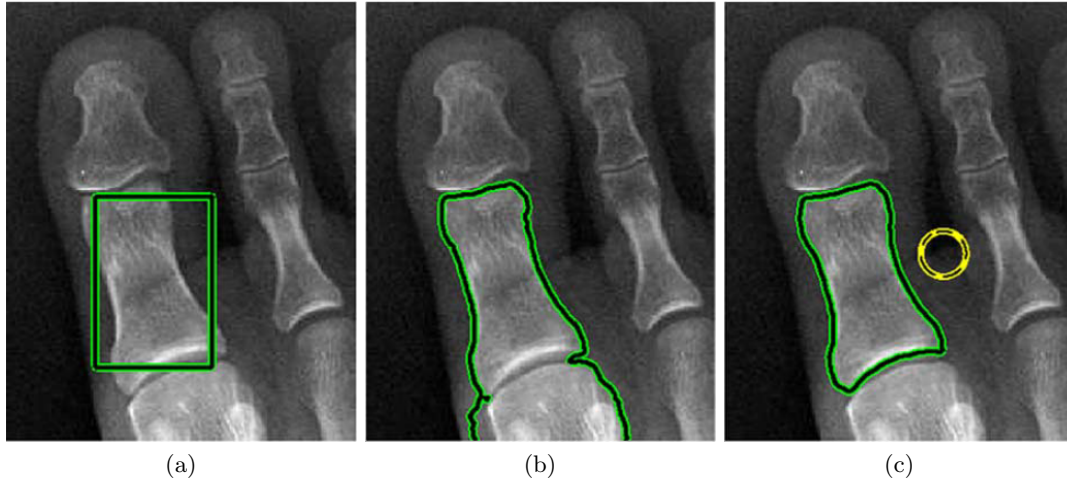
Figure 3.20: The comparison between the traditional GAC (b) and the localized GAC (c). Both methods were initialized by the curve (a). The circle in the localized GAC denotes the size the circular neighborhood used. The images come from [Lankton and Tannenbaum, 2008].

3.5 Grow Cut

Grow cut is a relatively new segmentation technique based on a cellular automaton. An interactive variant of grow cut algorithm is described in [Vezhnevets and Konouchine, 2004]. The initialization of this method is made by seed-points selection. Usually, two types of seed-points are used. One type for the foreground and second for the background, see Fig. 3.21. However, grow cut can be generally used for multiple class segmentation as shown in Fig. 3.22.

A cellular automaton A is defined as triplet $A = \{S, N, \delta\}$, where S denotes state set, N used neighborhood (von Neumann or Moor neighborhood) and δ transition function (or evolution rule). The state $S(p)$ of a cell p is also a triplet $S(p) = \{l_p, \Theta_p, C_p\}$, where l_p denotes the label or connection to an object class, Θ_p cell strength, and C_p feature vector. In our case, the feature vector consists of just one feature - corresponding voxel density. However, in the general case, this vector can consist of many features. For example, the cell's location or a number of cells belonging to the same class can be taken into consideration.

Let $I = \{p : p \in \Omega\}$ denotes our image defined on image domain Ω and let E be set of seed-points $E = \{p : p \in I\}$ and $E(p) = 1 \ldots K, \forall p \in E$, where K is a number of

(a)   (b)   (c)
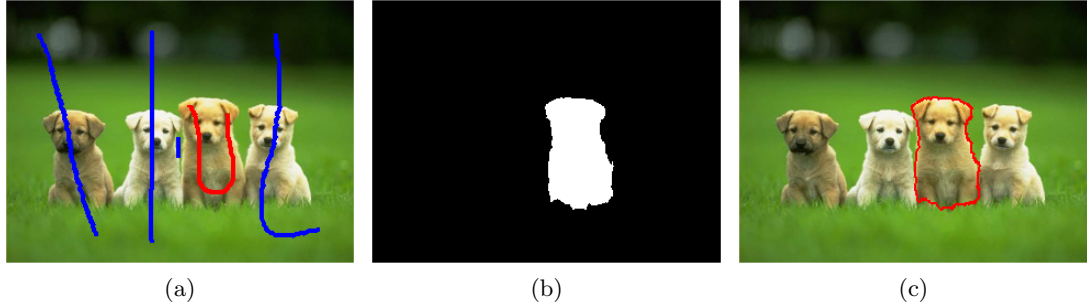
Figure 3.21: Binary segmentation using grow cut. Input image (a) contains red object seeds and blue background seeds. The segmented object is shown in (b) and (c).

(a)   (b)   (c)
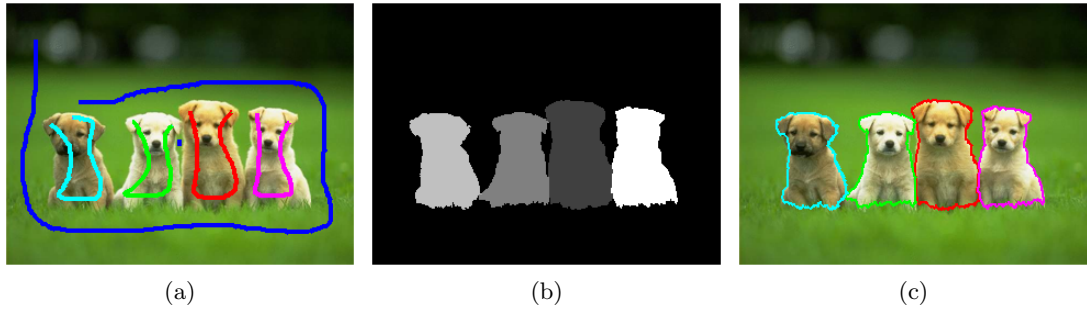
Figure 3.22: Multiclass segmentation using grow cut. Input image (a) contains blue background seeds and multiple object seeds. The final segmentation is shown in (b) and (c).

classes in which the segmentation is made. The algorithm can then be initialized as follows:

$$\forall p \in I \setminus E : l_p = 0, \Theta_p = 0, C_p = I(p), \tag{3.63}$$
$$\forall p \in E : l_p = E(p), \Theta_p = 1, C_p = I(p), \tag{3.64}$$

At each discrete time step, each active cell (cells with non-zero label) tries to attack its neighbors and both labels and strengths are updated according to the following scheme:

Function g denotes a monotonous decreasing function defined by Equation 3.65.

$$g(x) = 1 - \frac{x}{max||\vec{C}||_2}, g(x) \in< 0; 1 > \tag{3.65}$$

The algorithm can be interpreted with inspiration in biology as growth and struggle of K types of bacteria for domination according to their strengths. The strength of current cell depends on the strength of cell, that took control over it and the distance between their feature vectors.

The algorithm is of iterative character and ends if no change in labels or strengths is made in the current iteration step. Of course, it can be terminated when a predefined number of iterations is achieved.

```
begin
    // for each cell
    for ∀p ∈ I do
        // Copy previous state
        l_p^{t+1} = l_p^t
        θ_p^{t+1} = θ_p^t
        // neighbors try to attack current cell
        for ∀q ∈ N(p) do
            f_q = g(||C_p − C_q||_2) × θ_q^t
            if f_q > θ_p^t then
                l_p^{t+1} = l_q^t
                θ_p^{t+1} = f_q
            end
        end
    end
end
```

Algorithm 4: Evolutionary scheme of the grow cut

The final segmentation is often "hairy" (non-smoothed); therefore, post-processing of the final segmentation is recommended. A median filter can be used, for example. Another approach is to smooth the labels during the algorithm run, i.e. to perform a so called cells smoothing. If, for a given cell, the number of enemies (cells with a different label) inside an circular area with radius R is higher than a predefined threshold T, then this cell loses its label and is assigned to a class most frequent in its neighborhood. The threshold can be determined as a relative part of the circular mask. This approach eliminates isolated cells and smooths the boundary between classes. A comparison of smoothing methods is shown in Fig. 3.23.

| (a) | (b) | (c) |
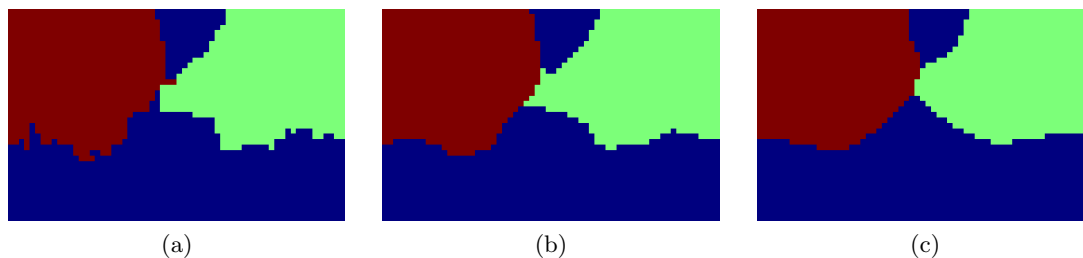
Figure 3.23: Comparison of smoothing a segmentation (a) by a median filter of the size 5x5 (b) and cells smoothing with $T = 0.5 * \text{circ\_area}$, $R = 7$ (c).

3.6 Saliency Maps

One of the biggest problems when analysing a complex visual scene is information overload. Human brain is able to almost instantaneously localize and identify interesting areas in an image. These areas very often differ in some way from the rest of the scene (global difference) or from its background (local difference). The human brain is making decisions which parts of the image contain interesting information and thus should be processed first. At the same time, the brain also suppresses image areas with low information gain. This process is called selective attention. The selective attention is why the human brain is capable of fast image preprocessing, quick reactions to movement in the field of vision.

Great contribution in this area was made in the work [Koch and Ullman, 1985]. Here, an approach is presented that combines several simple visual features into one single oriented map, the saliency map. This approach was further improved in paper [Itti and Koch, 2000] and is described in greater detail in the next section. The original paper uses visual features suited for color images, thus making them inappropriate for grayscale DICOM images. Therefore, new features more suitable for this task were proposed. Few features from the original paper (intensity, texture) were derived from a different approach. All visual features used for deriving the final map are described in greater detail in the following text. In the end, the final set of features makes the localization of tumors easier.

3.6.1 Itti-Koch Saliency

As already stated, this approach is based on a combination of several simple features competing for attention. The work [Itti and Koch, 2000] uses the total number of seven features: intensity, two colors (red-green and blue-yellow double-opponent channels) and four for encoding the orientation (0, 45, 90 and 135°). Each feature generates a so-called conspicuity map. Each point in the conspicuity map then describes its dissimilarity from its neighborhood. By using different feature types and normalizing individual conspicuity maps, the final saliency map is derived. The saliency map constructed this way then reflects the overall degree of attention. Individual steps for building the saliency map are described in the following text.

3.6.1.1 Center-surround computations

Each feature map (i.e. intensity) is computed at 9 different spatial scales by using Gaussian image pyramids, see for example [Sonka et al., 2007]. The center-surround computations are realized as a difference between a fine and a coarse scale. The center of the receptive field corresponds to a pixel in pyramid levels $c \in 2, 3, 4$. The surround to the corresponding pixel is located at pyramid level $s = c + \delta$, where $\delta \in 3, 4$. This gives us the total number of 6 feature maps for pyramid level pairs 2-5, 2-6, 3-6, 3-7, 4-7, 4-8. By using all of the 7 features, the total number of 42 feature maps is produced.

Figure 3.24: Normalization of image with a salient object. Feature map derived from the input image (a) at iteration 0 (b), iteration 4 (c) and iteration 10 (d), [Itti and Koch, 2000].

3.6.1.2   Normalization

One problem the algorithm needs to deal with is the normalization of the feature maps. Imagine an example where an object produces strong attraction just in one of all the features. Moreover, it is possible that the object would not be salient in all spatial scales. Therefore, there is a significant signal-to-noise problem. If the conspicuity maps were combined into the saliency map without any processing, the isolated saliency peak in a feature would be greatly weakened.

In the work [Itti and Koch, 2000] an iterative normalization process is described. A feature map $\mathcal{M}$ is transformed at each iteration by the following scheme:

$$\mathcal{M}(t+1) = |\mathcal{M}(t) + \mathcal{M}(t) * \mathcal{DOG} - C_{\text{inh}}|_{\leq 0}, \qquad (3.66)$$

where t denotes iteration step, C_{inh} is a constant inhibitory term, $||_{\leq 0}$ discards negative values and $\mathcal{DOG}$ is the Difference of Gaussian filter given by the formula 3.67.

$$\mathcal{DOG}(x,y) = \frac{c_{\text{ex}}^2}{2\pi\sigma_{\text{ex}}^2} e^{-(x^2+y^2)/(2\sigma_{\text{ex}}^x)} - \frac{c_{\text{inh}}^2}{2\pi\sigma_{\text{inh}}^2} e^{-(x^2+y^2)/(2\sigma_{\text{inh}}^x)}, \qquad (3.67)$$

where $\sigma_{\text{ex}} = 2\%$ and $\sigma_{\text{inh}} = 25\%$ of the input image width, $c_{\text{ex}} = 0.5$ and $c_{\text{inh}} = 1.5$. The total number of 10 iterations are computed. The normalization process of a salient object is shown in Fig. 3.24 and normalization of not-so-salient objects is shown in Fig. 3.25.

3.6.1.3   Conspicuity maps

After the normalization, the maps of a given feature are summed across scales to create a conspicuity map. Each conspicuity map is then normalized once again with another ten iterations of Eq. 3.66. This process simulates within-feature spatial competition.
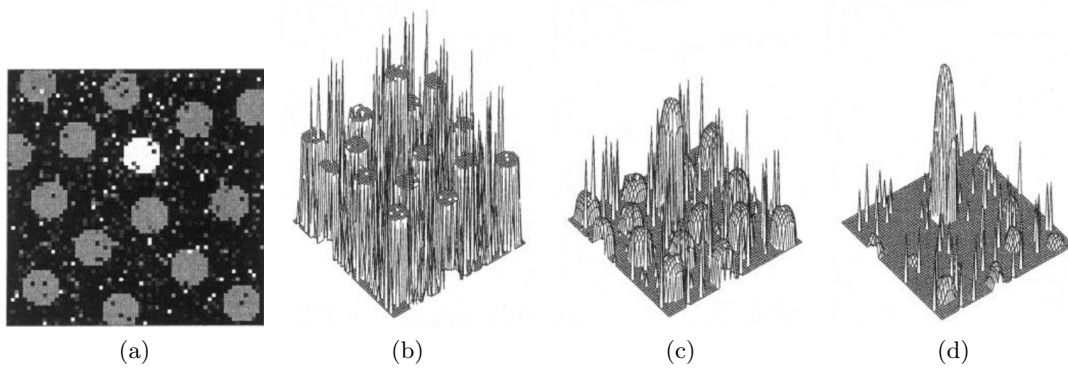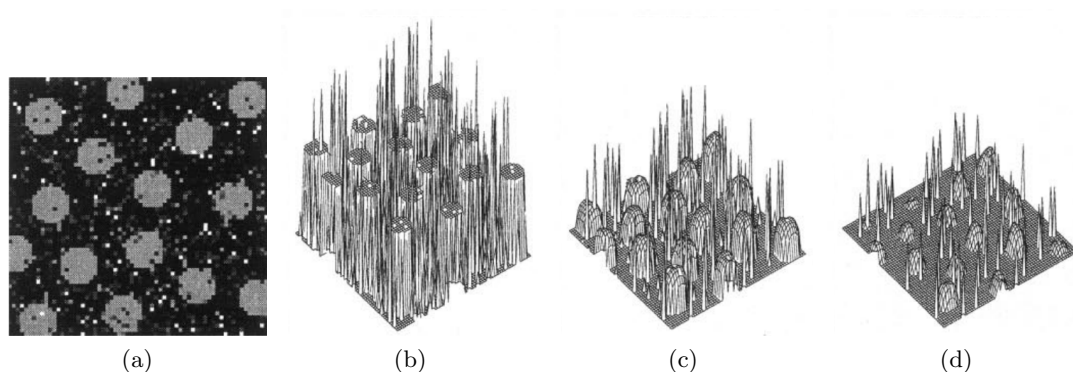
Figure 3.25: Normalization of image with not-so-salient objects. Feature map derived from the input image (a) at iteration 0 (b), iteration 4 (c) and iteration 10 (d), [Itti and Koch, 2000].

3.6.1.4 Saliency map

In the next step, all of the feature maps need to be combined into one final saliency map. The problem is how to combine features that come from different modalities. Therefore, the normalization of feature maps to the interval $[0, 1]$ is realized after the conspicuity map calculation. The individual conspicuity maps are then linearly summed into the final saliency map.

3.6.2 Intensity Conspicuity

In the described algorithm, the intensity conspicuity map (ICM) is designed for color images. Unfortunately, the input data in this work come from CT, i.e. in grayscale DICOM format. Therefore, deriving red-green and blue-yellow color features does not make sense and the ICM needs to be redefined. A total number of four intensity conspicuity maps is presented in this work and described in the following text. For a better signal-to-noise ratio, a logistic (sigmoid) function is used:

$$s(x) = \frac{1}{1 + \exp(-k * (x - c))},$$ (3.68)

where x defines the input signal (conspicuity map), c is the x-value of sigmoid's midpoint and k defines the steepness of the curve. An example of different sigmoids is shown in Fig. 3.26.

3.6.2.1 Intensity difference

If we accept a reasonable assumption that most of the liver parenchyma is formed by healthy tissue, the intensity of healthy tissue will correspond to a position p of a global maxima of intensity histogram. Different image areas could then be identified by subtracting the value
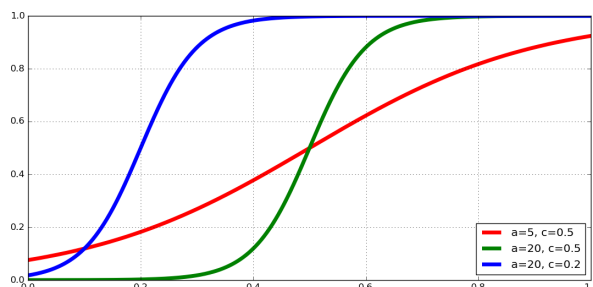
Figure 3.26: Sigmoids used for enhancing the conspicuity maps shown with different parameters.

p from the image: $\boldsymbol{I}_{\text{diff}} = \text{I} - \boldsymbol{p}$. The degree of dissimilarity is given by the absolute value of this difference, where the negative and positive differences defines hyperdense and hypodense areas, respectively. If we are interested in hypodense areas only, it is possible to reset points with negative difference. Figure 3.27 shows conspicuity map obtained by this approach.

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

Figure 3.27: The conspicuity map based on the intensity difference. The final map (b) is calculated from the input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively.

3.6.2.2   Histogram analysis

This approach adopts the same assumption as the previous one, i.e. most of the liver parenchyma is healthy. The algorithm is based on defining a so-called dominant class. The key role of its definition is the localisation of the most frequent intensity in the liver parenchyma, i.e. finding the global peak in the intensity histogram. After that, points with the intensity near to this peak form the dominant class. The whole process is visualized in Fig.3.28. The final binary mask is then inverted and the conspicuity map is defined as the distance transform of the inverted mask. The role of the distance transform is to suppress isolated points. Moreover, to make the conspicuity map to reflect the intensity difference, it is further multiplicated by the image $\boldsymbol{I}_{\text{diff}}$ defined in the previous section. The final conspicuity map is shown in Fig. 3.29.
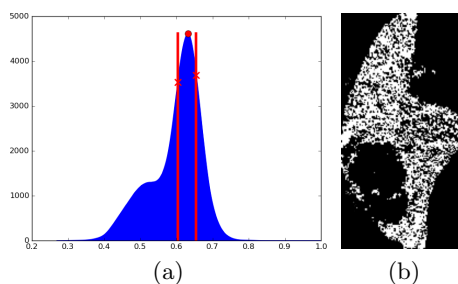
Figure 3.28: Dominant class (b) is derived from histogram (a). First, the global peak (circle) is localized; then, its surroundings are defined (between lines). The dominant class then consists of points with intensity in peak surroundings.

3.6.2.3 GLCM analysis

An obvious disadvantage of using the histogram is the absence of a spatial relationship. A possible solution is analysing the gray-level co-occurrence matrix (GLCM) C. More information about the GLCM can be found, for example, in [Theo, 1982]. For deriving the GLCM, the spatial relationship $\mathcal{R}$ needs to be defined first. A common choice is to use the first order neighborhood system shown in Fig. 3.2a. In that case, the set $\mathcal{R}(p)$ contains all neighbors of a point p.

$$C(i,j) = \sum_{p=1}^{n} \sum_{q \in \mathcal{R}(p)} \begin{cases} 1, & \text{if } I(p) = i \text{ and } I(q) = j \\ 0, & \text{otherwise} \end{cases}, \tag{3.69}$$

where p denotes an image point, $I(p)$ is intensity at image point p and n describes the total number of image points.

After the GLCM is computed, it is then further processed. The first step in this process is thresholding to filter out intensity-pairs that are not frequent. In the resulting image, blobs are found and analysed. Small blobs are discarded and blobs that are to eccentric are split into two smaller ones. Each left over blob then represents a class. Each class is described as a normal distribution with mean value corresponding to the center of the blob and standard deviation is proportional to the ratio of major and minor axes of fitted ellipse. For localizing hypodense tumors only, the blob representing the darkest class is taken into account. Similarly, only the brightest blob is considered when localizing hyperdense tumors. The whole process is shown in Fig. 3.30.

3.6.2.4 Sliding window

The last intensity-based conspicuity map involves a computation over a small image window. This is achieved using the sliding window approach. In the sliding window, points with lower/higher (lower for hypodense tumors, higher for hyperdense tumors) intensity than the mean intensity of the whole liver parenchyma are located. The response of the window is then defined as the ratio of these points inside the window. The response is subsequently assigned to an output image at a location corresponding with the sliding window:

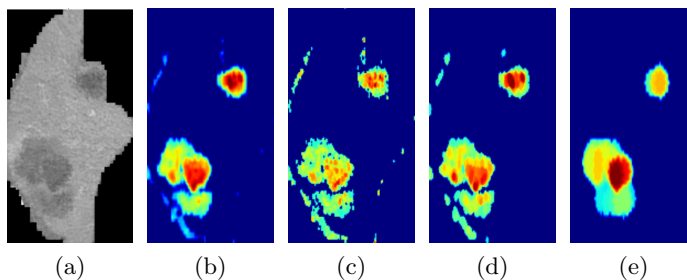(a)   (b)   (c)   (d)   (e)

(f)   (g)   (h)

Figure 3.29: The conspicuity map based on the histogram analysis. The final map (b) is calculated from the input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively. Figure (f) shows distance transform from dominant class, fig. (g) shows $\boldsymbol{I}_{\mathrm{diff}}$ and product of both is shown in Fig. (h).

$$n = \sum_{x=1}^{w} \sum_{y=1}^{h} \left\{ \begin{array}{ll} \boldsymbol{I_w} < \mathbf{mean}(\boldsymbol{I}) & \text{for hypodense tumors} \\ \boldsymbol{I_w} > \mathbf{mean}(\boldsymbol{I}) & \text{for hyperdense tumors} \end{array} \right. , \qquad (3.70)$$

$$\boldsymbol{I}_{\mathrm{out}}(\boldsymbol{i}, \boldsymbol{j}) = \frac{\boldsymbol{n}}{\boldsymbol{w} \times \boldsymbol{h}}, \qquad (3.71)$$

where $\boldsymbol{w}$, $\boldsymbol{h}$ denotes the width and height of sliding window $\boldsymbol{I_w}$, $\boldsymbol{I}$ represents input image, $\boldsymbol{I}_{\mathrm{out}}$ output image (the response) and $(\boldsymbol{i}, \boldsymbol{j})$ are coordinates of the sliding window points in the input image. An example of three different positions of sliding window is illustrated in Fig. 3.31. The final conspicuity map is shown in Fig. 3.32.

3.6.3   Texture Conspicuity

To describe texture, the Local binary patterns (LBP) are used. The LBP algorithm was introduced in [Ojala et al., 1996] and futher extended in [Ojala et al., 2002]. The LBP compute a local representation of texture by comparing each pixel with its surroundings. The surroundings are most often defined as points that lies at a distance $\boldsymbol{R}$ from the investigated point. Another parameter is the number of sampling points $\boldsymbol{P}$. A common choice is to choose $\boldsymbol{P} = \boldsymbol{8} \cdot \boldsymbol{R}$. If $\boldsymbol{R} = \boldsymbol{1}$ we obtain the neighborhood of $\boldsymbol{P} = \boldsymbol{8}$ pixels. We use the center pixel to threshold the neighbors. After the thresholding, the LBP are encoded as a decimal

Figure 3.30: The conspicuity map based on the GLCM analysis. The final map (b) is calculated from the input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively. The GLCM of the input image is thresholded (f) and analysed (g). The normal distribution of the darkest blob is then used for creating a conspicuity map of the hypodense tumors (h)

number following the next equation:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(I(c) - I(p))2^p, \qquad (3.72)$$

where $I(c)$ is intensity of the center point c and function s represents the thresholding defined by the following equation:

$$s(x) = \begin{cases} 1, & \text{if } x \leq 0, \\ 0, & \text{otherwise.} \end{cases} \qquad (3.73)$$

An example illustrating the $LBP_{8,1}$ encoding is shown in Fig. 3.33.

The possible binary patterns that can occur for a given radius R and number of points P are called prototypes. Another extension to the classical LBP is the uniformity. A prototype is considered to be uniform if it has at most two $0 - 1$ or $1 - 0$ transitions. The number of uniform prototypes n_p depends on the number of points P: $n_p = P + 1$. When creating a histogram of the LBP, the number of its bins n must be specified. Most often, the number of the bins is $n = P + 2$, where the first $P + 1$ prototypes represents all uniform prototypes and the last bin tabulates all prototypes that are not uniform.

In this work, we use the values $P = 24$ and $R = 3$ which gives us the total number $n_p = 25$ uniform prototypes and the number of the bin in the histogram $n = 26$. Given this histogram it is possible to choose prototypes that describes a feature, e.g. edge-like,
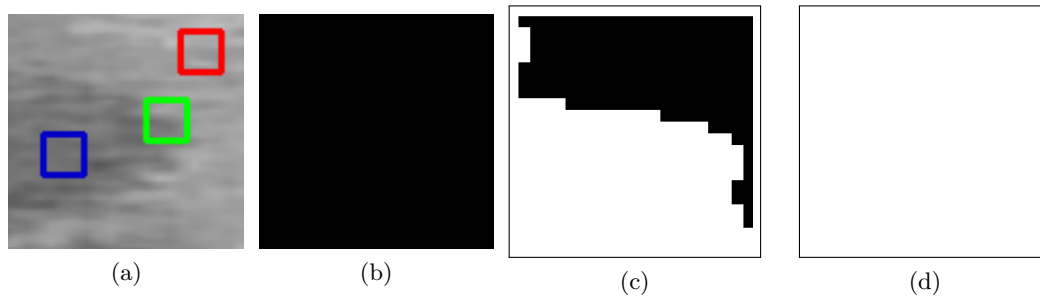
Figure 3.31: Figure shows sliding windows at different position (a). Image (b) shows masked red window, (c) is for green and (d) for blue window.
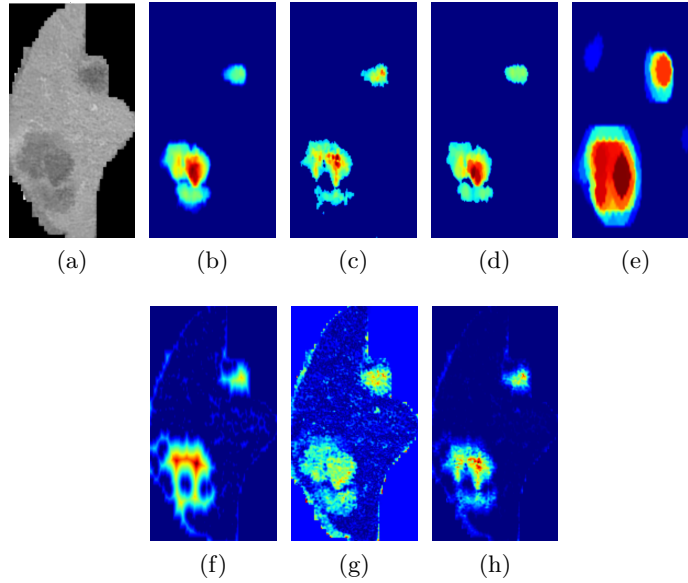

Figure 3.32: The conspicuity map based on the sliding windows. The final map (b) is calculated from the input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively.

Figure 3.33: Example of $\boldsymbol{LBP_{8,1}}$ encoding. Surrounding neighbors (a) are thresholded against their center (b). The LBP value is then given by Eq. 3.72 (c).

corner-like or flat. From our point of view, only the flat prototypes are important because they represent homogeneous regions. From the flat prototypes we can choose only those that describe darker or brighter areas only. The process of deriving texture conspicuity map is shown in Fig. 3.34.
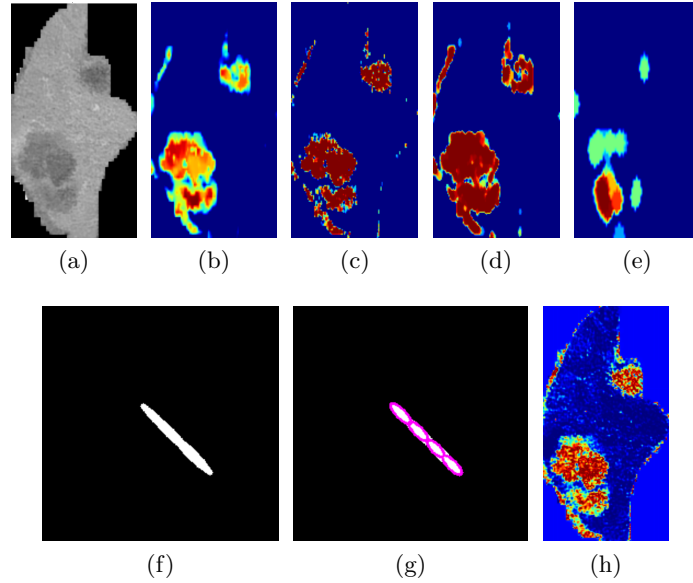

Figure 3.34: The conspicuity map based on the LBP. The final map (b) is calculated from the input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively. A histogram of LBP with highlighted flat prototypes is shown in (f) and visualized masked hypodense flat prototypes are shown in (g).

3.6.4   Shape Conspicuity

A vast majority of tumors are spherical by nature. To use this fact, a saliency measure describing the circularity is proposed in this work. The approach is very similar to texture analysis by the Gabor filters. We defined a set of filters whose response is the measurement of how elliptical the concrete image area is. The filters differ in size of major axis and orientation. To be consistent with the approach of the Gabor filters we call the filter's set the bank of filters (or just a bank).

At first, we calculate the response as intensity difference between the ellipse in the center and the rest of the filter (background). But this approach generates a positive response even if the analysed area was not elliptical. The reason for this is that by calculating the mean value of the entire filter background, we are ignoring any information about intensity distribution. This leads to a high number of false positive responses. Imagine an image area where the upper/lower part of the background mask is of higher/same intensity than the core (ellipse), respectively. The mean value of the background is slightly higher and the response of the filter would be positive (when looking for hypodense tumors), see Fig. 3.35. To overcome

this problem, we split the mask's background into four areas (quaternions) thus creating non-global background mask. Now, the filter response is positive only if the mean value of intensity in each quaternion is higher than the mean intensity of the core (when localizing hypodense tumors). In this case, a large number of false positive responses are suppressed, see Fig. 3.35.

| (a) | (b) | (c) |
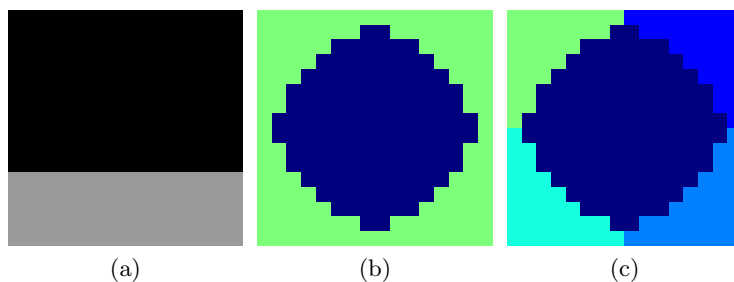
Figure 3.35: The input image (a) is filtered by the global (b) and non-global (c) filter. By using the global filter mask, we obtain a false positive response. On the other hand, by using the non-global mask we obtain a true negative response thanks to the upper quaternions, whose mean intensity is lower than the intensity of the center.

The bank of filters consists of the total number of 5 filters as shown in Fig. 3.36. The first one is a circle and the others are ellipses differing in the size of the major axis and orientation. By using image pyramids with the same filter's bank, this approach is able to localize tumors of different sizes.

The final conspicuity map, overall responses on different pyramid levels and individual filter's responses are shown in Fig.3.37

| (a) | (b) | (c) | (d) | (e) |
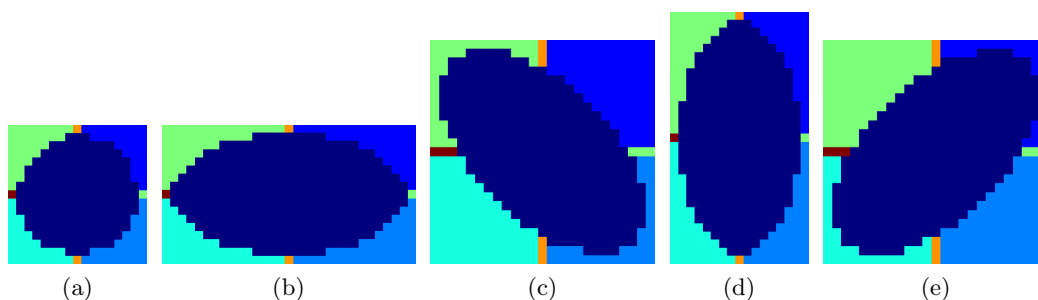
Figure 3.36: Bank of filters used for calculating conspicuity map based on shape. The horizontal and vertical axis of the mask is common for both neighboring quaternions; therefore, they have different labels.
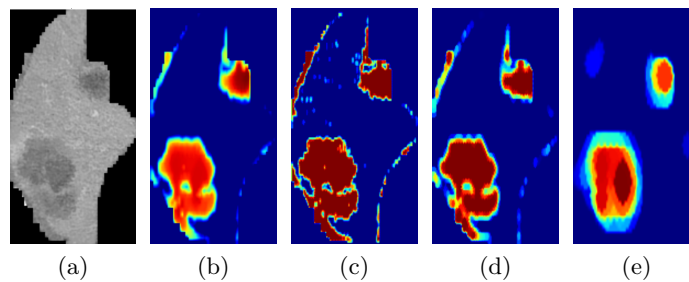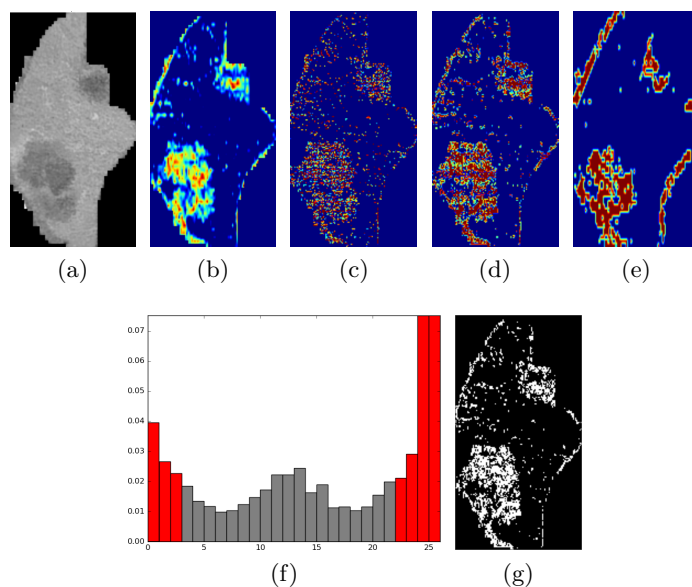
| (a) | (b) | (c) | (d) | (e) |
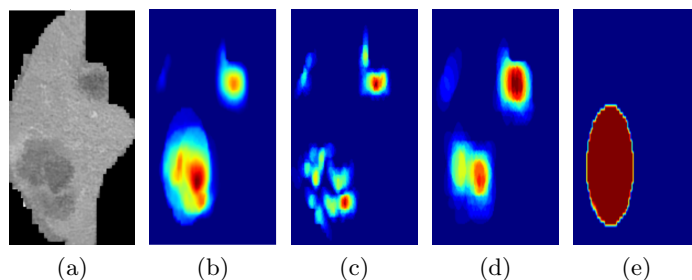
Figure 3.37: Conspicuity map based on shape. The final map (b) is calculated from input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively.

3.6.5 Blob Conspicuity

The previous approach searches the image for elliptical areas that differ in image intensity. Basically, one could look at the approach as a blob detector and it could be described in this section. In this work, this section is dedicated to classic blob detectors combined to a final conspicuity map. The individual algorithms are briefly described in the following text.

3.6.5.1 Laplacian of Gaussian (LoG)

The most accurate and slowest approach is based on a convolution of an image by a Gaussian kernel at a certain scale t:

$$L(x, y; t) = g(x, y; t) * f(x, y), \tag{3.74}$$

$$g(x, y, t) = \frac{1}{2\pi t^2} \exp(-\frac{x^2 + y^2}{2t^2}). \tag{3.75}$$

Then, the Laplacian operator $\nabla^2 L = L_{xx} + L_{yy}$ is applied, which yields an image with strong responses for dark blobs of radius $r = t\sqrt{2}$ and strong negative responses for bright blobs of a similar size.

3.6.5.2 Difference of Gaussians (DoG)

The LoG operator can be approximated by the difference between two Gaussian smoothed images:

$$\nabla^2 L(x, y; t) \approx \frac{t}{\delta t}(L(x, y; t + \delta t) - L(x, y; t)). \tag{3.76}$$

This yields a faster blob detector. On the other hand, both the LoG and the DoG are extremely slow for detecting larger blobs.

3.6.5.3 Determinant of Hessian (DoH)

To overcome the problem with speed of the previous approaches, the DoH can be applied. First, the the determinant of Hessian must be derived. Then, the blobs are found as maxima of

this matrix. The speed of detection is independent on the blobs' size because the convolution is replaced by the box filters.

3.6.5.4 OpenCV (OCV)

OCV blobs are derived following an algorithm from the OpenCV library, see [Itseez, 2015]. First the image is thresholded on different intensity levels. Each binary image is then searched for connected components that are further analysed and filtered by several features (area, color, convexity etc.).

Each of the blob detectors depends on several parameters. For a better localization, each approach is applied with more parameter values and all results are then combined into the final detector response. These responses are further combined in a final conspicuity map. Individual detector responses and the final conspicuity map are shown in Fig. 3.38.

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

(f) (g) (h) (i)
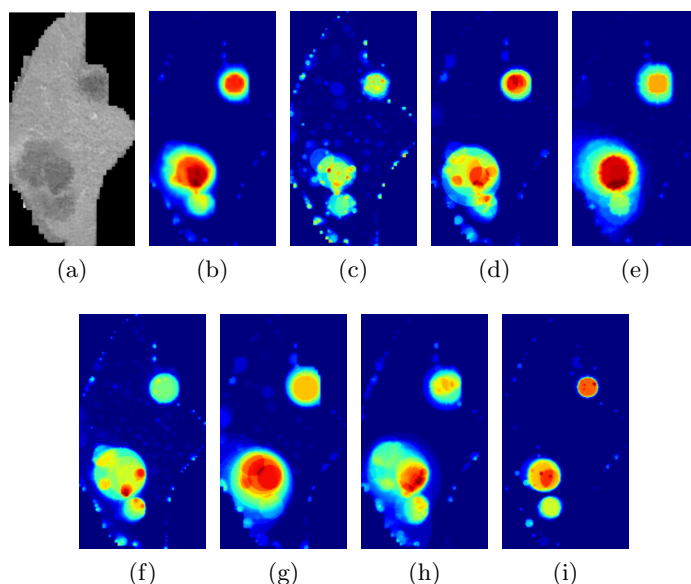
Figure 3.38: Conspicuity map based on blob detection. The final map (b) is calculated from input image (a). Figures (c), (d) and (e) show feature maps at pyramid level 1, 4 and 8, respectively. The final map is a combination of outputs from LoG (f), DoG (g), DoH (h) and OCV (i) blob detectors.

3.6.6 Conclusion

The total number of 7 conspicuity maps is derived - 4x intensity, texture, shape and blob. Some of them are adopted from the work [Itti and Koch, 2000] and mostly redefined; the rest is designed for the task of the localization of the liver tumors. Fig. 3.39 shows all of them.

The next step is to use these maps for the localization itself. In this step, the Markov random fields are used and this process is described in Section 6.1.
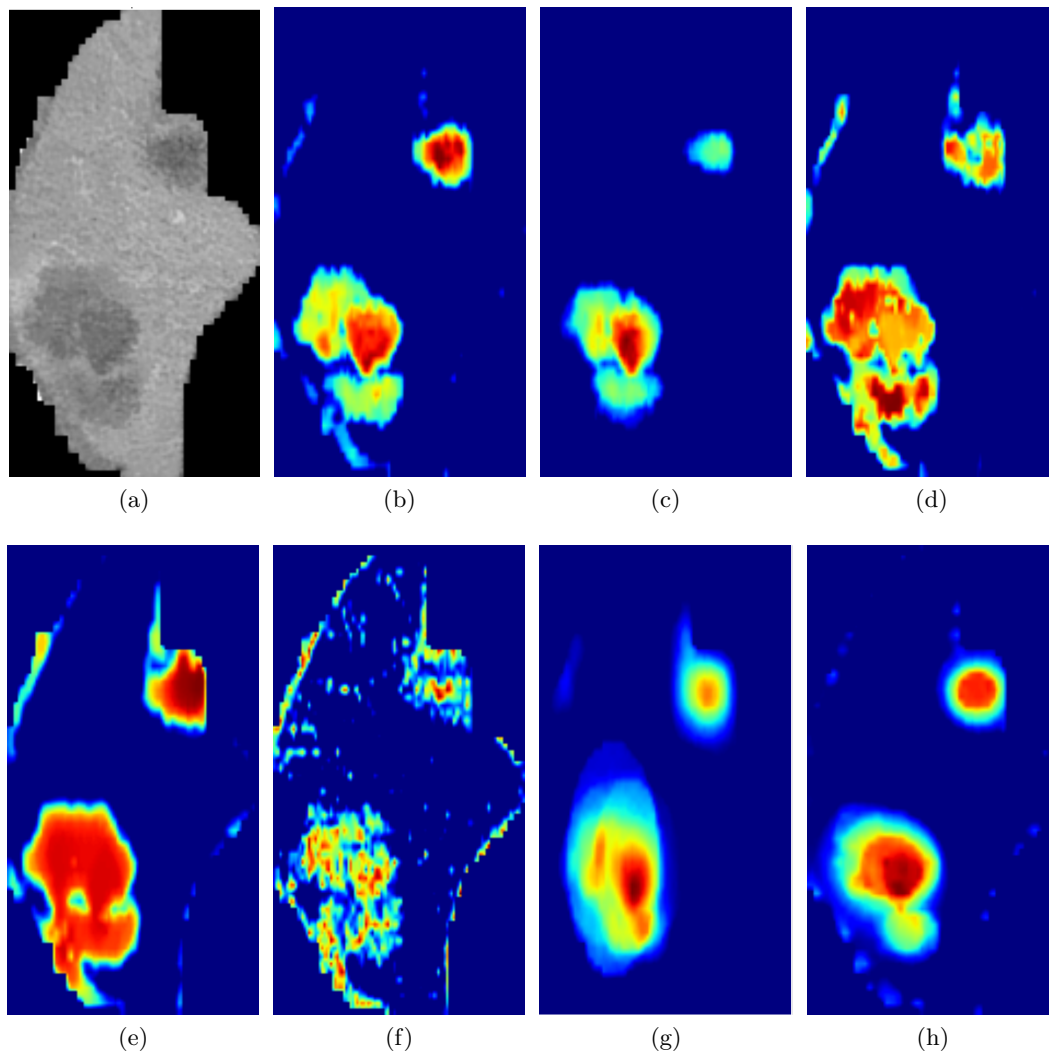
Figure 3.39: The input image (a) and individual conspicuity maps: intensity difference (b), intensity from histogram analysis (c), intensity from GLCM analysis (d), sliding window (e), texture (f), shape (g) and blobs (h).

3.7   Shortest path basins

The motivation for creating this algorithm is to develop an alternative approach for deriving seed points, which is fully autonomous. The method was introduced in [Ryba and Zelezny, 2015] as an autonomous algorithm for image segmentation. Thereafter, in [Ryba and Zelezny, 2016] the method was used as an initialization procedure for segmentation methods requiring seed point. The algorithm is described in more details in following text.

3.7.1   Algorithm overview

The principle of the algorithm is iterative searching for seed points representing individual image classes in the best way. For such an algorithm to be effective it is necessary to define a heuristic function that will control the selection of the seed points. The heuristic used in this paper is based on finding seed points that are maximally different. In other words, each new seed should be as different from the previously chosen seeds as possible. That way, the corresponding classes should be different as well.

When new seed point is derived it is appropriate to define image points that are close to the seed. This area is then penalized so that it is less possible for a new seed point to be chosen from this area, thus making it similar to an already created seed point. The area that contains points that are similar to a seed point is called the basin. Certainly, there are many approaches for defining such a basin. We use the shortest path in a graph; therefore, the basin is called the shortest path basin (SPB).

The core for creating a SPB is the inverse Dijkstra's algorithm. It propagates from a starting point (source) to its neighborhood and for each point derives a value that reflects the closeness of that point to the source. This value is determined as the cost of the shortest path from the source to this point. The algorithm terminates when there are no points that are closer than a predefined maximal distance threshold $\boldsymbol{T}$. In other words, all points that are closer from the source than the threshold $\boldsymbol{T}$ create the SPB of the source. By using the algorithm for computing the shortest path in a graph it is possible to define the proximity of points that depends not only on intensity difference but on geometric distance as well. This is achieved by defining graph edges in a common way:

$$w_{i,j} = 1/\exp\left(-\frac{(I_i - I_j)^2}{2\sigma^2}\right),\qquad(3.77)$$

where $\boldsymbol{I_j}$ is the intensity of a pixel $\boldsymbol{i}$ in an image $\boldsymbol{I}$ and $\boldsymbol{\sigma}$ could be interpreted as expected intensity difference inside the segmented object, [Boykov and Funka-Lea, 2006].

If there are two points with similar intensity that are geometrically far away, there is noticeable chance that these points will be classified into different classes - the geometrical distance outweighs the similarity of intensities. To overcome this problem, a system of siblings is used. Each point becomes a sibling of the source if it lies in a basin and has the same (or very similar) intensity as the source. Each sibling is then taken as another source of the same class and the algorithm can propagate to farther image areas.

A possible alternative to creating basins using the shortest path is to use Euclidean distance, which will be much more efficient. Using this approach yields to the suppression of

intensity values and the basins would be spherical. The main drawback is that points in such a basin could easily belong to different classes, i.e. they could be of very different intensity.

3.7.2   Penalizing energy

The above-mentioned heuristic is based on iterative updating of a so-called overall energy $E_o(t)$, where t denotes the iteration index. This energy reflects the penalization of each seed point that was derived so far. In each iteration, a new seed s is derived and the corresponding SPB needs to be penalized. The energy of the basin of the seed s is denoted $E(s)$ and defined by the costs of shortest path of each point in the basin:

$$E(s) = T - dist(p, s), \ \forall p \in \text{Im} \tag{3.78}$$

where T represents distance threshold for algorithm propagation and $dist(p, s)$ is the distance of an image point p to a seed point s in the sense of the shortest path.

A problem that can easily arise is that a new seed point will be derived near an edge in the image. To overcome this problem, the energy can be calculated as a weighted mean of the energy and image gradient at the same point:

$$E_f(s) = \alpha \cdot E(s) + (1 - \alpha) \cdot \nabla \text{Im} \tag{3.79}$$

Once a seed point is chosen, its energy $E_f(p)$ is added to the overall (accumulative) energy:

$$E_o(t) = E_o(t - 1) + E_f(s) \tag{3.80}$$

where t defines current iteration.

The next seed point will be chosen as a point with the smallest value of this accumulative energy:

$$s(t + 1) = \arg \min_{\forall p \in \text{Im}} (E_o(t)) \tag{3.81}$$

The algorithm terminates after a certain number of iterations or if a minimum of the accumulative energy exceeds the predefined threshold value. The process of iteratively deriving seed points and updating energy $E_O(t)$ is shown in Fig. 3.40.

3.7.3   Experiments

It is generally known that image segmentation is a ill-defined process. There is not only one ground truth in most cases, which makes the objective comparison of segmentation methods a challenging task. Due to this fact, it is reasonable to compare the tested method with several manual segmentations from different users. Fortunately, the Probabilistic rand index (PRI) is a measure for this kind of evaluation and was introduced in [Unnikrishnan and Hebert, 2005].

The main principle of this approach lies in the analysis of the labels of each pixel pair. Consider an input image of N pixels $X = x_1, x_2, \ldots x_i, \ldots, x_N$ and a set of manually segmented (ground truth) images $S_1, S_2, \ldots, S_K$. Let S^t be the segmentation that is to be compared and let the label of a point x_i be denoted as l_i in the segmentation S^t and $l_i^{(k)}$
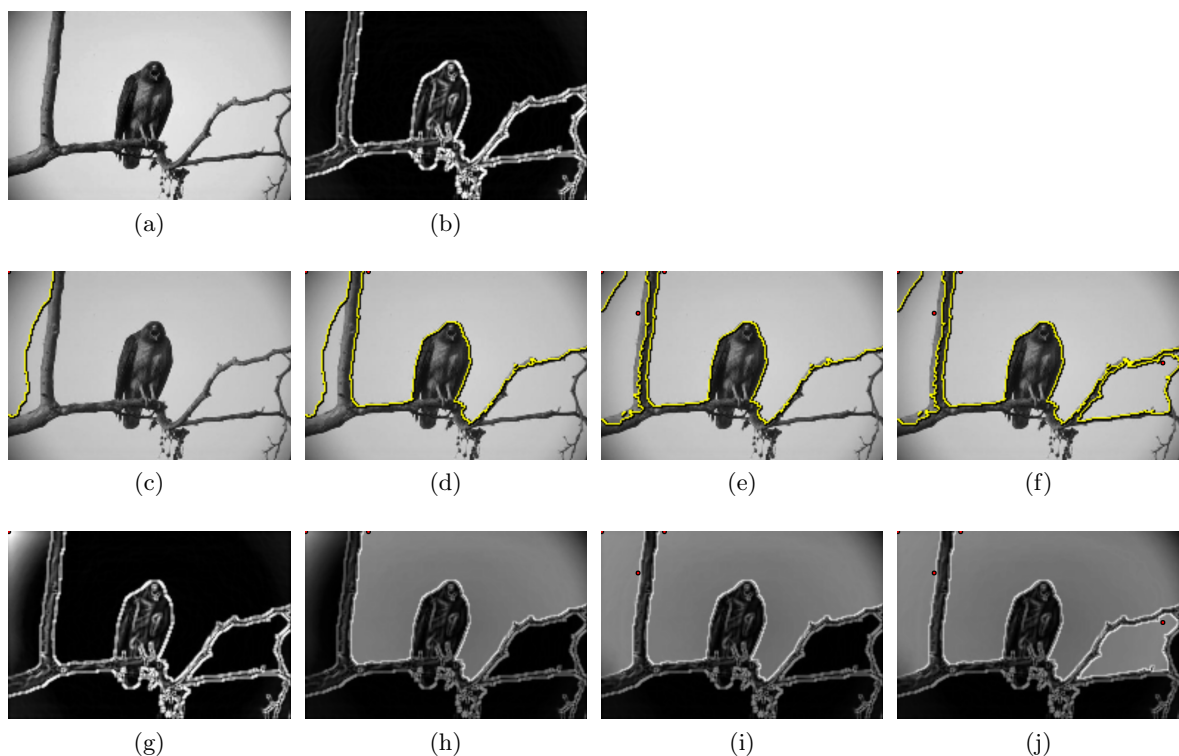
Figure 3.40: Example of deriving seed points in input image (a) with initial overall energy $E_o(t)$ in (b). Figures (c) - (f) show derived seed points and their basins. Figures (g)-(j) show corresponding iteration of the overall energy.

in the manual segmentation S_k. To be able to objectively compare a segmentation S with several manual segmentations, it is necessary to determine expected interrelation $P(l_i = l_j)$ and $P(l_i \neq l_j)$ for each pixel pair x_i and x_j following the next equation:

$$P(l_i = l_j) = \frac{1}{K} \sum_{k}^{K} \mathbb{I}\left(l_i^{(k)} = l_j^{(k)}\right) \tag{3.82}$$

$$P(l_i \neq l_j) = \frac{1}{K} \sum_{k}^{K} \mathbb{I}\left(l_i^{(k)} \neq l_j^{(k)}\right) \tag{3.83}$$

Equation 3.82 represents an expectation that both pixels will have the same label. Equation 3.83 represents an expectation that the pixels will have different labels. These terms reflects the interrelation of pixel pairs through the whole set of manual segmentations.

Table 3.4: Comparison of methods that is based on Probabilistic rand index (PRI). The mean and std values are calculated over 50 different images.

	GC-SPB	RW-SPB	FHC	NC
'man'	0.832	0.849	0.831	0.867
'bird'	0.834	0.835	0.882	0.592
'woman'	0.697	0.745	0.749	0.737
'boy'	0.667	0.619	0.652	0.661
mean	0.733	0.709	0.662	0.718
std	0.127	0.141	0.249	0.144

The PRI is then defined as follows:

$$PRI(S, \{S_k\}) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} [\mathbb{I}(l_i = l_j)\, P(l_i = l_j) +$$

$$+ \mathbb{I}(l_i \neq l_j)\, P(l_i \neq l_j)] \tag{3.84}$$

The PRI reflects how many pair pixels have the same interrelation (same or different labels). This value is also weighted by the expected values given by Equations 3.82 and 3.83. The PRI measure takes values in the interval $[0, 1]$, where 0 means that tested and manual segmentations have no similarities and 1 means that all segmentations are identical.

The accuracy of presented method was tested on 50 images taken from the well-known Berkeley Segmentation Dataset and Benchmark described in [Martin et al., 2001]. As the tested segmentation methods we use the GC and RW that were initialized with the SPB algorithm. These algorithms were named GC-SPB and RW-SPB. Furthermore, another two methods were included in the final comparison, the Felzenszwalb's algorithm ([Felzenszwalb and Huttenlocher, 2004]) (FHC) and the Normalized cuts (NC) ([Shi and Malik, 2000]). The set of ground truth segmentations $\{S_k\}$ was formed by five manual segmentations from above mentioned benchmark. The resulting PRI on several concrete images as well as the mean and std values are listed in Table 3.4. These results show that the use of SPB initialization provides meaningful segmentations that is more than comparable with popular autonomous methods. The resulting image segmentation of compared algorithms is shown in Figures 3.41, 3.42, 3.43.

3.7.4   Conclusion

To compare the SPB algorithm we use it for the initialization of the Graph cut (GC-SPB) and the Random walker (RW-SPB). These methods were then compared with two fully autonomous algorithms, the Felzenszwalb's algorithm (FHC) and the Normalized cuts (NC). The experiments show that the GC-SPB method gives the best results among compared algorithms followed by NC, RW-SPB and FHC.
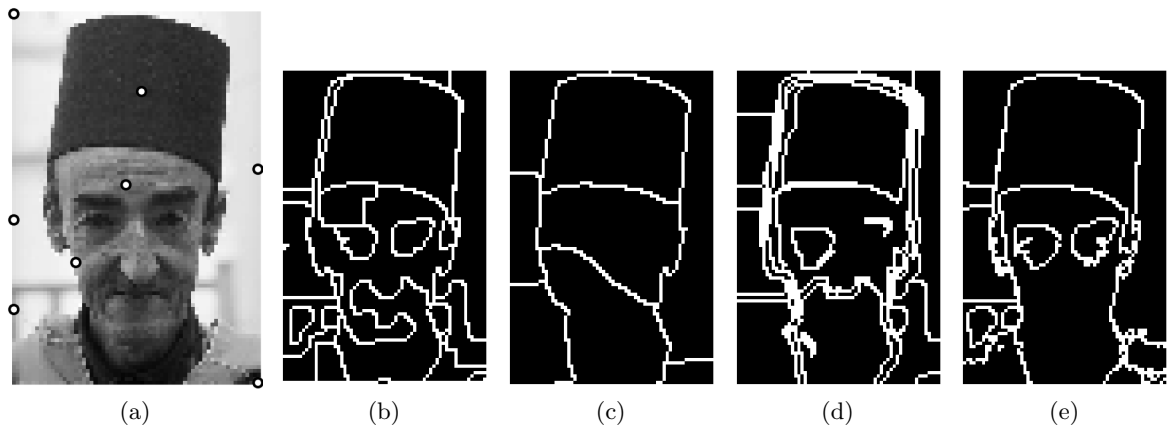
Figure 3.41: Segmentation of an image called 'man'. Fig. (a) shows the input image with derived seeds. Resulting PRI are: graph cut PRI = 0.832 (b), random walker PRI = 0.849 (c), FH-cut PRI = 0.831 (d) and N-cut PRI = 0.867 (e).
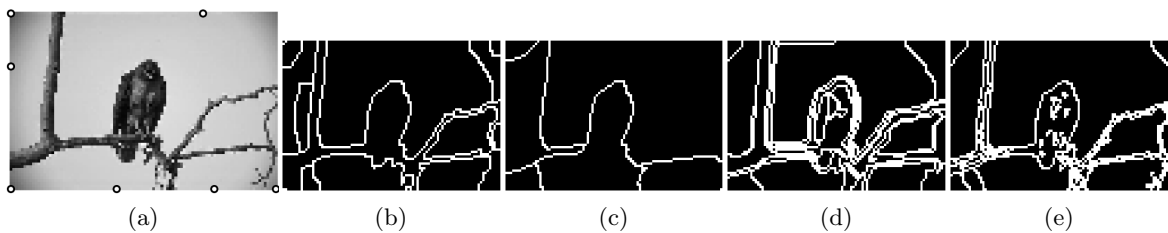
Figure 3.42: Segmentation of an image called 'bird'. Fig. (a) shows the input image with derived seeds. Resulting PRI are: graph cut PRI = 0.834 (b), random walker PRI = 0.835 (c), FH-cut PRI = 0.882 (d) and N-cut PRI = 0.592 (e).
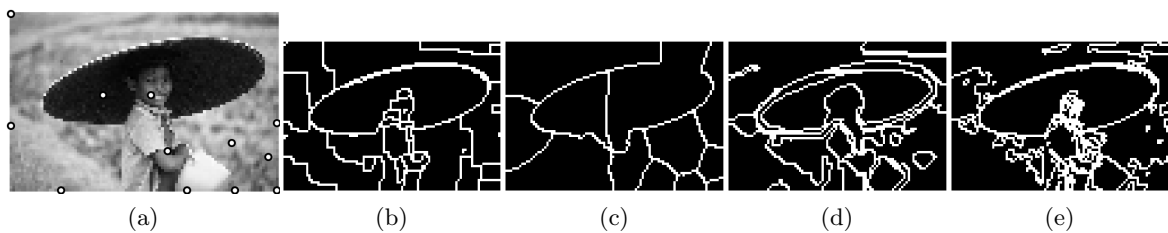
Figure 3.43: Segmentation of an image called 'boy'. Fig. (a) shows the input image with derived seeds. Resulting PRI are: graph cut PRI = 0.667 (b), random walker PRI = 0.619 (c), FH-cut PRI = 0.652 (d) and N-cut PRI = 0.661 (e)

A considerable disadvantage of the algorithm is time complexity. While the NC and the FHC take less than a second, the SPB needed tens of seconds. To overcome this problem, it is possible to reduce the spatial resolution using superpixels. On the other hand, using such an approach could degrade the image gradient, which is of great importance in the described approach.

Chapter 4

Automatic Liver Segmentation

Liver surgery is a field in which computer-based medical image processing can offer useful procedures with a significant impact on treatment. Because of the enormous individual anatomical variability of liver, a planning procedure in liver surgery is very important, [Meinzer et al., 2002]. This planning is based on pre-operative diagnostics. For a surgeon, it is difficult to mentally construct the three-dimensional model of an organ and important parts like its vessel system, [Selle et al., 2002]. Handmade volume segmentation is a relatively slow process and the large amount of prior knowledge available is underused, [Florin et al., 2007a]. However, computer vision algorithms can overcome this problem. Liver and its important parts can be visualized and may be used for planning procedure. Extracting liver from 3D imagery is usually a first step for planning liver surgery and diagnosis of liver disease.

Due to the nature of data and its acquisition there is little apriori information that could be used. Therefore, almost all of the initialization is done by data analysis with just a few assumptions. One of the assumptions is that the liver is one of the biggest organs in the abdomen. Another assumption is that most of the liver parenchyma is healthy. Therefore, it is reasonable to assume that the liver will form one of the biggest homogeneous blobs in the data. Of course, the whole parenchyma is not homogeneous - image points corresponding to tumors and vessel system differ in intensity. Therefore, it is not possible to easily extract the liver by simple thresholding the data, for example, and a more sophisticated method needs to be used.

In [Song et al., 2013] an adaptive fast marching method is proposed. This approach is based on fast implementation of active contour models that are initialized by and autonomous seed point placement. For more accurate initialization a nonlinear grayscale filter is used to enhance the contrast of the liver parenchyma. Another fully autonomous approach is presented in [Maier et al., 2008]. The authors propose an algorithm based on the random walker initialized by considering the ribcage position and voxel intensities. The work [Peng et al., 2015] uses graph cuts for optimizing an energy function derived from multiple region appearances. The authors claim that their algorithm is able to deal with liver parts with distinct appearances, e.g. liver tumors. Probably the most known use of the graph cut method in medical imaging is presented in [Boykov and Jolly, 2000]. This work describes use of the graph cut for interactive segmentation of medical images. Another class of methods can

be described as model based, e.g. [Florin et al., 2007b]. The authors incorporated a statistical analysis that helps them to deal with large shape variability of the liver parenchyma. Another interesting work is presented in [Dou et al., 2016]. Here, a convolutional neural network is used and the results are further refined by a conditional random field model.

In this chapter, we focus on an algorithm for automatic liver segmentation from CT images based on level sets and grow cut. For automatic seed-points selection, a density (image intensity) analysis is used. Coarse segmentation is achieved by the grow cut algorithm and is further refined by the localized active contours. The following section compares two approaches for seed selection: histogram analysis and GLCM analysis. The algorithm with initialization by histogram analysis was proposed in [Ryba et al., 2013]. The proposed algorithm is then compared with the shortest path basins algorithm described in Section 3.7.

4.1   Grow cut automation

Coarse segmentation of liver is achieved by using the grow cut algorithm described in Section 3.5. The algorithm needs to be initialized by defining a set of seed points. This initialization can be fulfilled by an expert (interactive version of the algorithm) or by an automatic procedure. Because the goal of the thesis is to design a fully automatic approach for tumor localization, two automatic initializations are presented. The first is based on histogram analysis, the second derives seed points from the GLCM.

4.1.1   Histogram analysis

The histogram is usually the first method used for data exploration. In a histogram it is possible to easily detect peaks in intensity distribution. From the assumption described above, we assume that one of the peaks will correspond to liver parenchyma. To be able to use grow cut algorithm, each significant peak is considered as a centroid of a class. Intensities near to these peaks/centroids then defines initialization of corresponding class.

For better peak localization, two operations are used in advance. The first one comes from the idea described in [Sonka et al., 2007]. The idea is based on suppressing image points with high image gradient when forming the histogram. That is because these pixels are often on the boundary between two or more objects and their intensities are often located in histogram valley between peaks of these objects. By suppressing these points, the peak-to-valley ratio will increase a bit as shown in Fig. 4.1. The second approach is histogram smoothing. It is achieved by convolving the original histogram by the Hanning window. The peaks are then localized as local maxima in a small region.

The final step is to define the peak's surroundings to be able to define a class. This is done by considering points with intensities near the corresponding peak. Derived seeds could be further processed by the median filter to remove isolated points. The whole process, as well as the derived seed points, is illustrated in Fig. 4.2.
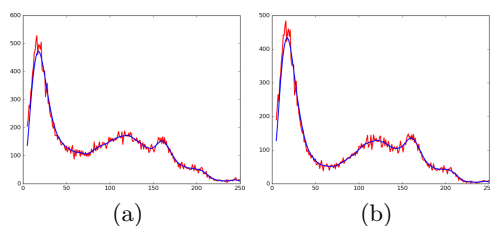
Figure 4.1: By increasing the peak-to-valley ratio in the original histogram (a), better peak isolation could be achieved (b). The original histogram is drawn red and its smoothed version is drawn blue.
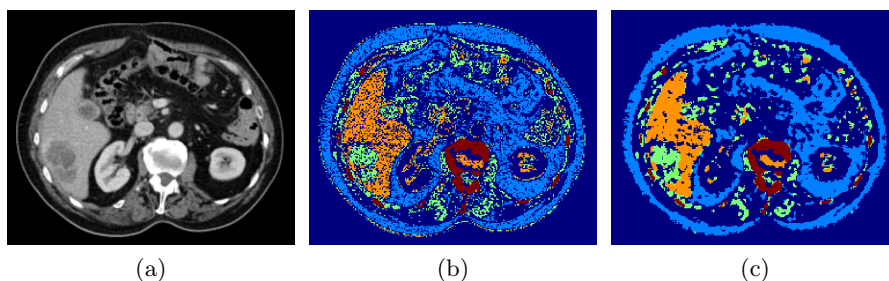
Figure 4.2: The seeds points (b) are derived from the input image (a) using the histogram analysis. The seed points are further filtered by the median filter of size 3x3 (c).

4.1.2   GLCM analysis

The disadvantage of using the histogram is that the obtained information ignores any spatial relation. Therefore, an approach based on the GLCM is presented. The GLCM contains identifiable modes corresponding to the homogenous regions (when located near the diagonal). By localizing these modes it is possible to use them as individual classes. One could try to fit a Gaussian mixture to the GLCM. Unfortunately, this approach does not produce precise results because of the noisy nature of the GLCM. Figure 4.3 shows how the algorithm creates classes that cut the modes instead of separating them, which yields to a splitting of the liver parenchyma.

Better results could be achieved using the mean shift algorithm described in [Comaniciu and Meer, 2002]. This algorithm looks for blobs in density of samples which makes it more suitable for our task. The final classes separate individual modes in the GLCM more precisely then the GMM estimation as shown in Fig. 4.3. The seeds derived by this approach are shown in Fig. 4.4.

4.1.3   Coarse segmentation

Both of the initialization approaches yield satisfactory results. Generally, slightly better results are achieved using the mean shift initialization with cell smoothing during the grow cut calculation. The segmentation is further processed and objects connected to the image
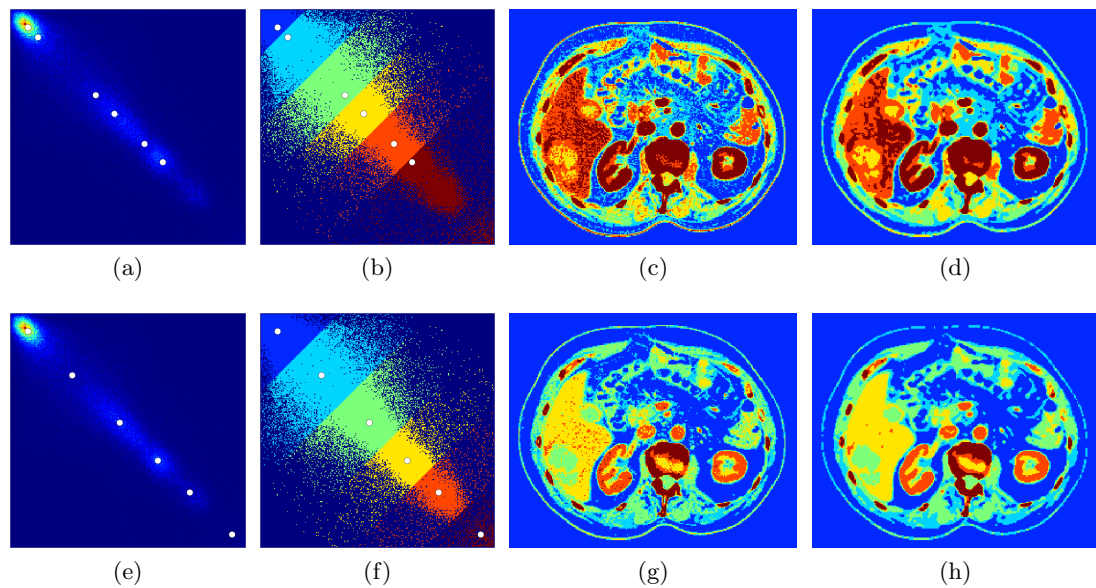
Figure 4.3: Comparison between the GLCM analysis by GMM (top row) and by mean shift (bottom row). Found classes (second column) are used to derive seed points (third column) that are further filtered by median filter of size 3x3 (fourth column).

borders are removed. The final coarse segmentation with both types of initialization and with or without the cell smoothing is shown in Fig. 4.5.
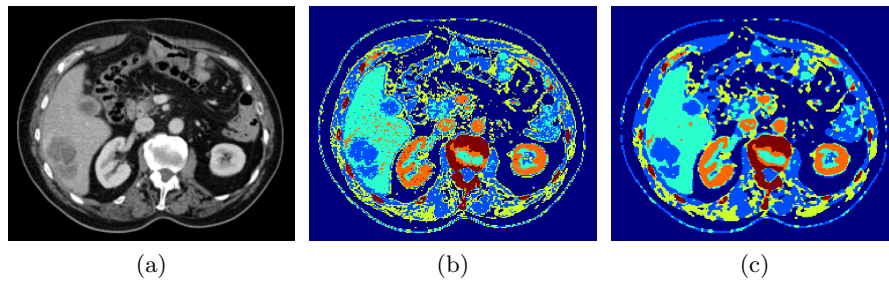
Figure 4.4: The seed points (b) are derived from the input image (a) using the GLCM analysis. The seed points are further filtered by the median filter of size 3x3 (c).
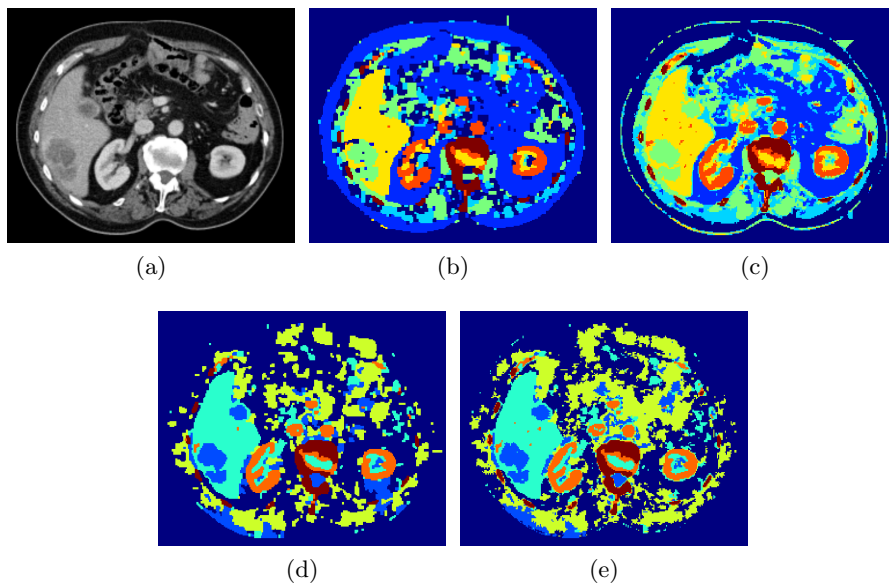
Figure 4.5: The input image (a) and final coarse segmentation using the grow cut algorithm. The segmentations (b) and (c) are from the histogram initialization with and without cell smoothing, respectively. Similarly, the segmentations (d) and (e) are from the GLCM initialization with and without cell smoothing, respectively.

4.2 Segmentation refinement using active contours

The segmentation obtained by the grow cut algorithm is often very coarse and inaccurate. The reason for that is the algorithm's natural disposition of creating rectangular boundaries. The true boundary of the liver parenchyma often has a slightly lower intensity than the rest due to the way the images are derived. In that case, the border is assigned to a different class and the overall segmentation accuracy decreases. It is, therefore, appropriate to use a segmentation method that will start at the coarse segmentation and will refine the result to better match the reality. For this purpose the active contours described in Section 3.4.3 are used.

Methods based on active contours need to be initialized by an initial curve that is going to be updated. This initial curve should be as close to the desired result as possible, which minimizes possibility of getting stuck in local energy minimum. In this work this initial curve corresponds to boundary of the coarse segmentation achieved by grow cut. An obvious problem that needs to be solved is choosing a blob that corresponds to the liver parenchyma.

4.2.1 Initialization

When determining the proper object for the initialization, a-priori information can be used. Four information items are considered when testing an object - its area, intensity, compactness and position. We know that density of the liver corresponds to he 40-60 HU in the native (non-contrast) CT. Unfortunately, the liver density differs in fatty or cirrhotic liver. For example, the density of fatty liver is lower than 40 HU ([Hamer et al., 2006]). To overcome the problem of comparing the intensity with a specific a-priori known value, several methods can be used. For example, one can localize the peak in histogram that lies in a specific yet wide range of intensities. Another approach that can be used is to identify the liver blob in the GLCM matrix. We used the GLCM approach in this work, because it is more robust and we already defined the GLCM in the previous step. Both methods are compared in Fig. 4.6. We set the range for peak localization to $[20 - 80]$HU. The histogram peak is located at 39 HU and the GLCM peak is located at 55 HU. In this case, the histogram peak could incorrectly diagnose the liver as fatty.

A problem that can arise when comparing the intensity is the similarity between the liver and the spleen. Fortunately, the spleen is significantly smaller then the liver and its compactness is lower as well. The liver is the largest organ in the abdomen, which makes the area the second most discriminative feature when filtering false objects. The compactness helps to filter out some connected structures, e.g. the closest surroundings of the liver. When looking at the position, the liver is located in the right part of the body, i.e. left part in CT images. Therefore we calculate the distance of each blob's centroid to the point $p = (\text{width}/4, \text{height}/2)$, where width and height correspond to the shape of the input image. By combining all four features, the object corresponding to the liver parenchyma is safely identified as shown in Fig. 4.7.

| (a) | (b) | (c) | (d) |
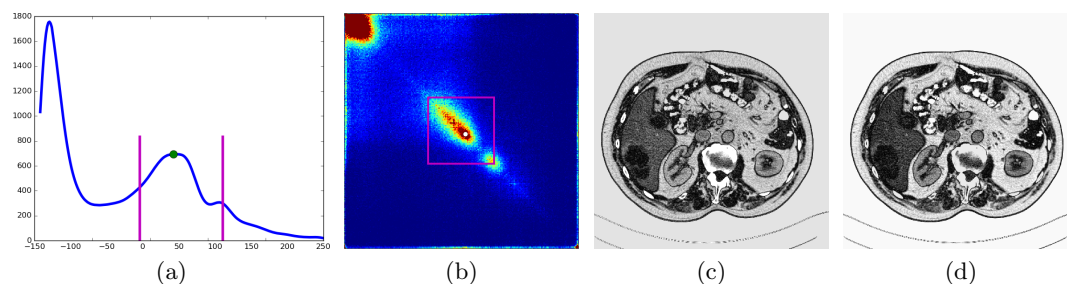
Figure 4.6: Liver blob identification. In the histogram, the peak is located at 39 HU (green circle) whereas the position of the peak in the GLCM corresponds to the 55 HU (white circle). The searching intervals are drawn in magenta. Image (c) shows absolute value of difference between the histogram peak and the image. The same goes for the image (d) and the GLCM peak.

4.2.2   Final segmentation

After coarse segmentation and the liver blob extraction it is possible to initialize and run the active contours algorithm. Because the density of the liver could change in different parts of the parenchyma, we choose the localized version of the active contours method described in Section 3.4.3.

The initialization is first processed by morphological operation to fill small holes and to smooth the border of the initializing blob. The smoothing of the borders is again applied to the resulting segmentation. The individual parameters of the method were set to the value derived from a training set. Comparison between the coarse and the refined segmentations is shown in Fig. 4.8.

A problem that can arise is caused by peripheral tumors, i.e. tumors located on the boundary of the liver parenchyma. By the nature of the grow cut algorithm, such a tumor would not be assigned to the same class as the liver, therefore it would not be a part of the initialization mask. The active contours very often stop their evolution on the boundary of the tumor assuming it is an outer object. This yields a false negative segmentation and the tumor wouldn't be localized, of course. Some of these cases can be solved by increasing the inner force of the level set algorithm. On the other hand, this sometimes also yields the propagation of the curve outside the liver parenchyma lowering the overall precision of the segmentation.

This problem can be partly solved in the intialization step by further postprocessing of the identified liver blob. The proposed solution lies in the analysis of objects neighboring on the liver blob. These neighbors are filtered by area and compactness. For each object in the filtered image, the compactness of the liver blob with the object is calculated. If the compactness (as defined in [Sonka et al., 2007]) decreases, then we assume that the object is a peripheral tumor, see Fig. 4.9. The segmentation refinement is then applied as usual due to the ability of localized active contours used to segment heterogeneous objects, see Fig. 3.18.
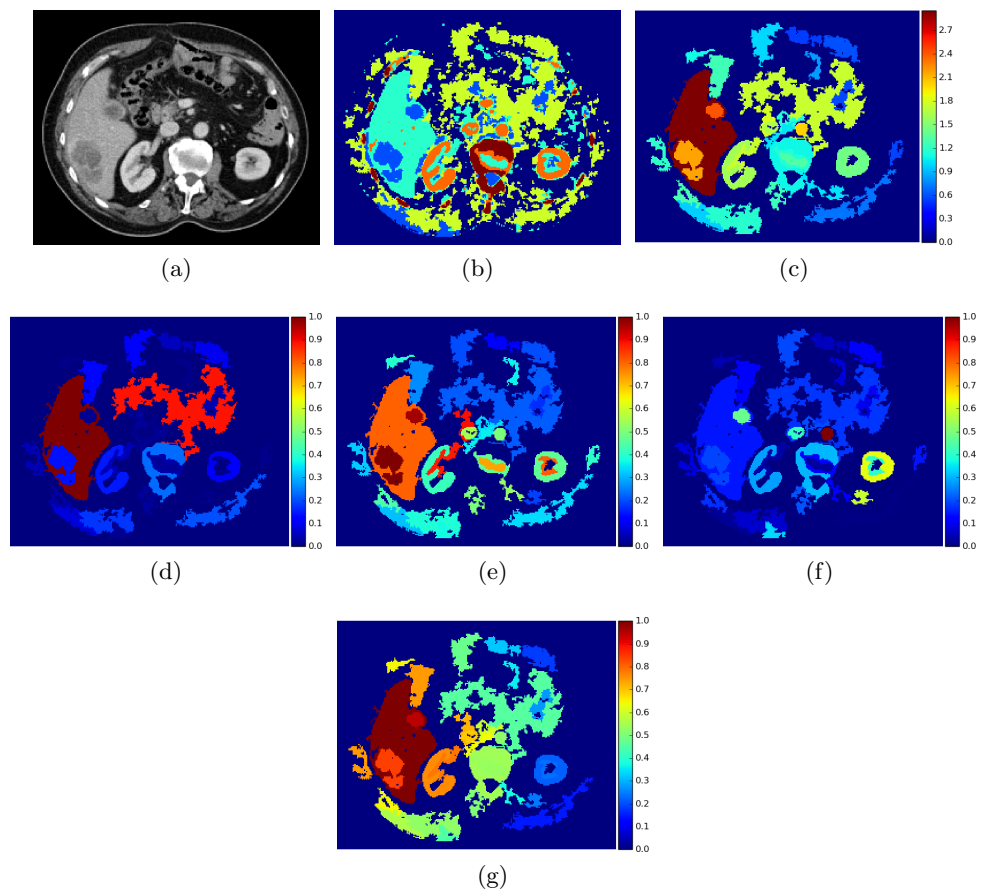
Figure 4.7: The input image (a) and its segmentation (b). The scores of area (d), intensity (e), compactness (f) and position (g) are combined into the final score (c).

(a)   (b)   (c)

(d)   (e)   (f)
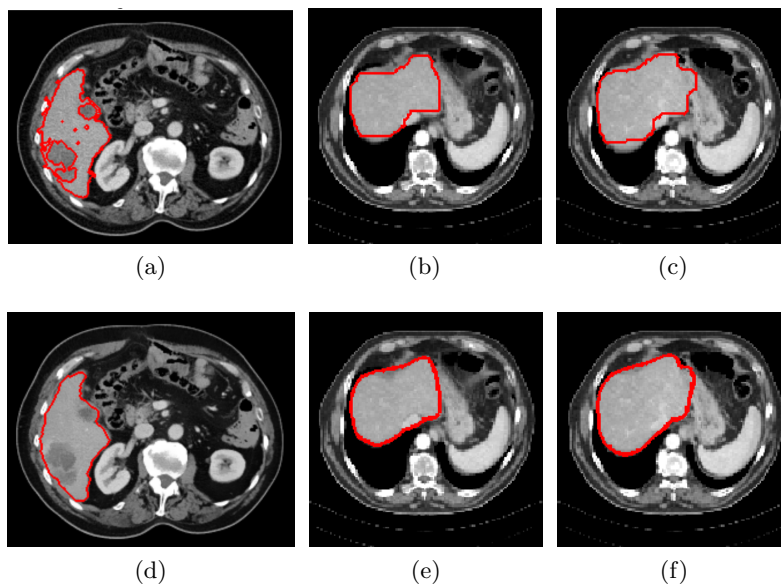
Figure 4.8: Comparison between coarse (a)-(c) and refined segmentations (d)-(f).

(a)   (b)

(c)   (d)   (e)
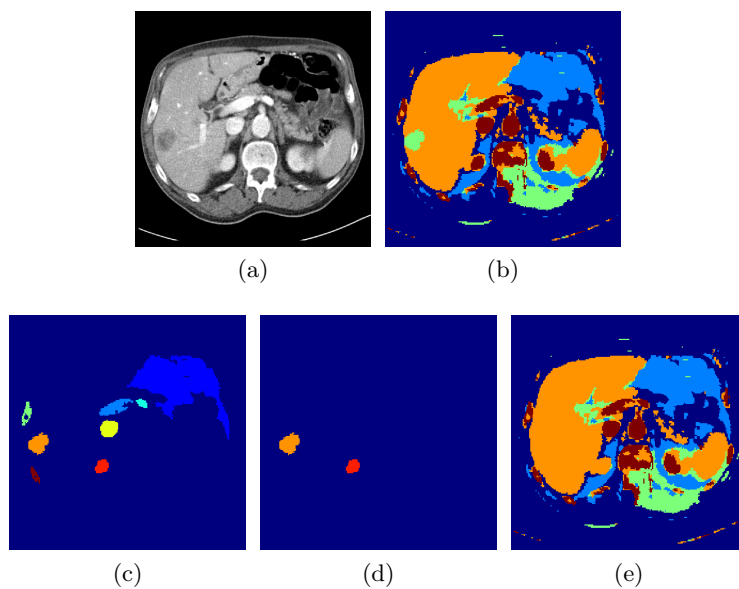
Figure 4.9: The input image (a) with a peripheral tumor is segmented by the grow cut in (b). All neighboring objects are then filtered by their area and compactness (c). After the analysis only two objects are considered as the peripheral tumor - a true tumor (orange) and a part of the right kidney (red) which yields the updated segmentation (d).

4.3   Segmentation by autonomous seed placement

The first algorithm that was tested for the liver segmentation was the shortest path basins algorithm (SPB) described in Section 3.7. The algorithm derives seed points in an iterative manner and places them on points with low energy. The energy is updated each time a new seed point is added to reflect the change of the segmentation. The final region from the labeled image is chosen following the same scoring algorithm as described in the previous section.

As shown in Fig. 4.10(c), the algorithm leaked out on the upper part of the liver. At the same time, it missed the lower part of the tumor. However, the biggest problem is that the algorithm cut off a hyperdense tumor located on the upper right part of the liver parenchyma. Although a region merging procedure is defined inside the SPB, two objects that significantly differ in intensity will never be merged as in the case presented.

Another segmentation achieved by the SPB is shown in Fig. 4.11. In this example, the algorithm labeled the lower half of the liver as the background. This time, the merging procedure failed to connect two parts of healthy parenchyma. This is caused by the intensity comparison between the two objects done by comparing the intensity of the seed points. The seed point of the lower part is located on a hyperdense region; therefore, the whole part is considered as hyperdense.

(a)   (b)

(c)
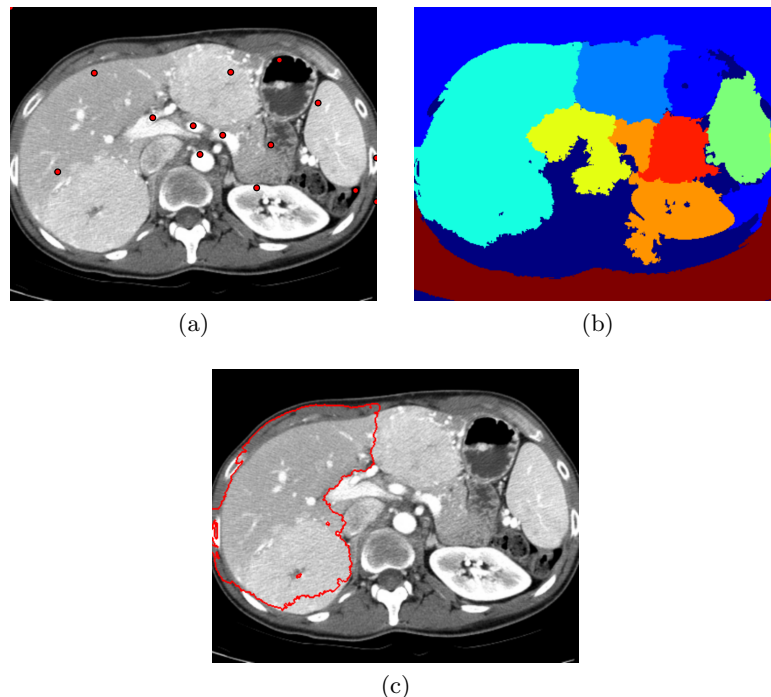
Figure 4.10: The SPB places seed points as shown in (a). This yields the segmentation (b). The liver blob (c) is chosen according to a scoring algorithm.
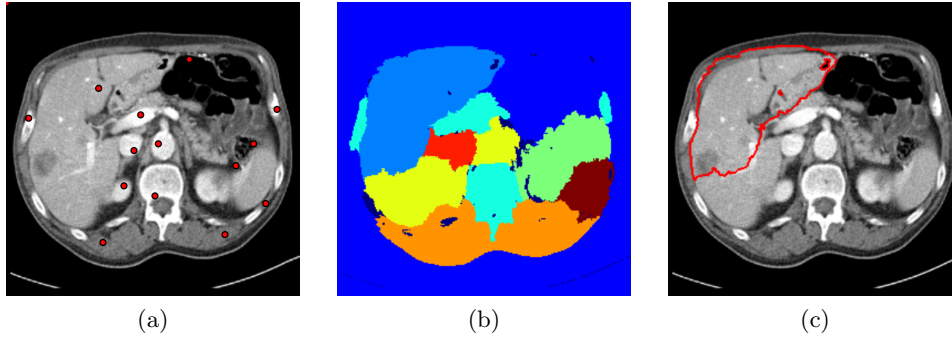
Figure 4.11: The SPB places seed points as shown in (a). This yields the segmentation (b). The liver blob (c) is chosen according to a scoring algorithm.

4.4   Accuracy of liver segmentation

The accuracy of the algorithm was evaluated over 164 images in the sense of precision, recall and F-measure. The precision p describes how many selected points are relevant; the recall r says how many relevant points are selected. The F-measure F then combines both of them by a harmonic mean:

$$p = \frac{\text{relevant points} \cap \text{selected points}}{\text{selected points}} = \frac{TP}{TP + FP}, \tag{4.1}$$

$$r = \frac{\text{relevant points} \cap \text{selected points}}{\text{relevant points}} = \frac{TP}{TP + FN}, \tag{4.2}$$

$$F = 2 \cdot \frac{p \cdot r}{p + r}, \tag{4.3}$$

where TP denotes true positive, FP false positive and FN false negative.

The results are shown in Tab. 4.1 and visualized in the form of box-plots in Fig. 4.12. The lowest accuracy was achieved using the SPB - the mean F-measure of value 0.708. Significantly more accurate results are obtained using the grow cut algorithm and are further improved by the segmentation refinement - the mean F-measure increases from 0.82 to 0.89. This shows the use of the proposed approach appropriate.

	SPB			Grow Cut			Active Contours		
	F-meas.	prec.	rec.	F-meas.	prec.	rec.	F-meas.	prec.	rec.
mean	0.708	0.688	0.754	0.82	0.79	0.87	0.89	0.86	0.95
median	0.732	0.704	0.777	0.85	0.82	0.1	0.92	0.88	0.99
std	0.113	0.107	0.122	0.13	0.12	0.14	0.14	0.13	0.15

Table 4.1: Accuracy of liver segmentation.

(a)   (b)   (c)
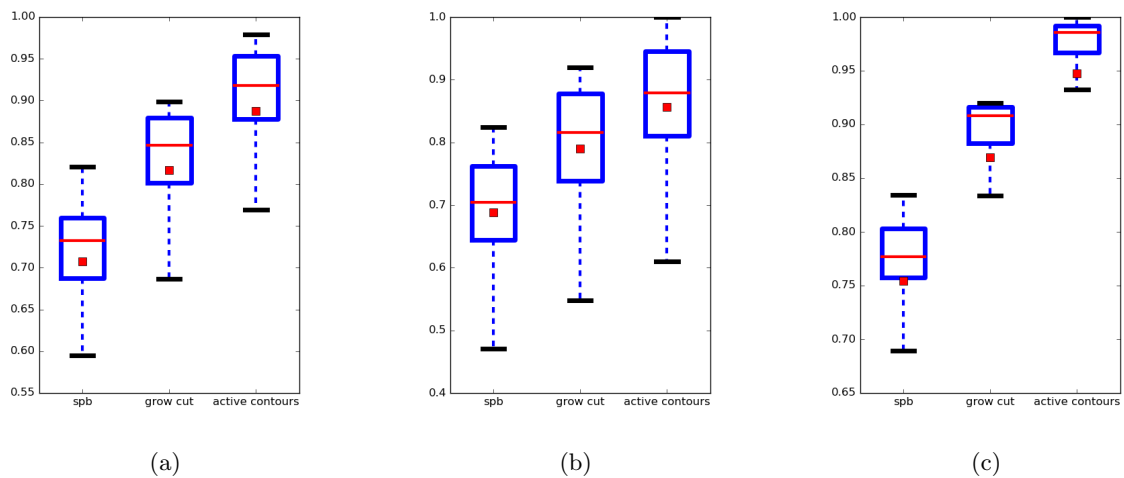
Figure 4.12: The accuracy of liver segmentation is evaluated over 164 images by the F-measure (a), the precision (b) and the recall (c). Values between the first (Q1) and the third quartile (Q3) are shown in blue boxes; red lines represent median and red squares represents mean values. The whiskers bounds the data considered as non-outliers. The minimal value is Q1 - 1.5(Q3 - Q1), the maximal value is Q3 + 1.5(Q3 - Q1).

Chapter 5

Liver Registration

5.1 Introduction

During the CT scanning, a patient lies at a CT table while the system is building the final image series. As stated at the beginning of the work, the delay between the start of the venous phase and the end of the arterial phase is about 1 minute. Ideally, the patient should not move at this time, which is almost impossible to guarantee. Therefore, the acquired data differs in its spatial domain and need to be aligned.

The goal of image registration is to find a geometric transformation that optimally maps one image onto a second image. The optimality of the process depends mainly on the estimated transformation type and the similarity measure used. Using different similarity measures often yields significantly different results; therefore, the proper type of similarity measure is essential.

The transformation type as well as the similarity measure can be chosen in advance, depending on the input data. To do that properly, one needs to know the nature of the input data and the way of their acquisition. The following text describes classification of registration methods followed by the reasoning about the parameters suitable for our case.

5.2 Classification of registration methods

There are numerous registration methods designed for general to very specific tasks. All of these methods can be classified from several aspects. In the paper [Maintz and Viergever, 1998] a classification according to 9 criteria is presented. This classification is reduced to 8 criteria in work [Fitzpatrick and Sonka, 2000]. We will use the first approach and describe all 9 criteria: dimensionality, nature of registration basis, nature of transformation, domain of transformation, interaction, optimization procedure, involved modalities, subject and object.

5.2.1 Dimensionality

This criterion categorizes the methods w.r.t. the number of dimensions of input data. Not all of the dimensions need to be spatial - a dimension could represent time as well. Moreover,

methods for registering data with different dimensionality are also described in this category. Probably the most frequent case is 2D-2D registration, e.g. individual slices of CT. To register the whole CT scan, a 3D-3D registration needs to be used. An example of 2D-3D registration is the mapping of an intra-operative X-ray image to a pre-operative CT image.

Methods where one of the dimension represents time can be further classified w.r.t. the time interval between individual data. Short intervals are most frequent in the case of post-operative monitoring, medium-length intervals are used when monitoring tumor growth and long intervals are typical for monitoring growth of bones.

5.2.2 Nature of registration basis

The registration basis defines the type of features that are used when estimating transformation parameters. Following this criterion, methods are divided into extrinsic and intrinsic. The former are based on foreign objects synthetically added into the imaged space, whereas the latter use image features derived from the anatomical structures only. The foreign objects can be any kind of artificial object attached to the patient. The image features are represented by significant points, edges and/or homogeneous areas. For more information see the book [Fitzpatrick and Sonka, 2000].

5.2.3 Nature of the transformation

As already stated, the proper use of transformation type has a great impact on the whole process of image registration. Theoretically, to use a general transformation describing most of the possible deformation may look like a good idea. However, such a transformation would certainly have many parameters whose estimation would be very time-consuming. A reasonable approach is to define all transformations that can occur in the current task and choose a transformation type based on this analysis.

The most used transformations are:

- rigid body - preserves all distances,

- affine - maps parallel lines onto parallel lines,

- elastic/curved - maps lines onto curves.

5.2.4 Domain of the transformation

Following this aspect, the methods are classified into global and local; see Fig. 5.1. We call a transformation global if it is applied to an entire image. Local methods are applied only on a part of an image. Theoretically, it is possible to define different local transformations for different image parts.

5.2.5 Interaction

Three levels of interaction can be recognized in registration methods - automatic, semi-automatic and interactive. The automatic methods does not require any interaction at all, the
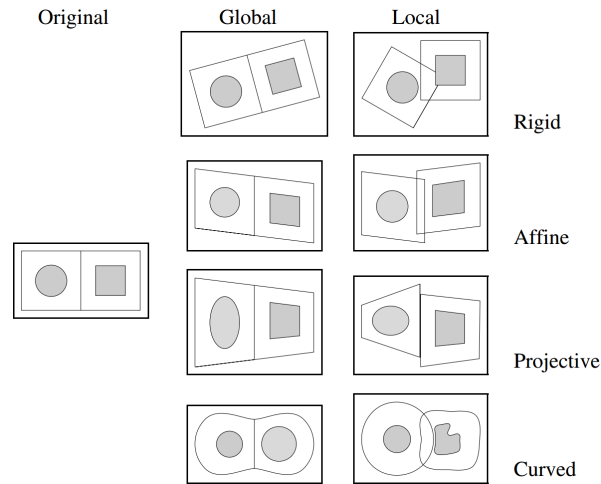
Figure 5.1: Examples of transformation types as shown in [Maintz and Viergever, 1998].

user only supplies input data. Some interactivity is needed in the semi-automatic methods. This interactivity could be represented by first segmenting the input data or by accepting or rejecting suggested results. The fully interactive methods only assists the user who registers the data himself.

5.2.6   Optimization technique

The optimization is the core of each registration method. The transformation parameters are iteratively estimated based on searching for an extreme in a cost function. The cost function describes the similarity between the fixed and the moving images. The moving image is then transformed according to the estimated transformation.

An important step of the optimization procedure is defining image similarity, i.e. defining a similarity measure. The most frequently used methods are described in Section 5.3.

5.2.7   Modalities involved

A modality can be interpreted as a way of acquiring the data, eg. X-ray, CT, MRI, PET, SPECT etc. The registration can be monomodal or multimodal. The former occurs if both types of data come from the same modality. The latter describes cases where the fixed image is of a different modality than the moving image. Besides these common tasks another less frequent exists. For example, modality-model describes registration of an anatomical part to a mathematical model of its structure.

5.2.8   Subject

According to this aspect, it is possible to classify the methods to intrasubject, intersubject and atlas. If both of the input data came from the same patient, we call it intrasubject. On the other hand, if each data came from a different patient, the registration is called

intersubject. The third class describes cases when one of the input images originates in an information database obtained using the imaging of many subjects. This database is called an atlas.

5.2.9   Object

In this case, the object means a part of the human body whose registration is our goal. Therefore, we distinguish the registration of the brain, liver, kidney, vertebra etc. For each object a different similarity measure as well as a different transformation type can be defined.

5.3   Similarity measures

After defining a transformation type, the next step is to determine the proper similarity measure. Again, using different measures can yield very different results; therefore, to choose a proper measure is essential for a precise registration. The most frequent similarity measures are briefly described in the following text.

5.3.1   Sum of squared differences

The sum of squared differences measure (SSD) is used in monomodal registration or if the data differs only in Gaussian noise. In that case, it is proven that the SSD is an optimal measure to use, see [Viola and Wells, 1997]. If both images are perfectly aligned, the value of the SSD is near zero. With increasing disalignement, the measure increases. The SSD between two images is defined as follows:

$$\mathbf{SSD} = \frac{1}{N} \sum_i^N |\boldsymbol{A}(i) - \boldsymbol{B'}(i)|^2, \forall i \in \boldsymbol{A} \cap \boldsymbol{B'}, \tag{5.1}$$

where $\boldsymbol{B'} = \boldsymbol{T}(\boldsymbol{B})$ is the transformed image and $\boldsymbol{T}(.)$ denotes the actual estimated transformation.

The assumption of the Gaussian noise is significantly restrictive and in the case of medical images often broken. For example, data from MRI exhibits the Rician distribution as stated in [Gudbjartsson and Patz, 1995]. This distribution significantly differs from the Gaussian distribution at low signal intensities. A registration based on the SSD measure is used, for example, in [Hajnal et al., 1995] and [Friston et al., 1996].

5.3.2   Correlation Coefficient

The correlation coefficient (CC) is very often used if the intensities of the input images are linearly dependent. Using the same notation as in the SSD and defining $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ as the

mean intensity of input images $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively, the CC is defined as follows:

$$\mathbf{CC} = \frac{\sum\limits_{i}^{N}(A(i) - \bar{A})(B'(i) - \bar{B'})}{\left(\sum\limits_{i}^{N}(A(i) - \bar{A})^2 \sum\limits_{i}^{N}(B'(i) - \bar{B'})^2\right)^{1/2}}, \forall i \in \boldsymbol{A} \cap \boldsymbol{B'}. \tag{5.2}$$

5.3.3   Mutual information

Using SSD or CC when registering multimodal images usually yields a very poor results. This is due to direct comparison of intensities that can significantly differ even for the same anatomical part, especially in the case of multimodal registration. The mutual information (MU), on the other hand, compares mutual histogram and joint probability. It was independently proposed in [Wells et al., 1996] and [Maes et al., 1997] and became the most used measure for multimodal registration tasks.

The first step towards determining the MI is the calculation of a joint probability density function (PDF) for images $\boldsymbol{A}$ and $\boldsymbol{B'}$:

- For each point $\boldsymbol{i} \in \boldsymbol{A} \cap \boldsymbol{B'}$ determine intensity partitions $\boldsymbol{a}$ and $\boldsymbol{b}$ and increment histogram $\mathbf{HIST}[\boldsymbol{a}, \boldsymbol{b}]$.

- Normalize the histogram to calculate the PDF:

$$\mathbf{PDF}[j, k] = \frac{\mathbf{HIST}[j, k]}{\sum_{j,k} \mathbf{HIST}[j, k]}. \tag{5.3}$$

In the next step, the joint entropy is calculated:

$$H(A, B') = -\sum_{j,k} \mathbf{PDF}[j, k] \log \mathbf{PDF}[j, k]. \tag{5.4}$$

Using the following equations the marginal entropies $\boldsymbol{H(A)}$ and $\boldsymbol{H(B')}$ are defined:

$$H(A) = \sum_{j}\left(\sum_{k} \mathbf{PDF}[j, k] \log \sum_{l} \mathbf{PDF}[j, l]\right) \tag{5.5}$$

$$H(B') = \sum_{k}\left(\sum_{i} \mathbf{PDF}[i, k] \log \sum_{j} \mathbf{PDF}[j, k]\right) \tag{5.6}$$

Finally, the MI is given using the next equation:

$$\mathbf{MI}(A, B') = H(A) + H(B') - H(A, B') \tag{5.7}$$

The MI can produce inaccurate results if the data are more noisy. In that case it is better to use its normalized version (NMI) proposed in [Studholme et al., 1999]:

$$\mathbf{NMI}(A, B') = \frac{H(A) + H(B')}{H(A, B')} \tag{5.8}$$

5.4   Choosing the transformation type and similarity measure

In our case, both inputs came from the CT but are acquired at different phases. The fixed image corresponds to the portal venous phase, while the moving image is acquired at the arterial phase. Although both images originate in the same modality (CT), their density/intensity differs depending on the enhancing by the contrast agent. Therefore, we can look at the process as a multimodal registration.

When registering images of the brain, one could safely use a rigid body transformation due to the fact that the brain is inside a skull that does not normally deform itself. Unfortunately, the abdomen does not have any kind of thick capsule. Therefore each movement deforms the inner organs in a non-rigid manner.

For better performance, only data inside a bounding box of corresponding segmentation are considered. This reduces the number of points to be registered to approximately 25% of the original size. Although we are trying to register the data in an optimal way, the accuracy of the process does not need to be pushed to its limits at any cost. In fact, for the tumor classification only a partial overlay of corresponding tumors is required. This allows us to pair-up localized objects for further analysis.

To find the right transformation type and similarity measure, we examined the registration process when using different parameter values. We compared the similarity, the affine and the spline transformation and the SSD and the MI similarity measures over 100 images. As discussed above, our data differs in the density distribution, therefore using the SSD measure is not recommended. Because the abdomen exhibits elastic transformations, we assume that the best results will be achieved using the spline transformation. This assumption was confirmed as shown in Fig. 5.2. The similarity transform achieved the SSD = 2523 and the MI = 0.451, the affine transform ends up with SSD = 1491 and the MI = 0.581 and the spline transform achieves the SSD = 1038 and the MI = 0.779. An example of the registration process of two images is shown in Fig. 5.3 where you can see that the similarity measure together with the SSD measure totally failed to register the two input images.

Although the best accuracy was achieved using the spline transformation (when using the MI measure as discused above) two drawbacks are linked to this approach - duration and deformations. The optimization of the spline transform is very time-consuming - almost 10x the time of registration with the affine or the similarity measure. Moreover, the obtained images are sometimes unnaturally deformed. Therefore, it could be advantageous to use the affine transform to avoid such deformations and to speed up the process.
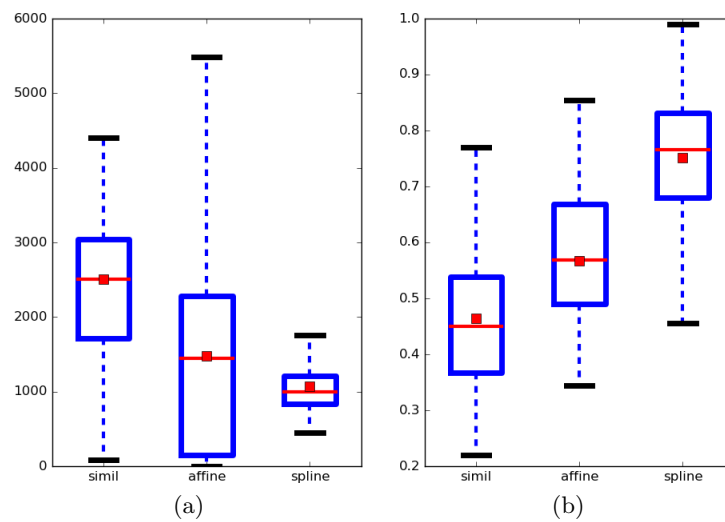
Figure 5.2: The accuracy of liver registration is evaluated over 100 images with the SSD (a) and the MI measure (b). Each measure was used with different transformation types. Values between the first (Q1) and the third quartile (Q3) are shown in blue boxes; red lines represent median and red squares represents mean values. The whiskers bound the data considered as non-outliers. The minimal value is Q1 - 1.5(Q3 - Q1); the maximal value is Q3 + 1.5(Q3 - Q1).
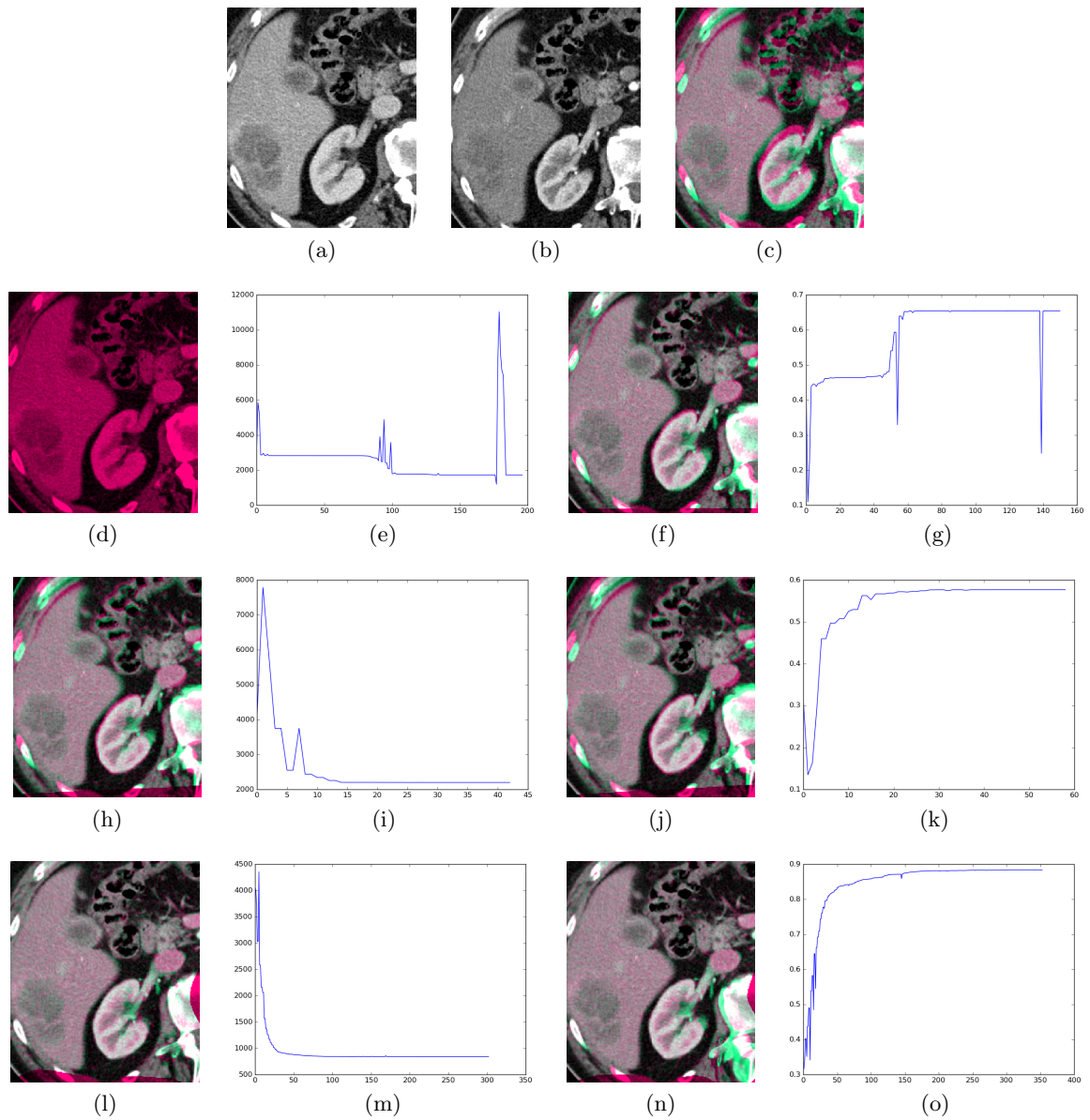
Figure 5.3: Registration of the fixed image (a) and the moving image (b) with their initial alignement (c). The other images visualize the result and the process of registration with the following set-up: similarity + SSD (d)-(e), similarity + MI (f)-(g), affine + SSD (h)-(i), affine + MI (j)-(k), spline + SSD (l)-(m) and spline + MI (n)-(o).

Chapter 6

Lesion Analysis

This chapter describes the process of localization and classification of liver tumors. The tumors are localized only inside a mask that defines the liver parenchyma. To be able to identify the individual tumors, we need to analyze their intensity distribution at both the arterial and the portal venous phase. For that, it is necessary to register the data so that the individual objects occupy the same image position. The liver segmentation and registration process were described in the previous chapters.

A serious problem in analysing medical images is lack of data. The medical data are protected by law and therefore they are rarely freely available. Fortunately, thanks to a cooperation with the Teaching Hospital Pilsen, we were able to get access to some data. Unfortunately, only the total number of 346 images with a tumor were available to us. Therefore, we were forced to use augmentation of these data to create more of them. We augmented the data with random geometrical transformation (flip, rotation, scale). Moreover, we added noise to change the intensity distribution as well. This way we were able to obtain the total number of 900 images - 100 images per a tumor.

A popular approach for liver tumor segmentation is the fuzzy C-means algorithm (FCM); for example see [Mala and Sadasivam, 2005], [Mala and Sadasivam, 2006] and [Kumar and Moni, 2010]. The FCM is able to localize tumors, but for our purposes it is not robust enough. The work [Choudhary et al., 2008] presents an adaptive multi-thresholding method based on an entropy measure. Another interesting group of methods are based on fitting a Gaussian. For example, the paper [Soler et al., 2001] describes an algorithm based on fitting the histogram by a Gaussian followed by topological and geometrical analysis. A similar histogram analysis is used in [Massoptier and Casciaro, 2008]. This time, the total number of k Gaussian functions is fitted and the results are used in the k-means classification method. The work [Zhou et al., 2008] is based on extracting a tumorous region from one slice by a trained SVM classifier. The extracted region is then projected to its neighboring slices and further processed.

6.1   Localization

The density of a significant number of liver tumors is not very contrasting from the density of surrounding liver tissue. Therefore, the use of conventional segmentation methods did not provide satisfactory results. Some examples of such liver tumors are shown in Fig. 6.1. Thresholding-based algorithms are not able to detect the right threshold value, region growing algorithms cannot deal with the noisy-nature of data etc. Even more advanced algorithms cannot deal with such problems. For example, the active contours are often leaking out from the tumorous area due to the low contrast at its border. Thus a new approach was designed that is based on the saliency maps and the Markov random fields.

The algorithm defines the MRF over the input image and the saliency maps described in Section 3.6 are used as the unary/data term. For efficiency, the MRF is defined in the sense of the Ising model (Sec. 3.3.5.1) and optimized by the graph cut with large moves (Sec. 3.3.6.3).

6.1.1   Individual saliency maps

For the MRF, the unary term for the object as well as for the background needs to be specified. At first, the object's unary term was defined as a saliency map. For the background's unary term the same but inverted saliency map was used. This defines a setup of the localization system. The total number of 7 setups was tested, each corresponding to one of the 7 saliency maps described in Sec. 3.6: the intensity difference (idiff), the histogram analysis (hist), the GLCM analysis (GLCM), the sliding window (sliwin), the texture (LBP), the shape and the blobs. The accuracy of each setup was calculated pixel-wise in the sense of the precison, the recall and the F-measure. The results are written in Table 6.1 and visualized with boxplots in Fig. 6.2.

The results show that the best mean F-measure was achieved using intensity difference (0.673). On the other hand, the worst accuracy was achieved using the shape saliency (0.614). Nevertheless, all results are comparable and usable. The question is: is it possible to increase accuracy by combining individual saliency maps? The answer to this question is discussed in the next section.

6.1.2   Combined saliency maps

Although all saliency maps lead to very similar localization accuracy, it is possible to determine the most and the least informative ones. The reason for doing this is to choose a combination of individual saliency maps that will increase localization accuracy. The value of the resulting saliency maps may correspond to the mean value of individual saliencies at each image point. However, much better accuracy can be achieved by using the median value instead.
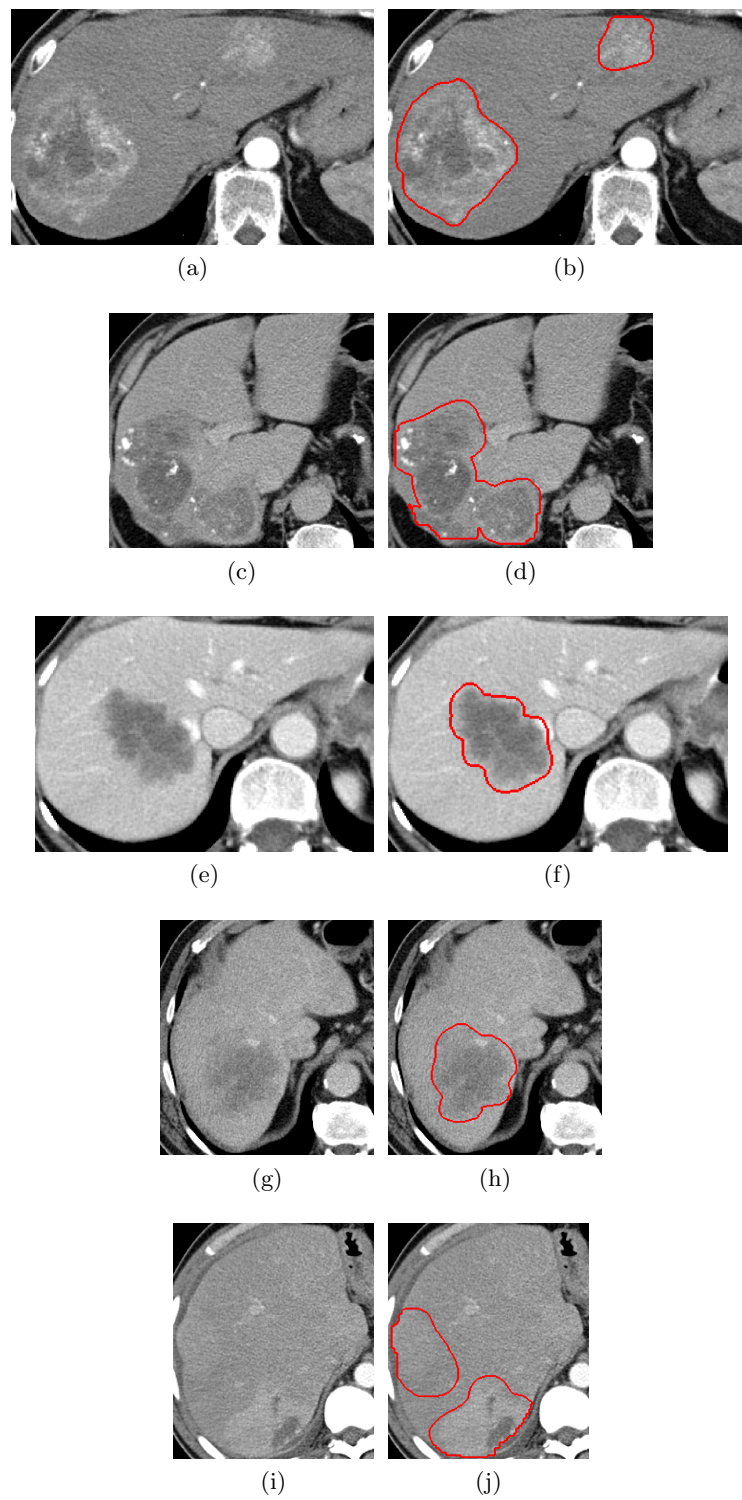
Figure 6.1: Example of different liver tumors with ground-truth segmentation.
.

	F-measure			precision			recall		
	mean	median	std	mean	median	std	mean	median	std
idiff	0.673	0.712	0.19	0.663	0.716	0.247	0.77	0.784	0.158
sliwin	0.67	0.736	0.209	0.654	0.737	0.267	0.787	0.814	0.16
blobs	0.658	0.743	0.219	0.643	0.733	0.272	0.785	0.82	0.173
GLCM	0.639	0.718	0.218	0.609	0.671	0.263	0.773	0.815	0.176
hist	0.639	0.713	0.226	0.633	0.724	0.283	0.761	0.8	0.189
LBP	0.615	0.673	0.212	0.562	0.601	0.246	0.793	0.826	0.167
shape	0.614	0.667	0.226	0.598	0.668	0.219	0.764	0.8	0.187

Table 6.1: Accuracy of tumor localization with individual saliency maps
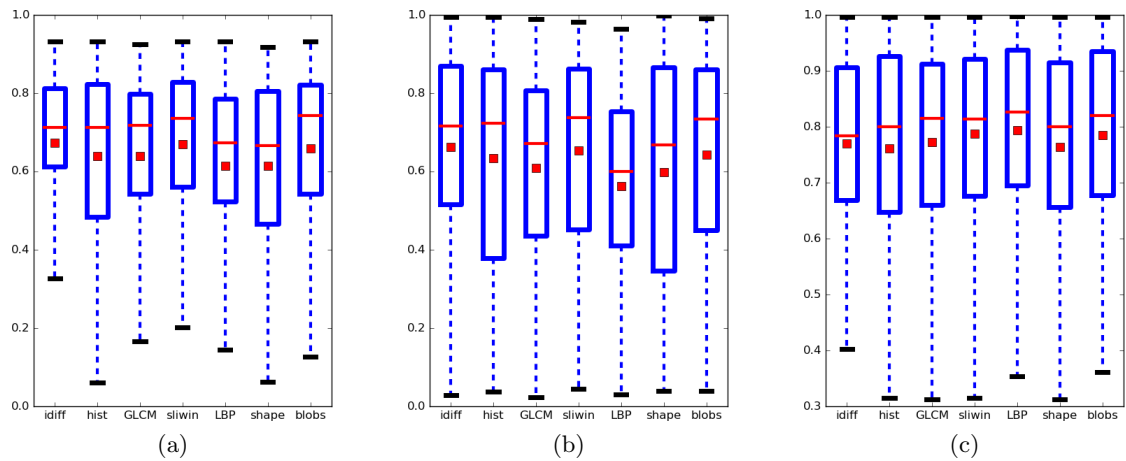
Figure 6.2: Accuracy of tumor localization with individual saliency maps: the F-measure (a), the precision (b) and the recall (c).

Over 20 combinations were tested but only the 6 most interesting combinations are described in this work:

- *all* ...all saliency maps combined,

- *no_LBP* ...all saliency maps except the LBP,

- *dwsb* ...intensity difference, sliding window, shape, blobs,

- *gwsb* ...GLCM, sliding window, shape, blobs,

- *ws* ...sliding window, shape,

- *dwb* ...intensity difference, sliding window, blobs.

The accuracy of all above-listed combinations is written in Table 6.2 and visualized in Fig. 6.3. It shows up that the best combination of saliency maps is *dwb*. Comparing mean values of the F-measure, the best saliency map achieved accuracy of 0.673, whereas the best combination achieved accuracy of 0.743. An interesting fact when comparing the accuracy is that the most combinations are less accurate than using all of the saliency maps. Only the combination of intensity difference, sliding window and blobs is better. Also, using the GLCM (*gwsb*) achieved better results than using intesity difference (*dwsb*). Therefore, one can argue why there is not the the GLCM, sliding window and blobs combination (*gwb*) in the comparison instead of the *dwb*. The reason is that this is not the final version of the algorithm. The next section describes a modification of the algorithm, where using intensity difference yields better results then the GLCM analysis.

6.1.3   Merging procedure

Sometimes, the tumor may be split into more regions. This may be caused by its heterogeneous character and/or a peripheral part with different density. An example of such a situation is shown in Fig. 6.4. To deal with this problem, a merging procedure is defined that is run if more regions are located next to each other. This process merges individual regions into a single one that ideally corresponds to the real tumor.

The approach is based on the compactness measure and the assumption that two different tumors are not located next to each other in a compact way. In other words, if we localize more regions next to each other and their overall compactness would be significantly low, then we will merge these regions into a single object. The compactness is defined as in [Sonka et al., 2007], i.e. the lower the value the more compact the object is.

6.1.4   Segmentation tuning

The graph cut method has two parameters: α, β. These were tuned over 200 images and the best results were obtained with the values $\alpha = 5$, $\beta = 1$. Although the active contours are not appropriate for the localization process, they can be used for segmentation refinement (in similar manner as in the liver segmentation problem). To avoid problems with leaking the curve out of the tumorous area, only a few iterations are allowed. This way, the mean

	F-measure			precision			recall		
	mean	median	std	mean	median	std	mean	median	std
all	0.718	0.763	0.164	0.826	0.945	0.247	0.662	0.667	0.165
np_LBP	0.712	0.764	0.169	0.82	0.951	0.259	0.657	0.668	0.165
dwsb	0.711	0.769	0.183	0.796	0.918	0.261	0.688	0.716	0.171
gwsb	0.716	0.782	0.189	0.801	0.924	0.266	0.696	0.725	0.161
ws	0.63	0.731	0.251	0.662	0.851	0.348	0.7	0.718	0.172
dwb	0.743	0.799	0.17	0.835	0.939	0.233	0.707	0.742	0.176

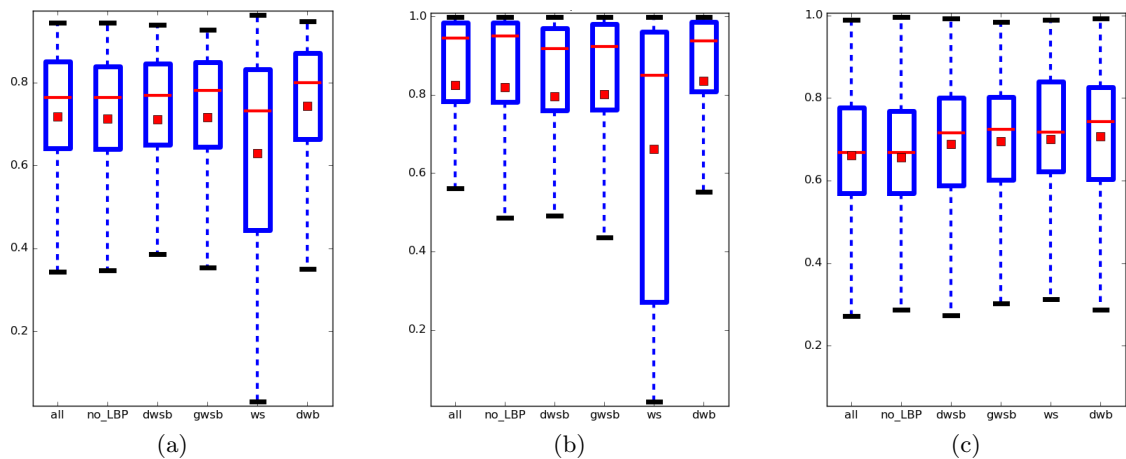Table 6.2: Accuracy of tumor localization with combination of saliency maps

Figure 6.3: Accuracy of tumor localization with combination of saliency maps: the F-measure (a), the precision (b) and the recall (c).
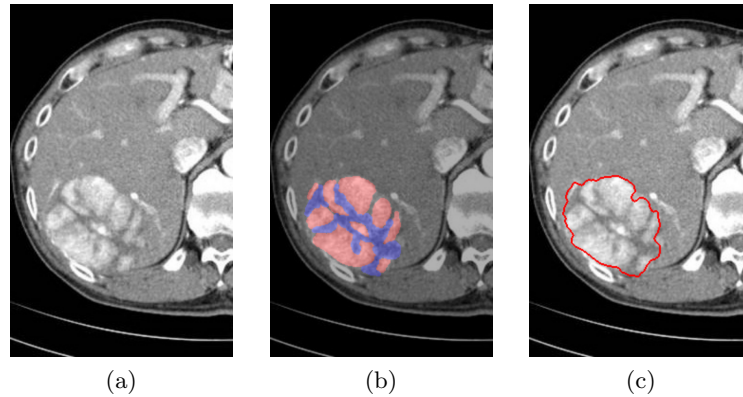
Figure 6.4: A tumor (a) is split into a several contrast regions (b). After the merging procedure and morphological postprocessing a single tumor is localized (c).

value of the F-measure is improved from 0.743 to 0.853. In contrast with the non refined localization, using the LLS pushes the accuracy of the intensity difference above the GLCM analysis as already mentioned in the previous section. The accuracy of localization refined with the LLS is listed in Table 6.3 and visualized in Fig. 6.5.

Localization results are visualized in Fig. 6.6. The basic MRF represents algorithm, where the unary terms are derived from the intensity histogram. This approach is described in greater detail in [Ryba et al., 2013]. Typical for this algorithm is its general under-segmentation. The cases where a tumor is not localized at all are not exceptional. On the other hand, intensity difference saliency map often over-segments the data as shown in Fig. 6.6(d). This feature could be sometimes helpful, because the expert-made ground truth data usually includes a small region around the tumorous tissue sometimes called safety capsule/region. This region does not differ from the healthy tissue but is still considered as a part of the tumor. Unfortunately, the native versions of the algorithms described in this work cannot deal with this problem. A possible solution is a morphological postprocessing (dilation, closing) that would expand the boundary of localized objects. More results of the lesion localization can be found in appendix A.

6.1.5 Object-wise accuracy

To this point, the accuracy was evaluated in the pixel-wise manner. For the purpose of a CAD system providing the doctor with lesion candidates, the object-wise accuracy is more informative. In this case, the recall represents how many lesions were localized and the precision describes how many candidates correspond with a true lesion. To be able to compute these statistics, we need to define a correctly localized lesion first. Let say that a candidate corresponds to a lesion if they overlap by more than 75%. In that case, we will say that the lesion was correctly localized.

Here, the presented approach is compared with the MRF method described in the previous section. This method forms our baseline. As shown in Tab. 6.4, the F-measure achieved by

	F-measure			precision			recall		
	mean	median	std	mean	median	std	mean	median	std
all	0.844	0.924	0.179	0.834	0.951	0.218	0.885	0.933	0.109
no_LBP	0.844	0.921	0.181	0.836	0.952	0.218	0.879	0.916	0.111
dwsb	0.832	0.92	0.195	0.801	0.921	0.231	0.899	0.952	0.104
gwsb	0.825	0.908	0.203	0.797	0.911	0.235	0.901	0.952	0.102
ws	0.74	0.862	0.265	0.707	0.823	0.297	0.913	0.958	0.098
dwb	0.853	0.932	0.17	0.821	0.913	0.207	0.912	0.956	0.103

Table 6.3: Accuracy of tumors localization with combination of saliency maps and LLS refinement

(a)   (b)   (c)
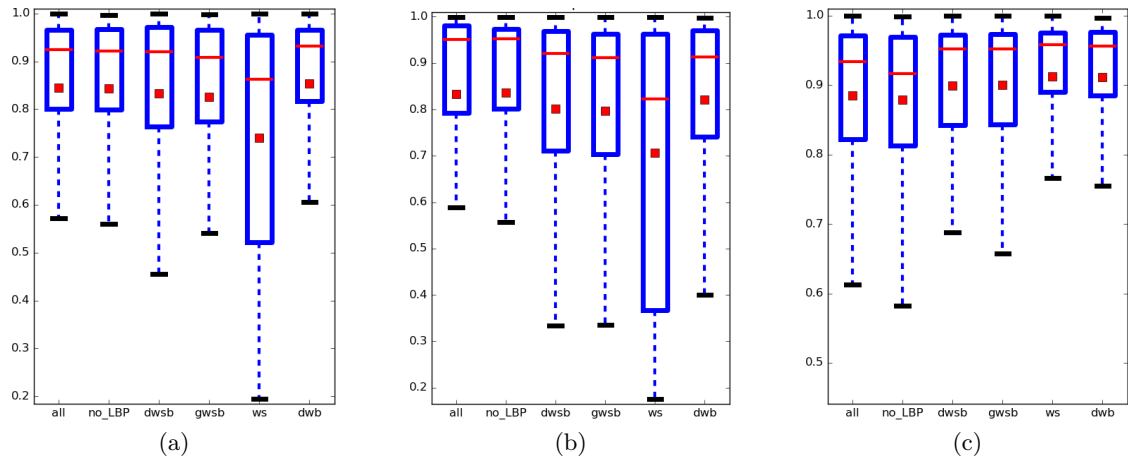
Figure 6.5: Accuracy of tumor localization with combination of saliency maps and LLS refinement: the F-measure (a), the precision (b) and the recall (c).

the presented approach is **0.838**, whereas the F-measure of the baseline method is **0.758**. Although the precision of the baseline method is higher, the recall is more important in our case. The CAD system should be able to localize as much lesions as possible even at the cost of reasonably lower precision. In other words, a reasonable amount of false candidates are tolerated as long as the CAD localizes most of the lesions.

	F-measure	**precision**	**recall**
dwb	0.838	0.775	0.913
MRF	0.758	0.801	0.719

Table 6.4: Object-wise accuracy of lesion localization.

Figure 6.6: Example of localization results: input data (a), ground-truth (b), basic MRF (c), intensity difference saliency map (d), *dwb* saliency map (e).

6.2   Classification

Once the data are registered and tumors are localized it is time for the final step - tumor classification. Three situations may arise regarding the tumor localization:

- corresponding tumors overlap . . . the ideal situation,
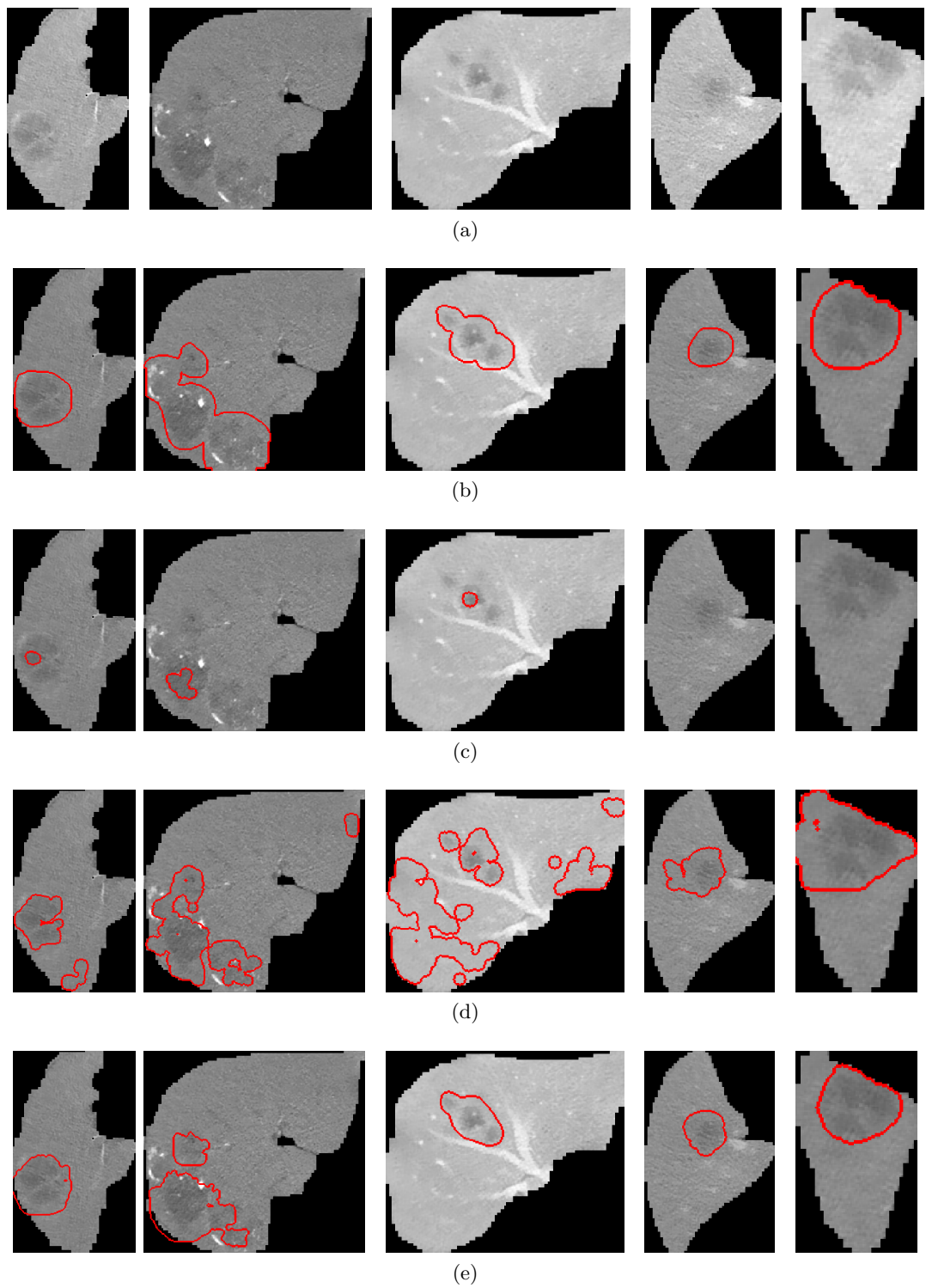
- non-corresponding tumors overlap . . . two different tumors were localized at the same place,

- only one tumor was detected.

The first situation is the most common and the classification can be applied directly. On the other hand, is hard to distinguish if the tumors correspond or not. Although situations where the system detects different tumors in the same place are rare, we should be aware that it can happen. For example, in the case of multiple lesions, the registration process can misalign them and pair up diferrent lesions. Once that happens, there is not much we can do. Fortunately, its extremely rare to find different types of lesions occupying nearby space. Moreover, the registration process did not allow any extreme transformation that would align tumors localized far away in the original data. Therefore, we can assume that the tumors are aligned well.

The last case is not so rare. There are tumors that are isodense in one phase and therefore cannot be localized. A simple solution to this problem would be to assume that the mean density of the missing tumor is isodense. Unfortunately, the missing tumor can be present/visible but only with pure contrast that makes it undetectable by the system. Therefore we will not assume isodensity a priori but examine the aligned space for any type of inhomogeneity. This way, we are sometimes able to detect the tumor a posteriori. Only if not even this analysis detects the missing tumor we will use the mean density of registered space corresponding the the detected tumor in the other phase.

The main step in the classification is to analyze the localized tumor in both phases. The density of the tumor may be homogenous or heterogeneous (Fig. 2.7b). Unfortunately, the algorithm cannot produce good results using this information only. To be able to identify individual tumors more precisely, more features needs to be derived. For example, detecting possible calcifications (Fig. 2.7a) and/or a capsule (Fig. 2.6c) around the tumor made it possible to distinguish between FLC and HCC. Other significant features are the presence of a central scar inside the tumorous tissue (Fig. 2.13b), hemorrhage (bleeding) (Fig. 2.9c) and the size of the tumor.

Despite all, the biggest problem of the classification still remains - the diversity of individual tumors. Often a representative of a tumor type may present itself drastically different and it is easy to misclassify it. Therefore, the results of tumor classifications needs to be reviewed by an expert. Moreover, a better knowledge of a patient is always advantageous. For example, a typical problem is to distinguish between HA, HCC, FNH and hypervascular metastasis. But more typical tumor for otherwise healthy woman using contraception is HA, whereas a patient with a history of cirrhosis will more likely have an HCC. Therefore, a connection to a patient database may be helpful for the tumor identification.

From Section 2.7 a decision tree shown in Fig. 6.7 was created. As already mentioned, the diversity of the liver tumors is enormous. Therefore, the decision tree contains the most common cases.

6.2.1   Feature identification

As already described, we need to identify features that will help us to distinguish between individual types of tumors. The first thing to do is to check whether a detected object lies inside another one. If not, we need to differentiate between homogenous and heterogeneous tissue. Once again, it is possible to use a histogram or a GLCM. As usual in this work, we prefer the latter option where the spatial relationship is considered. By thresholding the GLCM and measuring the width of resulting blob we are able to distinguish between **homogeneous** and **heterogeneous** object, see Fig. 6.8.

In situations, where an object is detected inside another object, wee need to specify the relationship of these objects, i.e. we need to check if one of the objects is a scar, a calcification, a capsule or a hemorrhage. At first we are looking at the area of the inner object $area_i$ w.r.t. the area of outer object $area_o$. If the ratio $r = \frac{area_i}{area_o} < 0.5$ then the inner object is most likely a scar, a calcification or a hemorrhage. A **scar** (Fig. 6.9a) is a small hypodense object whereas a **calcification** (Fig. 6.9b) is a small hyperdense object. A **hemorrhage** could be a small or a large hyperdense object. In the case of a small hemorrhage, it can be differentiated from a calcification by the density - a calcification is much more dense then a hemorrhage, it is often visualized as a pure white object.

If the ratio $r = \frac{area_i}{area_o} > 0.5$ then the inner object would be a hemorrhage or the outer object would be a capsule. A **capsule** (Fig. 6.9d) can be hyperdense or hypodense. For distinguishong between these two features a compactness of the inner objects may be used - the tumor with a capsule is more compact than a hemorrhage. A decision tree for identifying a feature of a tumor is shown in Fig. 6.10.

6.2.2   Results

The classification accuracy was again computed in the sense of the F-measure, the precision and the recall. The results are listed in Table 6.5.

The results show that the best accuracy is achieved in the case of classifying **cysts**. This is an expected result due to the nature and low diversity of the liver cysts.

Less accurately were classified the hepatic **metastasis**. This class combines the case of both hypervascular and hypovascular cases. Both of them are easily mistaken if the capsule is not localized.

The localization of a capsule is also a problem when classifying the **HCC**. When the capsule is not localized in this case, the HCC is often classified as the HA.

An interesting result is that the **FLC** is mostly mistaken for the FNH. This is due to the fact, that the FNH is often visualized in the portal phase as isodense or even hypodense tissue. In addition, if the FLC does not have a capsule then it is very similar to the FNH.

A characteristic feature of the **ICC** is the presence of the central scar. Unfortunately, the scar is sometimes poorly visible and therefore not localized as a feature. On the other hand,
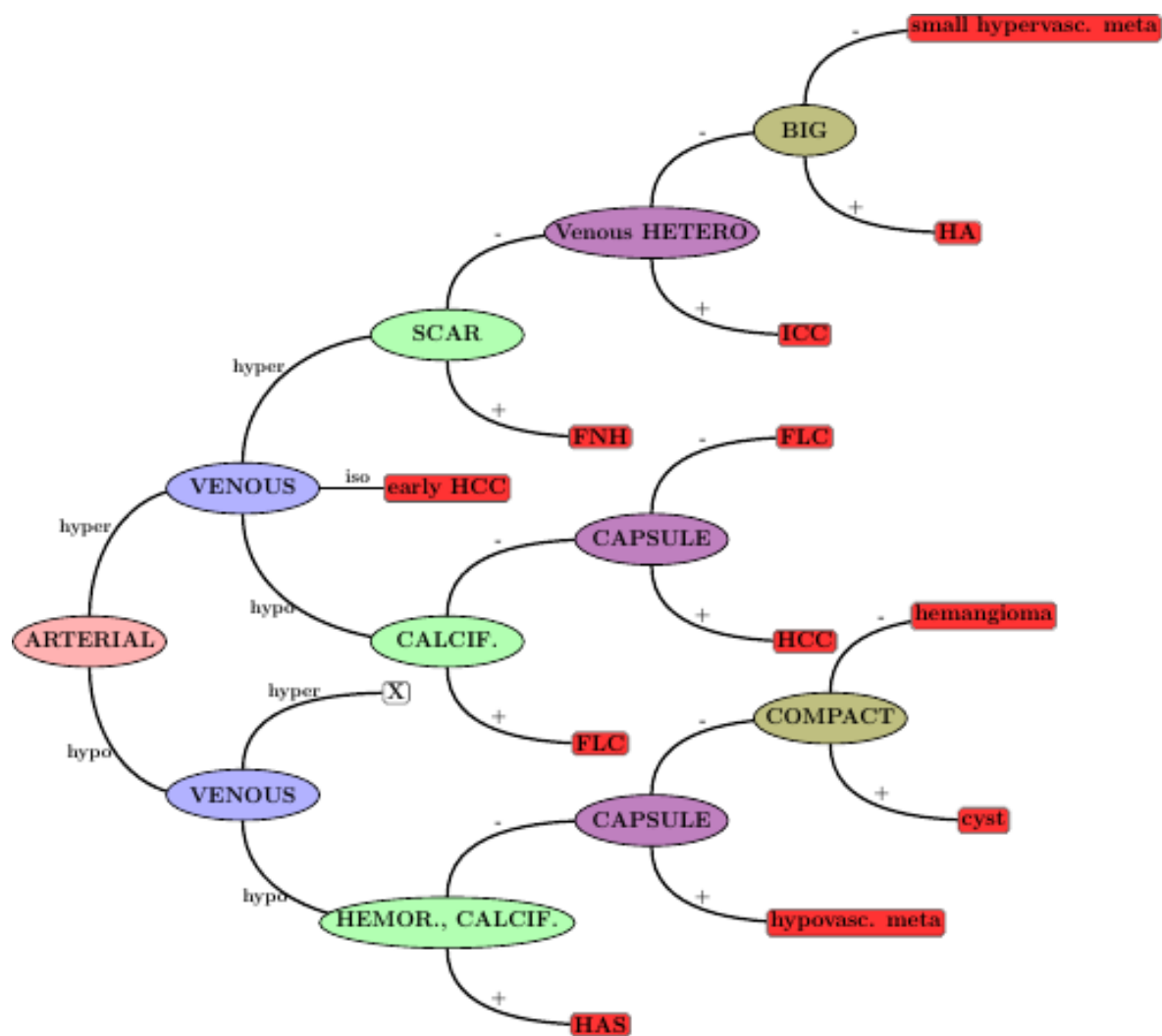
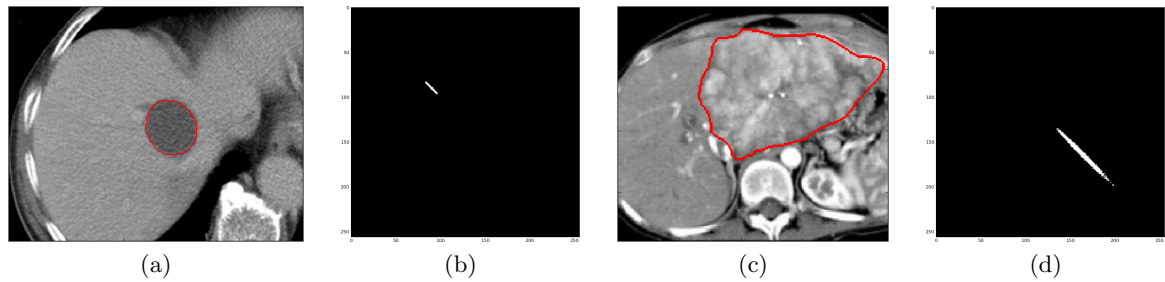Figure 6.7: Decision tree for lesion classification

Figure 6.8: A homogeneous tumor (a) can be differentiated from a heterogenous tumor (c) analysing their thresholded GLCM. The blob in the GLCM of the heterogeneous tumor (d) is much wider than the blob in the GLCM of the homogeneous tumor (b).
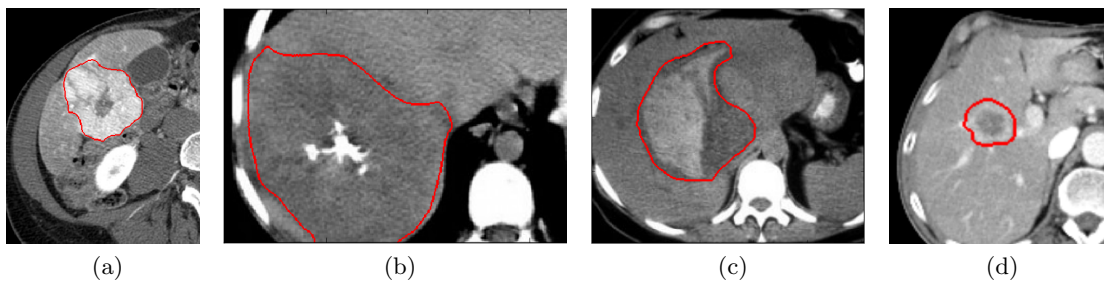
Figure 6.9: A tumor can be classified based on a feature present: a scar (a), a calcification (b), a hemorrhage (c) and a capsule (d) enclosing a tumor.

the corresponding GLCM will still be affected by the scar and therefore the tumor may be identified as heterogeneous.

In the case of a small **HAS** with enormous hemorrhage, the whole tumor looks just like an hypervascular metastasis. The hypodense tissue of the HAS is then misclassified as the hypodense capsule of the metastasis.

A very similar problem arises in the case of the **FNH** and the hypovascular meta. An enormous scar in a small FNH can be identified as the hypodense center of a hypovascular metastasis. Therefore, the FNH is mostly mistaken for the hypovascular metastasis.
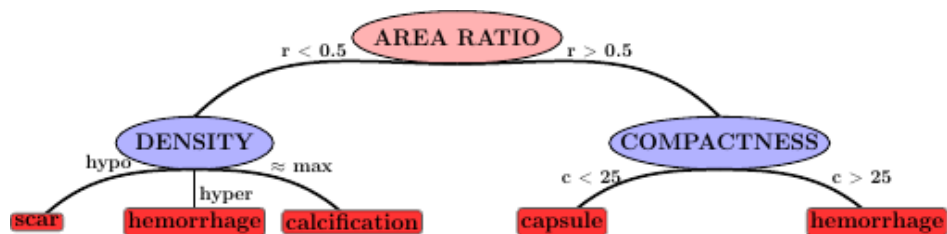
Figure 6.10: Decision tree for identification of a tumor feature.

A calcification and/or a hemorrhage of the HAS is the only feature that distinguish it from the HG. If they are not localized then the **HG** will be misclassified as the HAS.

The reason for a frequent misclassification of the **HA** and the HCC is that the small HCC looks just like the HA. The only difference lies in their area. If the HA is atypically small, it can be easily classified as a small HCC.

	HCC	FLC	ICC	HAS	meta	HG	FNH	HA	cyst
F-measure	0.736	0.759	0.697	0.749	0.646	0.804	0.677	0.808	0.861
precision	0.733	0.748	0.704	0.768	0.653	0.788	0.673	0.816	0.853
recall	0.74	0.77	0.69	0.73	0.64	0.82	0.68	0.8	0.87

Table 6.5: Accuracy of tumor classification

	HCC	FLC	ICC	HAS	meta	HG	FNH	HA	cyst
HCC	74	7	2	0	6	1	5	5	0
FLC	6	80	0	2	0	3	8	4	0
ICC	1	1	69	6	7	2	8	2	4
HAS	1	3	4	73	8	6	1	0	4
meta	5	2	9	6	64	3	6	1	4
HG	2	2	3	5	3	82	0	0	3
FNH	6	7	4	0	8	1	68	6	0
HA	6	3	4	1	0	1	5	80	0
cyst	0	1	3	2	2	5	0	0	87

Table 6.6: Confusion matrix of tumor classification.

Chapter 7

Conclusion

Despite all the progress that modern medicine is undergoing, cancer remains a disease hard to cure. Liver cancer is the sixth most common type of cancer worldwide as stated in [American Cancer Society, 2016]. This is caused by the high blood volume that flows through the liver, which makes it easier for secondary tumors to develop. The most common treatment procedure after the cancer was diagnosed is liver resection, i.e. surgically removing a part of the liver containing the tumor. The decision whether such an operation is possible is determined by a radiologist and a surgeon. The task of the radiologist is to acquire and describe the CT data of the patient. That means the doctor has to localize all tumors inside the liver parenchyma and identify them as specifically as possible. The surgeon then decides if the tumors are resectable and if the risks of the operation are not too high.

The work of the doctors is supported by numerous systems that help them to increase their effectiveness, accuracy and objectivity. Such systems are called computer-aided diagnostic systems (CAD). The goal of this thesis is to design and implement a CAD system that will localize and further identify liver tumors.

The input of the system is a pair of CT data obtained from a so called differential diagnosis, i.e. two serial screenings of a patient with a time interval of several seconds. Because identical screening conditions can not be guaranteed, it is necessary to first register the data. The issue of the registration process is described in Chapter 5, where the total number of six approaches for solving this task are presented. These approaches differ in the similarity measure used and the transformation type estimated by the registration process. We tested the mean squares and the mutual information similarity measures and estimated the similarity, the affine and the spline transformation. The results show that the best accuracy is achieved for mutual information and the spline transformation. The results are no surprise. Because the human body is not a rigid object, its transformation would be elastic as well. The data differs in intensity distribution, therefore using the mutual information similarity measure provides us with the best accuracy. On the other hand, from the nature of pairing up the tumors in the following processing, demands on accuracy are not critical. The results show that by using the similarity transform and the MS similarity measure, the registration produces comparable precise results in 10x less time.

To be able to localize liver tumors, the liver parenchyma needs to be segmented first. There are a large number of methods that can be used to solve this task. In this work, a cellular automaton called Grow Cut is used. This algorithm needs to be initialized by a set of seed points that can be marked interactively by a user or derived in an autonomous way. Because this work focuses on a fully automatic approach, the latter option was used. Chapter 4 describes two possibilities for the initialization - the histogram analysis and the GLCM analysis. The GLCM analysis combines color and spatial features; therefore, it describes individual blobs more precisely. For the GLCM analysis the Mean Shift algorithm was used. After the blobs were separated, the one representing the liver needs to be chosen. For this task, a scoring system was developed. This system assigns a score to each blob that reflects the blob's area, intensity, compactness and position. The blob with the highest score is chosen as the one that corresponds to the liver. Due to the nature of the Grow Cut algorithm, the results is more a coarse segmentation. To achieve higher accuracy, the coarse segmentation is further refined using the localized active contours. This refinement increases the mean F-measure from 0.82 to 0.89. The shortest path basins algorithm was tested for liver segmentation as well but with significantly worse results - it achieved the mean F-measure of value 0.708.

Once the liver is segmented, the system can proceed to the lesion localization. Because the lesions are often non-contrast or their boundary is partly indistinct, conventional methods mostly fail when solving this task. Moreover, using only area-based or boundary based methods yields very poor results. Therefore, an approach based on the Markov random fields was designed. The MRF is defined as the Ising model and optimized by the graph cut algorithm. Therefore, only cliques of the first order needs to be defined. As shown in Section 6.1 using the common approach of defining the unary term by a histogram analysis does not produce accurate results. Therefore an approach based on saliency maps was adopted and extended. The total number of seven different saliency maps were extracted and used - the intensity difference, the sliding window, the blob detection, the GLCM analysis, the histogram analysis, the texture analysis and the shape. The experiments show that best results are achieved using the intensity difference (F-measure = 0.673), the sliding window (F-measure = 0.67) and the blobs detection (F-measure = 0.658). Next, more than 20 combinations of the individual saliency maps were tested. The best result was achieved using the combination of the intensity difference, the sliding window and the blob detection (F-measure = 0.743). Even higher accuracy was achieved by refining the results again by the localized active contours. By this refinement the F-measure increased to 0.853.

The last step was to classify the tumors into one of the nine most common lesions in liver - hepatocellular carcinoma (HCC), fibrolamellar carcinoma (FLC), cholangiocellular carcinoma (ICC), hepatic angiosarcoma (HAS), secondary tumors (metastasis), hemangioma (HG), focal nodular hyperplasia (FNH), hepatocellular adenoma (HA) and cysts. There are two major problems with this task: lack of the data and the nature of the data. The former problem can be partly solved by augmenting the existing data but there is no chance to obtain enough data to train a modern classifier like a (convolutional) neural network. The latter problem is caused by a high diversity of the data. Same type of tumors can appear in several ways and an atypical case of a tumor can perfectly correspond to a definition of a different tumor. These cases are almost impossible to distinguish and therefore the overall accuracy of the classification cannot achieve high values. The classification is done by a decision tree that

analyses a feature at each level of its branching. The first discriminative feature is the density of the tumor in both the arterial and the portal venous phase. Several features are searched inside the tumorous tissue - a scar, a calcification, a hemorrhage, a tumor inside a capsule. Moreover, the tumorous tissue can be homogeneous or heterogeneous, large or small. The results show that the best accuracy of classification is achieved in the case of the cysts (F-measure = 0.861). The cysts are well-separated, contrast and homogeneous tumors. On the other hand, the lowest accuracy was achieved when classifying secondary tumors (F-measure = 0.646). This is caused by the frequently missing capsule, which is a characteristic feature of the metastasis.

The system presented is able to segment the liver, localize lesions and identify them. The weak link is the identification of the lesions caused by the lack of the data and especially the high diversity of the appearance of the lesions. The former problem could be at least partly solved by asking more hospitals for cooperation. This could increase the amount of the data used for the training and a modern classifier could be used. The latter problem is hard to solve. Connecting the system to a database of patients' anamnesis could help solve problems such as distinguishing similar looking tumors, e.g. HCC and HA. Both tumors can appear almost identical, but the former is more common in men using anabolic steroids or patients with a history of cirrhosis, whereas the latter is a more typical tumor for an otherwise healthy woman using contraception. Of course, to connect such a system to hospital system with these private data is a delicate task due to the nature of the data and the law protecting them.

Bibliography

[Adalsteinsson and Sethian, 1995] Adalsteinsson, D. and Sethian, J. A. (1995). A fast level set method for propagating interfaces. *Journal of Computational Physics*, 118(2):269–277.

[Amelio and Pizzuti, 2012] Amelio, A. and Pizzuti, C. (2012). An evolutionary and graph-based method for image segmentation. In *PPSN (1)*, volume 7491 of *Lecture Notes in Computer Science*, pages 143–152. Springer.

[American Association for Cancer Research,] American Association for Cancer Research.

[American Cancer Society, 2016] American Cancer Society (2016). Cancer facts & figures 2016.

[Bäck, 1996] Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, Oxford, UK.

[Besag, 1974] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236.

[Blake et al., 2011] Blake, A., Kohli, P., and Rother, C. (2011). *Markov Random Fields for Vision and Image Processing*. MIT Press.

[Bohatá et al., 2009] Bohatá, Š. et al. (2009). *Přínos kontrastního ultrazvukového vyšetření v diferenciální diagnostice ložiskových procesů jater*. PhD thesis, Masarykova univerzita, Lékařská fakulta.

[Boykov and Funka-Lea, 2006] Boykov, Y. and Funka-Lea, G. (2006). Graph cuts and efficient n-d image segmentation. *International Journal of Computer Vision*, 70(2):109–131.

[Boykov and Jolly, 2000] Boykov, Y. and Jolly, M.-p. (2000). Interactive organ segmentation using graph cuts. *October*, 1935:276–286.

[Boykov and Kolmogorov, 2004] Boykov, Y. and Kolmogorov, V. (2004). *An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*, volume 26, pages 1124–1137. IEEE.

[Boykov and Veksler, 2006] Boykov, Y. and Veksler, O. (2006). Graph cuts in vision and graphics: Theories and applications. *Handbook of Mathematical Models in Computer Vision SpringerVerlag*, 1(100):100–119.

[Boykov et al., 1998] Boykov, Y., Veksler, O., and Zabih, R. (1998). Markov random fields with efficient approximations. In *Computer vision and pattern recognition, 1998. Proceedings. 1998 IEEE computer society conference on*, pages 648–655. IEEE.

[Boykov et al., 2001] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239.

[Boykov and Jolly, 2001] Boykov, Y. Y. and Jolly, M. P. (2001). Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. *Proceedings Eighth IEEE International Conference on Computer Vision ICCV 2001*, 1(July):105–112.

[Caselles et al., 1993] Caselles, V., Catté, F., Coll, T., and Dibos, F. (1993). A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31.

[Caselles et al., 1997] Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79.

[Chan and Vese, 2001] Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277.

[Choudhary et al., 2008] Choudhary, A., Moretto, N., Ferrarese, F. P., and Zamboni, G. A. (2008). An entropy based multi-thresholding method for semi-automatic segmentation of liver tumors. In *MICCAI Workshop*, volume 41, pages 43–49.

[Chow, 1962] Chow, C. (1962). A recognition method using neighbor dependence. *IRE Transactions on Electronic Computers*, pages 683–690.

[Čihák, 2011] Čihák, R. (2011). *Anatomie 2*. Grada.

[Collier and Sherman, 1998] Collier, J. and Sherman, M. (1998). Screening for hepatocellular carcinoma. *Hepatology*, 27(1):273–278.

[Comaniciu and Meer, 2002] Comaniciu, D. and Meer, P. (2002). Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

[Doi, 2007] Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics the official journal of the Computerized Medical Imaging Society*, 31(4-5):198–211.

[Dou et al., 2016] Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., and Heng, P.-A. (2016). 3d deeply supervised network for automatic liver segmentation from ct volumes. *CoRR*.

[Felzenszwalb and Huttenlocher, 2004] Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181.

[Ferlay et al., 2015] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012. *International journal of cancer*, 136(5):E359–E386.

[Fitzpatrick and Sonka, 2000] Fitzpatrick, J. M. and Sonka, M. (2000). Medical image processing and analysis. *Image Rochester NY*, 2(1):1108.

[Florin et al., 2007a] Florin, C., Paragios, N., Funka-Lea, G., and Williams, J. (2007a). *Information Processing in Medical Imaging: 20th International Conference, IPMI 2007, Kerkrade, The Netherlands, July 2-6, 2007. Proceedings*, chapter Liver Segmentation Using Sparse 3D Prior Models with Optimal Data Support, pages 38–49. Springer Berlin Heidelberg.

[Florin et al., 2007b] Florin, C., Paragios, N., Funka-Lea, G., and Williams, J. (2007b). Liver segmentation using sparse 3d prior models with optimal data support. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 38–49. Springer.

[Ford and Fulkerson, 1956] Ford, L. R. and Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, 8(3):399–404.

[Ford and Fulkerson, 1962] Ford, L. R. and Fulkerson, D. R. (1962). *Flows in Networks*. Princeton University Press.

[Friston et al., 1996] Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., and Turner, R. (1996). Movement related effects in fmri time-series. *Magn Res Med*, 35:346–355.

[Geman and Geman, 1984a] Geman, S. and Geman, D. (1984a). Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

[Geman and Geman, 1984b] Geman, S. and Geman, D. (1984b). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.

[Goldberg and Tarjan, 1988] Goldberg, A. V. and Tarjan, R. E. (1988). A new approach to the maximum-flow problem. *J. ACM*, 35(4):921–940.

[Goldberg, 1989] Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition.

[Gudbjartsson and Patz, 1995] Gudbjartsson, H. and Patz, S. (1995). The rician distribution of noisy mri data. *Magnetic Resonance in Medicine*, 34(6):910–914.

[Hajnal et al., 1995] Hajnal, J. V., Saeed, N., Soar, E. J., Oatridge, A., Young, I. R., and Bydder, G. M. (1995). A registration and interpolation procedure for subvoxel matching of serially acquired mr images. *Journal Of Computer Assisted Tomography*, 19(2):289–296.

[Hamer et al., 2006] Hamer, O. W., Aguirre, D. A., Casola, G., Lavine, J. E., Woenckhaus, M., and Sirlin, C. B. (2006). Fatty liver: Imaging patterns and pitfalls 1. *Radiographics*, 26(6):1637–1653.

[Hammersley and Clifford, 1971] Hammersley, J. M. and Clifford, P. E. (1971). Markov random fields on finite graphs and lattices.

[Holland, 1975] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

[Ising, 1925] Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258.

[Itseez, 2015] Itseez (2015). Open source computer vision library. `https://github.com/itseez/opencv`.

[Itti and Koch, 2000] Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10):1489–1506.

[Jedlicka and Ryba, 2016] Jedlicka, P. and Ryba, T. (2016). Genetic algorithm application in image segmentation. *Pattern Recognition and Image Analysis*, 26(3):497–501.

[Kennedy and Eberhart, 1995] Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948 vol.4.

[Klener, 2011] Klener, P. e. a. (2011). *Vnitřní lékařství*. Galén.

[Koch and Ullman, 1985] Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–227.

[Kolmogorov and Boykov, 2005] Kolmogorov, V. and Boykov, Y. (2005). What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. *Tenth IEEE International Conference on Computer Vision ICCV05 Volume 1*, I(October):564–571.

[Kumar and Moni, 2010] Kumar, S. and Moni, R. (2010). Diagnosis of liver tumor from ct images using curvelet transform. *International Journal on Computer Science and Engineering*, 2(4):1173–1178.

[Lankton and Tannenbaum, 2008] Lankton, S. and Tannenbaum, A. (2008). Localizing region-based active contours. *IEEE Transactions on Image Processing*, 17(11):2029–2039.

[Lee et al., 2001] Lee, W. J., Lim, H. K., Jang, K. M., Kim, S. H., Lee, S. J., Lim, J. H., and Choo, I. W. (2001). Radiologic spectrum of cholangiocarcinoma: Emphasis on unusual manifestations and differential diagnoses. *RadioGraphics*, 21(suppl_1):S97–S116.

[Li, 2009] Li, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition.

[Llovet et al., 2003] Llovet, J. M., Burroughs, A., and Bruix, J. (2003). Hepatocellular carcinoma. *The Lancet*, 360:1907–1917.

[Maes et al., 1997] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198.

[Maier et al., 2008] Maier, F., Wimmer, A., Soza, G., Kaftan, J. N., Fritz, D., and Dillmann, R. (2008). Automatic liver segmentation using the random walker algorithm. In *Bildverarbeitung für die Medizin 2008*, pages 56–61. Springer.

[Maintz and Viergever, 1998] Maintz, J. B. and Viergever, M. A. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36.

[Mala and Sadasivam, 2005] Mala, K. and Sadasivam, V. (2005). Automatic segmentation and classification of diffused liver diseases using wavelet based texture analysis and neural network. In *2005 Annual IEEE India Conference-Indicon*, pages 216–219. IEEE.

[Mala and Sadasivam, 2006] Mala, K. and Sadasivam, V. (2006). Wavelet based texture analysis of liver tumor from computed tomography images for characterization using linear vector quantization neural network. In *2006 International Conference on Advanced Computing and Communications*, pages 267–270. IEEE.

[Malladi et al., 1993] Malladi, R., Sethian, J. A., and Vemuri, B. C. (1993). A topology independent shape modeling scheme. *SPIE*, 2031:246–258.

[Malladi et al., 1995] Malladi, R., Sethian, J. A., and Vemuri, B. C. (1995). Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175.

[Martin et al., 2001] Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423.

[Massoptier and Casciaro, 2008] Massoptier, L. and Casciaro, S. (2008). A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from ct scans. *European radiology*, 18(8):1658–1665.

[Meinzer et al., 2002] Meinzer, H.-P., Thorn, M., and Cardenas, C. E. (2002). Computerized planning of liver surgery-an overview. *Computers & Graphics*, 26(4):569–576.

[Mosaliganti et al., 2009] Mosaliganti, K., Smith, B., Gelas, A., Gouaillard, A., and Megason, S. (2009). Level set segmentation : Active contours without edges. *Insight*, pages 1–9.

[Ojala et al., 1996] Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59.

[Ojala et al., 2002] Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987.

[Park and Song, 1998] Park, Y. and Song, M. (1998). A genetic algorithm for clustering problems. In Koza, J. R., Banzhaf, W., Chellapilla, K., Deb, K., Dorigo, M., Fogel, D. B., Garzon, M. H., Goldberg, D. E., Iba, H., and Riolo, R., editors, *Genetic Programming 1998: Proceedings of the Third Annual Conference*, pages 568–575. Morgan Kaufmann.

[Pavlidis, 1986] Pavlidis, T. (1986). A critical survey of image analysis methods. *ICPR*, pages 502–511.

[Peng et al., 2015] Peng, J., Hu, P., Lu, F., Peng, Z., Kong, D., and Zhang, H. (2015). 3d liver segmentation using multiple region appearances and graph cuts. *Medical physics*, 42(12):6840–6852.

[Perez, 1998] Perez, P. (1998). Markov random fields and images. *CWI Quarterly*, pages 413–437.

[Potts, 1952] Potts, R. B. (1952). Some generalized order-disorder transformations. In *Mathematical proceedings of the cambridge philosophical society*, volume 48, pages 106–109. Cambridge Univ Press.

[Radiology Assistant,] Radiology Assistant. Liver masses i, ii.

[Radiopaedia,] Radiopaedia. Radiopaedia.

[Ressurreição, 1970] Ressurreição, J. (1970). Benign hepatic tumors: how to diagnose?-a multimodality approach with emphasis on mri. European Congress of Radiology 2011.

[Ryba et al., 2013] Ryba, T., Jirik, M., and Zelezny, M. (2013). An automatic liver segmentation algorithm based on grow cut and level sets. *Pattern recognition and image analysis*, 23(4):502–507.

[Ryba and Zelezny, 2015] Ryba, T. and Zelezny, M. (2015). An automatic image segmentation algorithm involving shortest path basins. volume 25, pages 89–95. Springer.

[Ryba and Zelezny, 2016] Ryba, T. and Zelezny, M. (2016). An automatic initialization of interactive segmentation methods using shortest path basins. *Pattern Recognition and Image Analysis*, 26(2):336–342.

[Selle et al., 2002] Selle, D., Preim, B., Schenk, A., and otto Peitgen, H. (2002). Analysis of vasculature for liver surgical planning. In *Proc. of IEEE Trans. Med. Imaging*, pages 1344–1357.

[Shi and Malik, 2000] Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888 –905.

[Siriwardena et al., 2014] Siriwardena, A. K., Mason, J. M., Mullamitha, S., Hancock, H. C., and Jegatheeswaran, S. (2014). Management of colorectal cancer presenting with synchronous liver metastases. *Nature Reviews Clinical Oncology*, 11(8):446–459.

[Soler et al., 2001] Soler, L., Delingette, H., Malandain, G., Montagnat, J., Ayache, N., Koehl, C., Dourthe, O., Malassagne, B., Smith, M., Mutter, D., et al. (2001). Fully automatic anatomical, pathological, and functional segmentation from ct scans for hepatic surgery. *Computer Aided Surgery*, 6(3):131–142.

[Song et al., 2013] Song, X., Cheng, M., Wang, B., Huang, S., Huang, X., and Yang, J. (2013). Adaptive fast marching method for automatic liver segmentation from ct images. *Medical physics*, 40(9):091917.

[Sonka et al., 2007] Sonka, M., Hlavac, V., and Boyle, R. (2007). *Image Processing, Analysis, and Machine Vision*, volume 3. PWS Publishing.

[Stipa et al., 2006] Stipa, F., Yoon, S. S., Liau, K. H., Fong, Y., Jarnagin, W. R., D'Angelica, M., Abou-Alfa, G., Blumgart, L. H., and DeMatteo, R. P. (2006). Outcome of patients with fibrolamellar hepatocellular carcinoma. *Cancer*, 106(6):1331–1338.

[Studholme et al., 1999] Studholme, C., Hill, D., and Hawkes, D. (1999). An overlap invariant entropy measure of 3d medical image alignment . *Pattern Recognition*, 32(1):71–86.

[Taylor, 2011] Taylor, C. R. (2011). Cirrhosis imaging.

[Terrier et al., 2000] Terrier, F., Grossholz, M., and Becker, C. D. (2000). *Spiral CT of the Abdomen*. Springer Science & Business Media.

[The Liver Imaging Atlas,] The Liver Imaging Atlas. The liver imaging atlas.

[Theo, 1982] Theo, P. (1982). Algorithms for graphics and image processing. *USA.: Computer Science Press, Inc.*

[Třeška et al., 2010] Třeška, V., Liška, V., Skalický, T., Sutnar, A., Šmíd, D., Ňaršanská, A., Vachtová, M., Třešková, I., Brůha, J., and Vyčítal, O. (2010). Jaterní metastázy jiného než kolorektálního původu. *Rozhledy v chirurgii*, 89(3):202–207.

[Unnikrishnan and Hebert, 2005] Unnikrishnan, R. and Hebert, M. (2005). Measures of similarity. In *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, volume 1, pages 394–394. IEEE.

[Válek et al., 2006] Válek, V., Kala, Z., and Kiss, I. (2006). *Maligní ložiskové procesy jater: diagnostika a léčba včetně minimálně invazivních metod*. Grada.

[Vezhnevets and Konouchine, 2004] Vezhnevets, V. and Konouchine, V. (2004). Growcut - interactive multi-label n-d image segmentation by cellular automata. *Cybernetics*, page 150–156.

[Viola and Wells, 1997] Viola, P. and Wells, W. M. (1997). Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):16–23.

[Wells et al., 1996] Wells, W. M., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51.

[Winkler, 1995] Winkler, G. (1995). Image analysis, random fields and dynamic monte carlo methods.

[World Cancer Research Fund International - American Institute for Cancer Research, 2015] World Cancer Research Fund International - American Institute for Cancer Research (2015). Liver cancer 2015 report: Diet, nutrition, physical activity and liver cancer.

[Xu et al., 2000] Xu, C., Pham, D. L., and Prince, J. L. (2000). *Medical Image Segmentation Using Deformable Models*, volume 2, pages 129–174. SPIE Press.

[Yezzi et al., 1997] Yezzi, A., Kichenassamy, S., Kumar, A., Olver, P., and Tannenbaum, A. (1997). A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16(2):199–209.

[Yezzi et al., 2002] Yezzi, A., Tsai, A., and Willsky, A. (2002). A fully global approach to image segmentation via coupled curve evolution equations. *Journal of Visual Communication and Image Representation*, 13(1-2):195–216.

[Zhou et al., 2008] Zhou, J., Xiong, W., Tian, Q., Qi, Y., Liu, J., Leow, W. K., Han, T., Venkatesh, S. K., and Wang, S.-c. (2008). Semi-automatic segmentation of 3d liver tumors from ct scans using voxel classification and propagational learning. In *MICCAI Workshop*, volume 41, page 43.

Bibliography

[1] Železný, M and Krňoul, Z and Jiřík, M and Ryba, T. **Simple Gesture Based Medical Data Interface**, *Proceedings of the 8th Open German-Russian Workshop*, Saint Petersburg, Russia, 2010.

[2] Ryba, T and Jiřík, M **Segmentace obrazu v lékařské diagnostice**, *Diagnostika, léčba a prevence závažných civilizačních onemocnění*, Plzeň, 2010, pp. 82-82.

[3] Jiřík, M and Ryba, T **Interaktivní segmentace pomocí metody Graph-Cut v prostředí Matlab**, *SVK 2010 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2010, pp. 23-24.

[4] Jiřík, M and Ryba, T and Železný, M **Texture based segmentation using graph cut and Gabor filters**, *Pattern Recognition and Image Analysis*, Volume 21, Issue 2, Springer, 2011, pp. 258-261.

[5] Ryba, T and Jiřík, M and Mírka, H. **Analýza jaterního parenchymu metodami zpracování obrazu**, *Experimentální výzkum v medicíně a jeho klinická aplikace*, Plzeň, 2011, pp. 23-23.

[6] Ryba, T **Registrace obrazu v úloze monitorování tumorů**, *SVK 2011 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2011, pp. 93-94.

[7] Jiřík, M and Ryba, T and Železný, M **Gabor filter and graph cut based texture analysis**, *Pattern Recognition and Image Analysis*, Volume 22, Issue 1, Springer, 2012, pp. 215-220.

[8] Ryba, T and Železný, M and Nedbal, L **Registration of Chlorophyll Fluorescence Emission and 3D Models of Plants**, *Proceedings of EPPN Developers' Workshop*, 2012.

[9] Ryba, T **Automatic Localization of Liver Tumors**, *Proceedings of the 1st Czech-Russian Forum of Young Scientists*, Plzeň, 2012, pp. 25-26.

[10] Ryba, T **Image Processing in Plant Science**, *Proceedings of the 1st Czech-Russian Forum of Young Scientists*, Plzeň, 2012, pp. 15-16.

[11] Ryba, T **Analýza rostlin metodami zpracování obrazu**, *SVK 2012 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2012, pp. 87-88.

[12] Ryba, T and Jiřík, M and Železný, M **An automatic liver segmentation algorithm based on grow cut and level sets**, *Pattern Recognition and Image Analysis*, Volume 23, Issue 4, Springer, 2013, pp. 502-507.

[13] Ryba, T and Železný, M **Co-registration of chlorophyll fluorescence emission and 3D models of plants**, *Proceedings of the 3rd annual Global Change and Resilience Conference*, 2013, pp. 264-267.

[14] Ryba, T **Analýza rostlin metodami zpracování obrazu**, *SVK 2013 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2013, pp. 79-80.

[15] Ryba, T **Lokalizace nádorů pomocí automatického vyhledávání zajímavých oblastí v obraze**, *SVK 2014 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2014, pp. 93-94.

[16] Timr, Š and Brabec, J and Bondar, A and Ryba, T and Železný, M and Lazar, J and Jungwirth, P **Nonlinear Optical Properties of Fluorescent Dyes Allow for Accurate Determination of Their Molecular Orientations in Phospholipid Membranes**, *Journal of Physical Chemistry B*, Volume 119, Issue 30, 2015, pp. 9706-9716.

[17] Ryba, T and Železný, M **An automatic image segmentation algorithm involving shortest path basins**, *Pattern Recognition and Image Analysis*, Volume 25, Issue 1, Springer, 2015, pp. 89-95.

[18] Ryba, T **Porovnání automaticky inicializovaných segmentačních metod**, *SVK 2015 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2015, pp. 109-110.

[19] Ryba, T and Železný, M **An automatic initialization of interactive segmentation methods using shortest path basins**, *Pattern Recognition and Image Analysis*, Volume 26, Issue 2, Springer, 2016, pp. 336-342.

[20] Jedlička, P and Ryba, T **Genetic algorithm application in image segmentation**, *Pattern Recognition and Image Analysis*, Volume 26, Issue 3, Springer, 2016, pp. 497-501.

[21] Ryba, T **Měření valgózního postavení dětských pat pomocí 3D skeneru**, *SVK 2016 - magisterské a doktorské studijní programy, sborník rozšířených abstraktů*, Plzeň, 2016, pp. 117-118.

Appendices

Appendix A

Lesion Localization

Results of lesion localization with *dwb* saliency map combination.
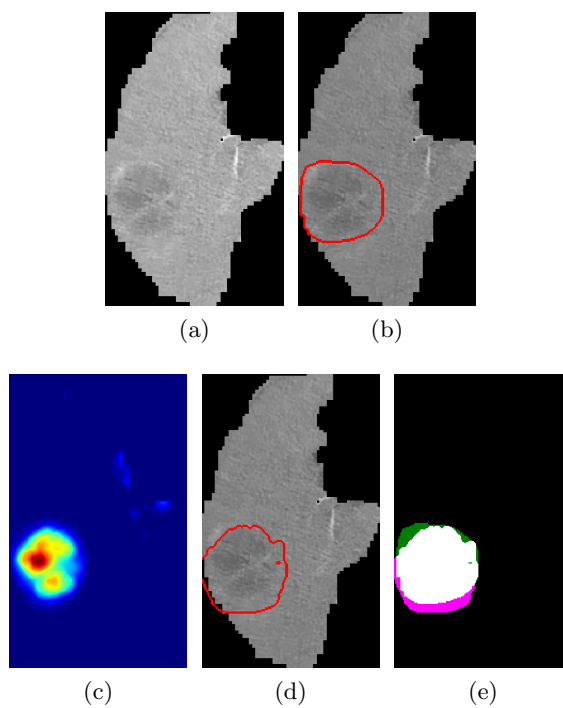
A.1 Example 1

Figure A.1: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).
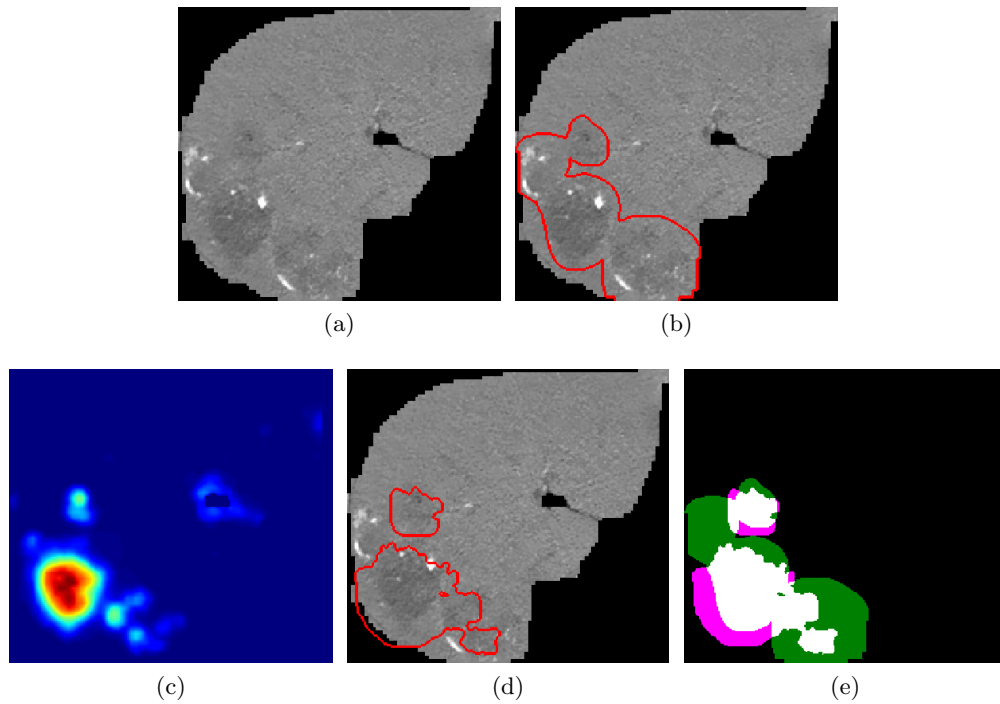
A.2　Example 2

Figure A.2: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).
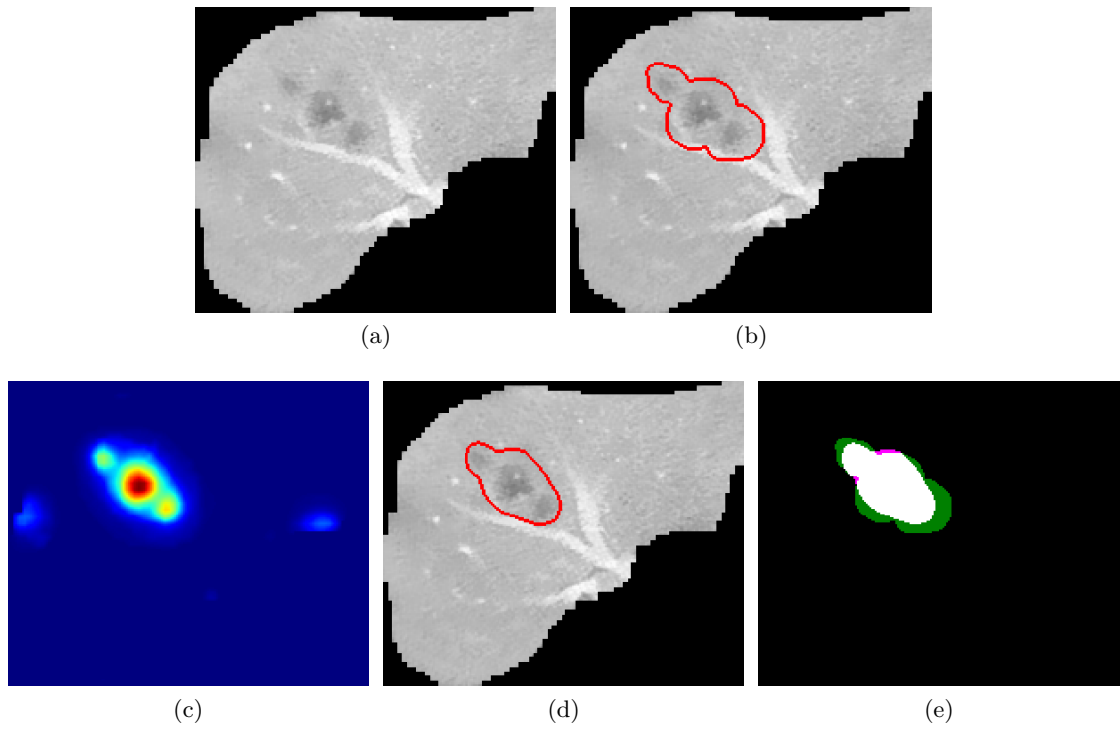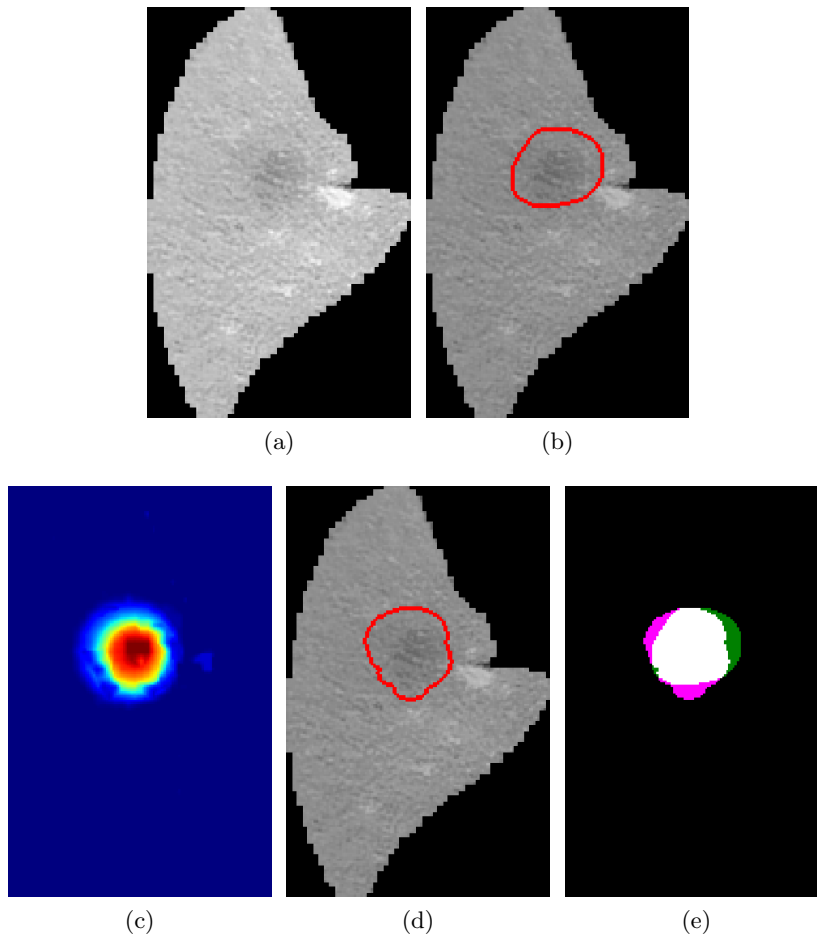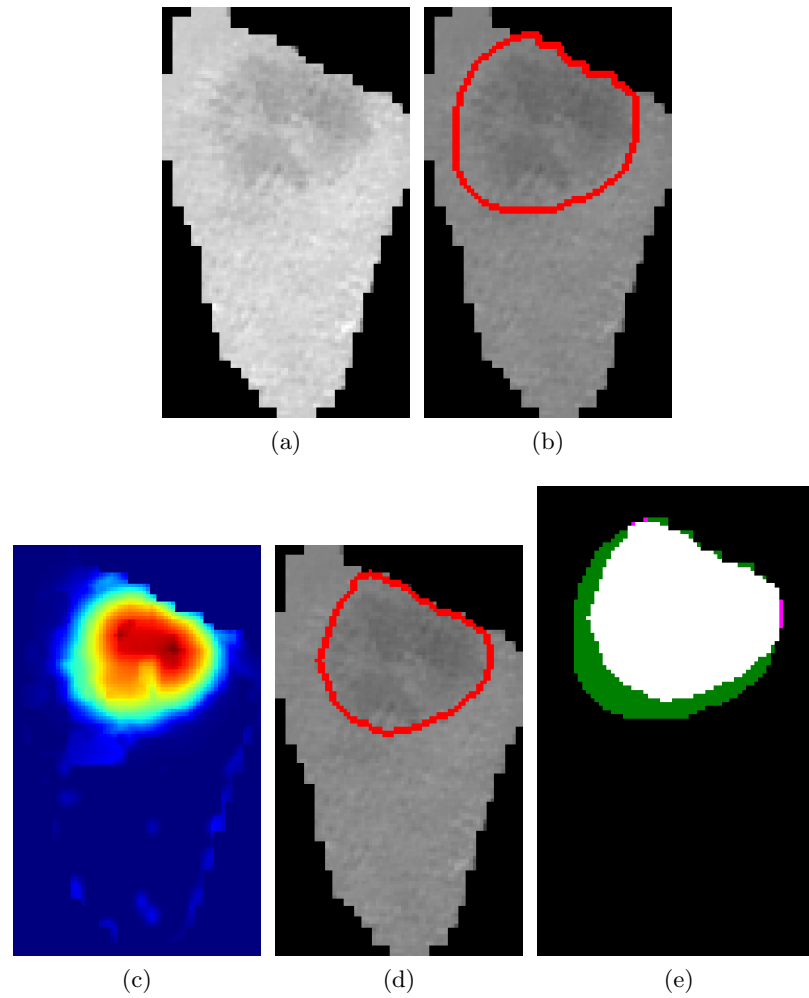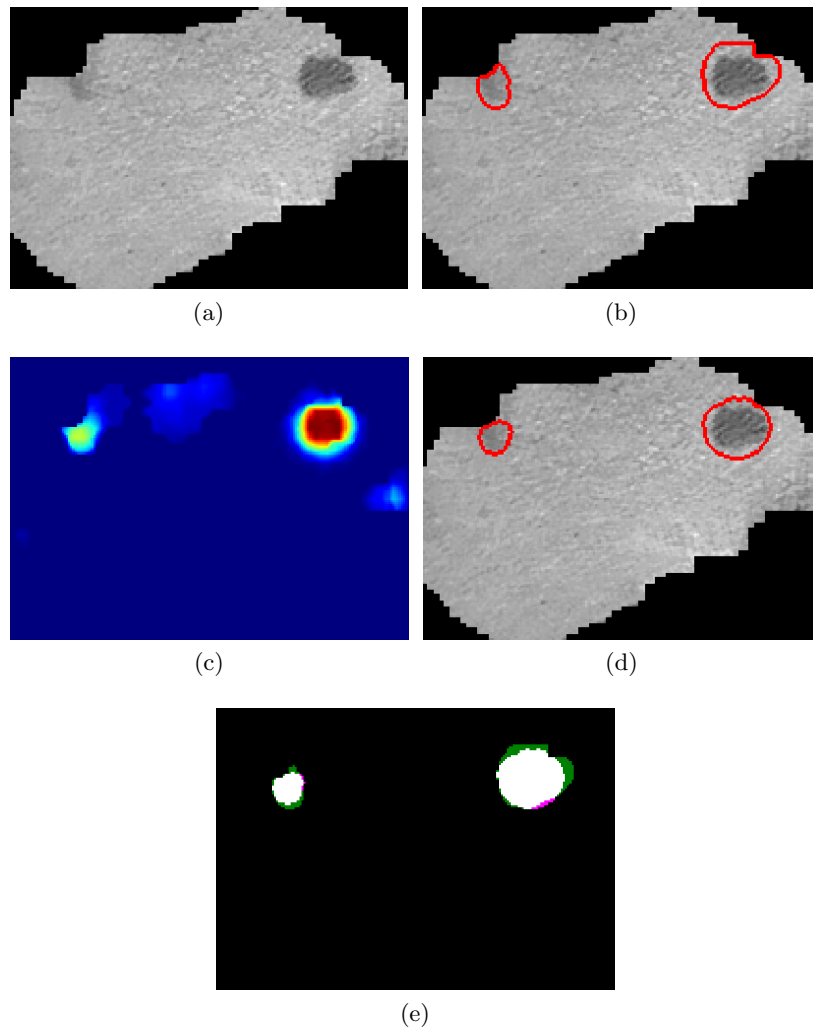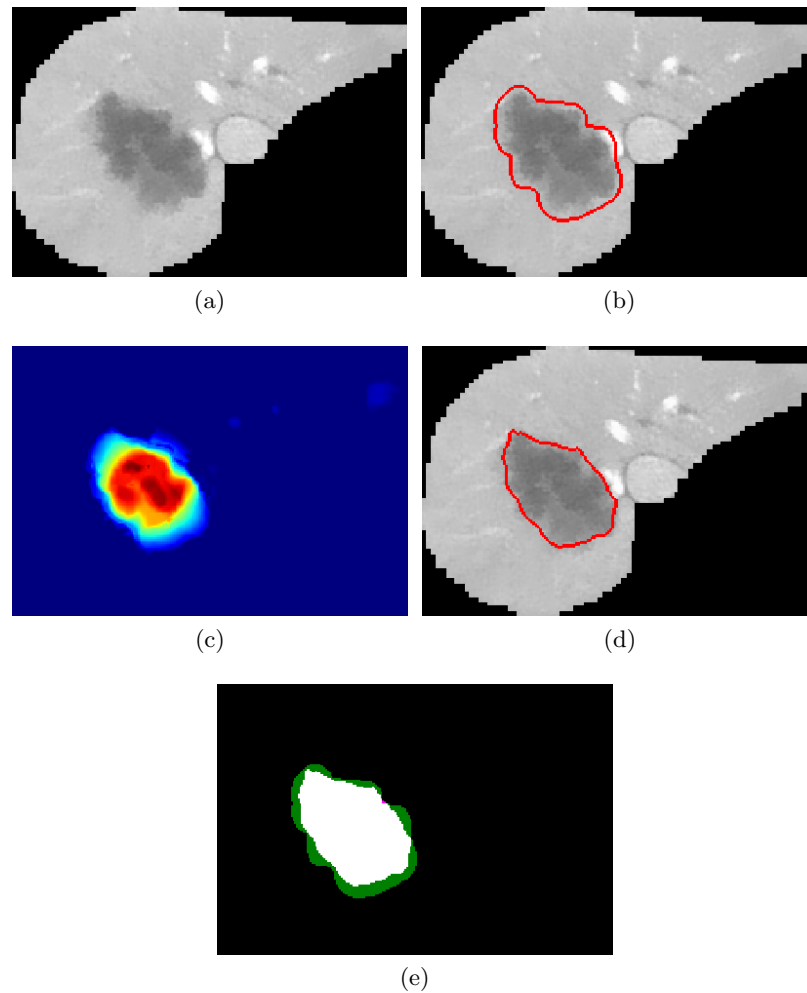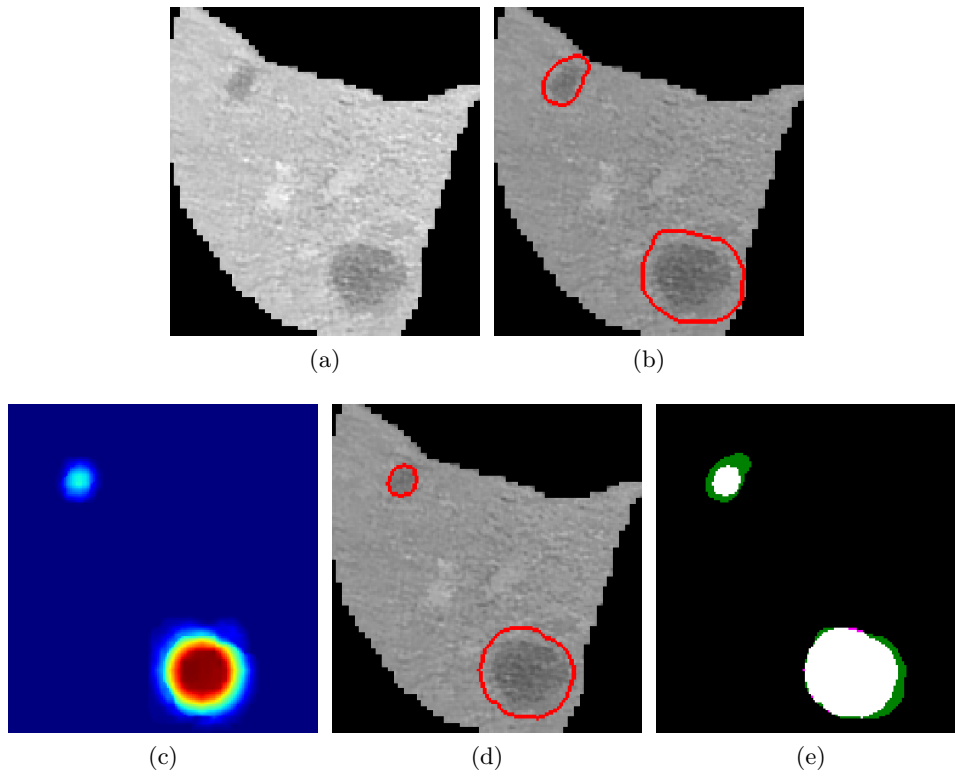
A.3   Example 3

Figure A.3: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).
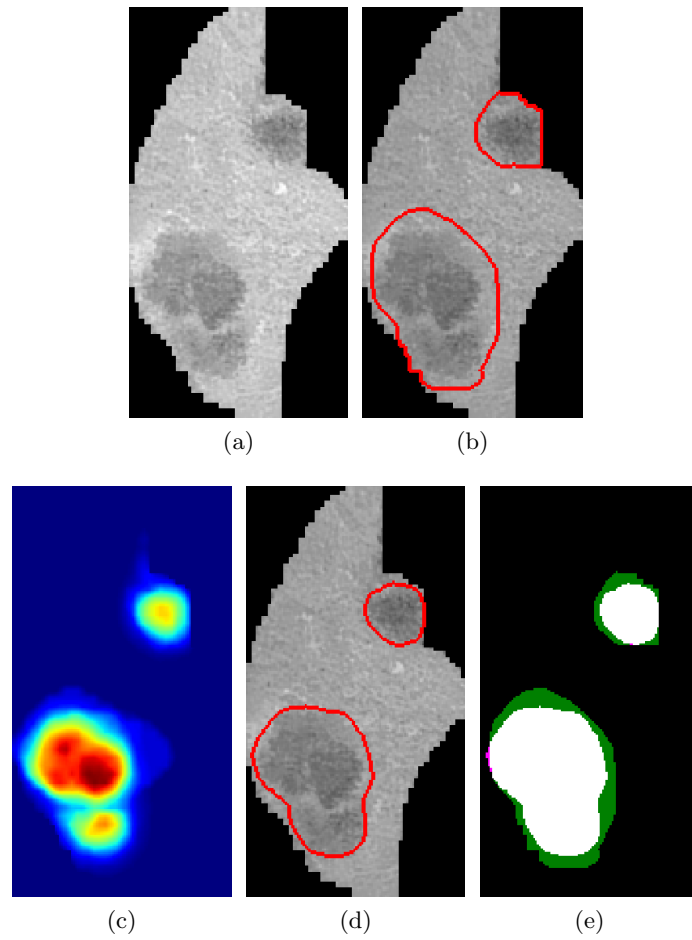
A.4   Example 4

Figure A.4: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).

A.5 Example 5

Figure A.5: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).

A.6   Example 6

(a)

(b)

(c)

(d)

(e)

Figure A.6: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).

A.7 Example 7

Figure A.7: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).
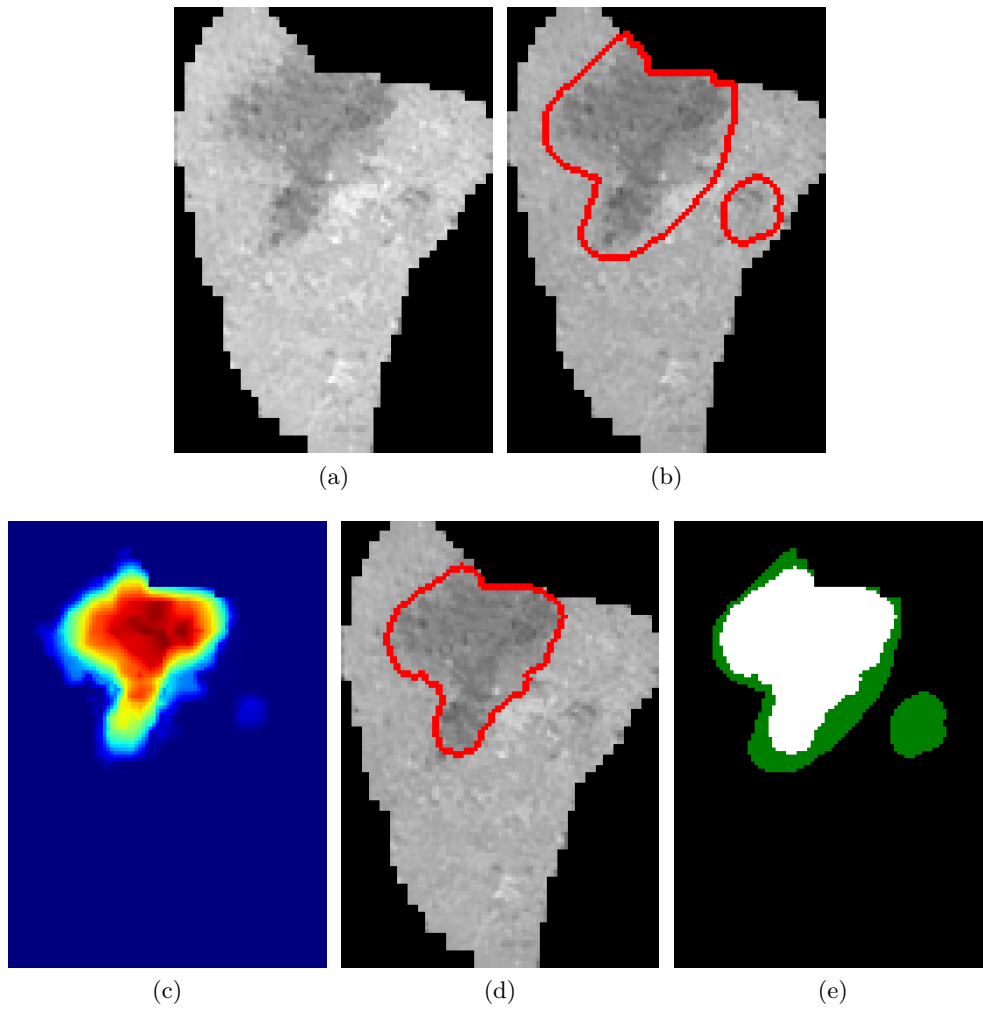
A.8   Example 8

Figure A.8: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).

A.9 Example 9

Figure A.9: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).

A.10   Example 10

(a)   (b)

(c)   (d)   (e)

Figure A.10: Input data (a), ground-truth (b), *dwb* saliency map (c), final segmentation (d), overlay with ground-truth (e).

Appendix B

Lesion Classification

This appendix contains several results of tumor classification.

B.1 Cyst

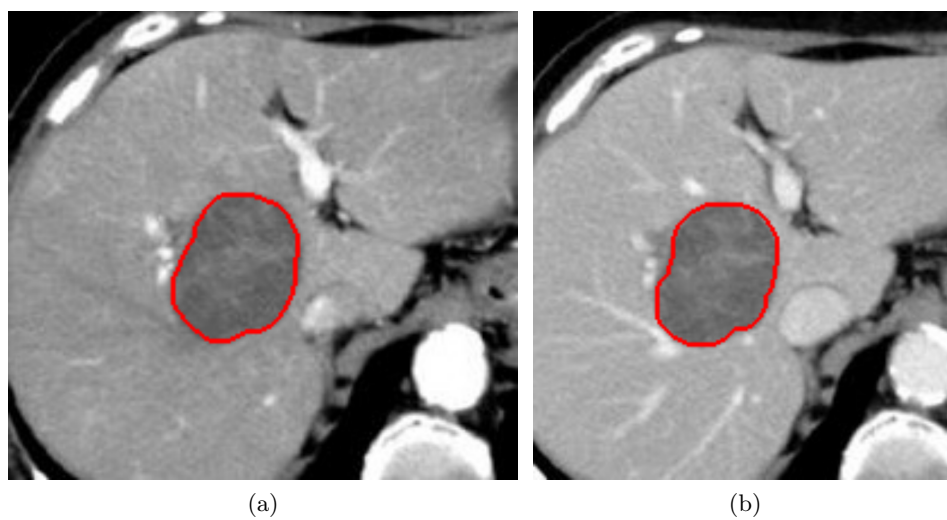A cyst with its typical visualization in both phases is shown in Fig. B.1. The cyst was classified correctly.

<center>(a)   (b)</center>

Figure B.1: The cyst visualized in the arterial (a) and the portal phase (b).

B.2   Focal Nodular Hyperplasia

A characteristic feature of the FNH is its central hypodense scar, see Fig. B.2. If the FNH is small and the scar occupies relatively big area, then the FNH can be missclassified as the hypovascular meta. This FNH was classified correctly.

(a)

(b)

Figure B.2: A FNH in the arterial (a) and the portal phase (b).

B.3   Angiosarcoma

A typical HAS contains a hemorrhage and/or a calcification. These features often correspond
to a relatively small hyperdense area as shown in Fig. B.3. In the case of a small tumor, where
the features are relatively large, the HAS can be mistaken for a hypervascular metastasis.
The HAS in Fig. B.3 was classified correctly.

(a)

(b)

Figure B.3: A HAS in the arterial (a) and the portal phase (b).

B.4   Hepatocellular Adenoma

The HA is often a large tumor. The only problem can arise if the classified HA is small, in which case it can be labeled as small HCC. The HA shown in Fig. B.4 was classified correctly.

(a)

(b)

Figure B.4: A HA in the arterial (a) and the portal phase (b).

B.5 Hemangioma

A HG is very similar to a cyst. The difference between them is the compactness - the cyst is much more compact than the HG. The HG can be labeled as the HAS as well, if the features of the HAS (hemorrhage, calcification) are not presented inside the tumorous tissue.
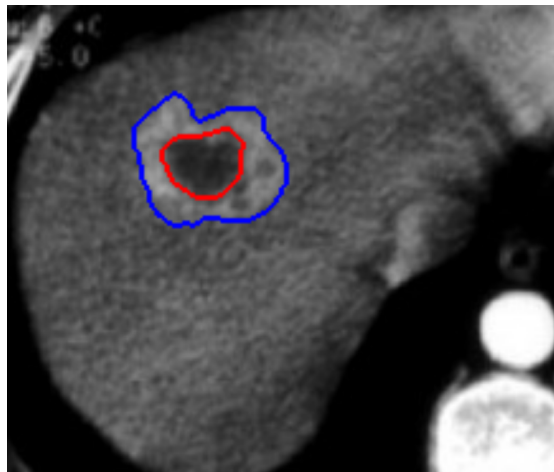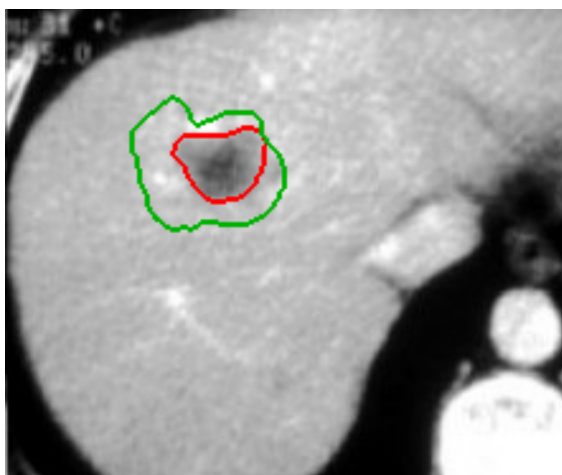

(a)

(b)

Figure B.5: A HG in the arterial (a) and the portal phase (b).

B.6   Metastasis

A hypovascular metastasis with clearly visible hyperdense capsule in the arterial phase (Fig. B.6(a)). The capsule is not visible in the portal phase (Fig. B.6(b)); therefore, the mask of the capsule from the arterial phase is used. This leads to the correct classification of the metastasis.
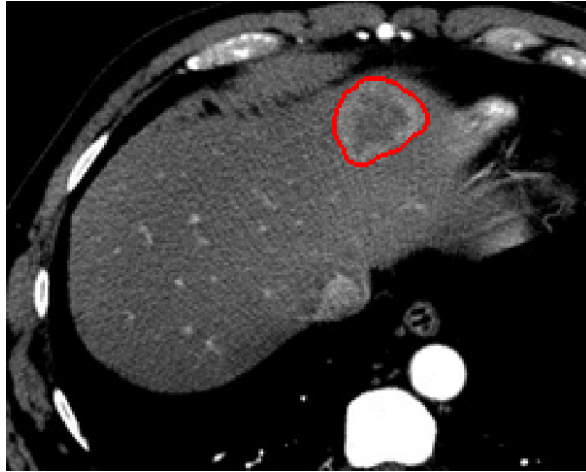

(a)

(b)

Figure B.6: A hypovascular metastasis in the arterial (a) and the portal phase (b).

B.7   Cholangiocellular Carcinoma

An atypical ICC that appears like a tumor with a hyperdense capsule. The algorithm wrongly classified this ICC as a hypovascular metastasis.

(a)

(b)

Figure B.7: An ICC in the arterial (a) and the portal phase (b).

B.8   Hepatocellular Carcinoma

The HCC shown in Fig. B.8 appears isodense in the portal venous phase with hypodense structures. Therefore, the algorithm uses the mask segmented in the arterial phase. Unfortunately, the HCC does not have a capsule and thus was wrongly classified as the FLC. A correctly classified HCC with clearly visible capsule is shown in Fig. B.9.
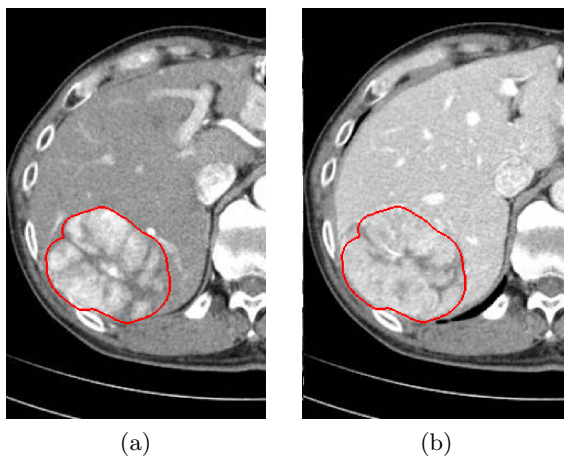

(a)   (b)

Figure B.8: A HCC in the arterial (a) and the portal phase (b).
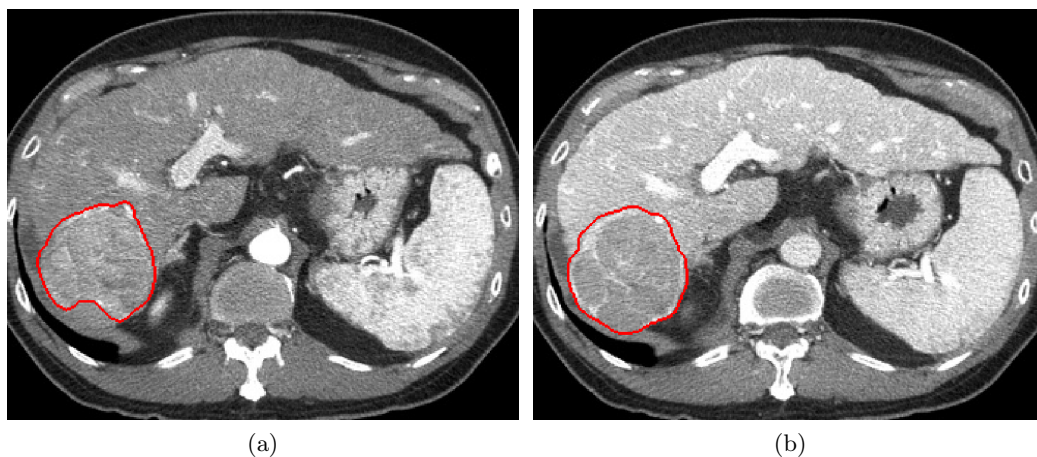

(a)   (b)

Figure B.9: A HCC in the arterial (a) and the portal phase (b).