UCL Université catholique de Louvain

École polytechnique de Louvain (EPL)

epl

Automated Lung Tumor Detection in Fluoroscopic Images for Breathing Motion Mitigation in Proton Therapy

Dissertation presented by
Anne-Sophie COLLIN

for obtaining the Master's degree in
Biomedical Engineering

Supervisor(s)
John LEE, Jose SEABRA

Reader(s)
Rudi LABARBE, Edmond STERPIN

Academic year 2017-2018

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Abstract

École Polytechnique de Louvain

Master in Biomedical Engineering

Automated Lung Tumor Detection in Fluoroscopic Images for Breathing Motion Mitigation in Proton Therapy

by Anne-Sophie COLLIN

Tumor motion is one of the major issues of radiotherapy that can lead to discrepancies between planned dose and dose absorbed by the patient. Due to the proton beam physics, proton therapy is particularly sensitive to this source of geometrical uncertainties. In the specific case of lung cancer, tumor motions induced by breathing can be substantial, which greatly complicate the treatment by proton therapy.
However, motion mitigation techniques offer the opportunity to deal with intrafractional motions and ensure conformity in the delivered dose distribution profile. Two of these techniques, beam gating and beam tracking, rely on a real-time monitoring of the tumor. For this purpose, dedicated image processing algorithms must be designed in order to detect lung tumors in images acquired during the treatment.

The purpose of this thesis is to investigate methods for lung tumor detection in fluoroscopic images in order to mitigate tumor motion in proton therapy. Here, detection has to be taken in the broad sense of the term and, depending on the strategy used to mitigate motion, information about either the contour or the location of the tumor center can be computed. Both segmentation and tracking methods are then under scope of this work.
A segmentation tool based on graph cut as well as a kernel mean-shift tracking method are investigated. These two approaches have been tested on digitally reconstructed radiographs after application of a contrast enhancement method implementing the contrast limited adaptive histogram equalization algorithm.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors, John Lee and Jose Seabra, for their advices and guidance throughout the year. Their deep interest in the topic was a great source of motivation.
I am extremely grateful for the time and effort they have dedicated to proofreading this thesis. Their constructive comments have helped greatly to improve the quality of the present document.

Then, a special thank to Kevin Souris for supplying the data on which this work was based. His help in the construction of a workable and consistent data set was precious. I would also like to thank Damien Dasnoy-Sumell for sharing his technical knowledge about the subject.

I would like to warmly thank Marc Green for his work during the proofreading of this document.

Finally, I am thankful to my friends and family for all their little attentions that were of a great help.

Contents

List of Figures

List of Tables

List of Abbreviations

4DCT	4-**D**imensional **C**omputed **T**omography
AHE	**A**daptive **H**istogram **E**qualization
CLAHE	**C**ontrast **L**imited **A**daptive **H**istogram **E**qualization
CNN	**C**onvolutional **N**eural **N**etwork
CPAP	**C**ontinuous **P**ositive **A**irway **P**ressure
CT	**C**omputed **T**omography
CTV	**C**linical **T**arget **V**olume
Dice	**Dice**'s similarity coefficient
DoG	**D**ifference of **G**aussian
DRR	**D**igitally **R**econstructed **R**adiograph
fps	**f**rame **p**er **s**econd
GTV	**G**ross **T**umor **V**olume
HE	**H**istogram **E**qualization
HU	**H**onsfield **U**nit
LFR	**L**ow-**F**rame-**R**ate
MRI	**M**agnetic **R**esonance **I**maging
OAR	**O**rgan **A**t **R**isk
pdf	**p**robability **d**ensity **f**unction
PTV	**P**lanning **T**arget **V**olume
RGB	**R**ed **G**reen **B**lue color model
ROI	**R**egion of **I**nterest
SNR	**S**ignal-to-**N**oise **R**atio
SOBP	**S**pread-**O**ut **B**ragg **P**eak
SPR	**S**topping **P**ower **R**atio

Chapter 1

Introduction

Cancer is a worldwide burden being one of the major causes of death. In 2000, almost 10 million people were diagnosed with cancer but approximately 22 million are estimated to suffer from the disease. In the same year, the number of deaths caused by cancer reached 6.2 million. According to the World Cancer Report 2014 [1], the number of people who will develop a cancer will be as high as 15 million by 2020, thus corresponding to an increase of 50% in 20 years.

Cancer treatment can be carried out by several modalities. Dominant means of treatment are surgery, chemotherapy and radiotherapy, which can be given alone or in combination. In particular, radiotherapy is involved in almost 50% of the treatments, both for curative and palliative intent.
The principle of radiotherapy is based on the way radiation affects living cells. Interaction of the ionizing radiations with the tissues results either in mutation or complete functional disruption of the targeted cells. Radiotherapy takes advantage of the lower DNA repair capabilities of cancerous cells in comparison to normal cells. This effect is all the more significant when the total dose delivered is divided into several fractions, i.e. independent sessions with lower-dose exposition [2]. This offers an efficient solution for increasing the damage caused to cancerous cells while giving time to the healthy tissues to repair from DNA damage caused by the radiation.

The aim of radiotherapy is damaging as much as possible cancerous cells while sparing surrounding healthy tissues. For this purpose, proton therapy offers many advantages over the use of conventional photon therapy [3]. Unlike photon beams which deposit a lot of energy at the entrance and then progressively decrease dose delivery, proton beams deposit most of their energy at the distal end of the trajectory. This part of the beam, where most of dose is absorbed by tissues, is known as the Bragg peak.
Figure 1.1 shows photon and proton depth-dose profiles. The Bragg peak is visible on the mono-energetic proton beam curve (red). Due to the narrow shape of the distal part of the depth-dose curve, only a small area of tissues can be treated by a single mono-energetic proton beam. In order to widen the treatment area, multi-energetic proton beams (blue) can been obtained by weighting several mono-energetic proton beams. The resulting sum of individual Bragg-peaks is called Spread-Out Bragg Peak (SOBP). The whole tumor can be completely exposed to radiation if optimal weights of individual Bragg-peaks are employed, while an homogeneous 3-dimensional dose distribution is achieved by using different irradiation angles.

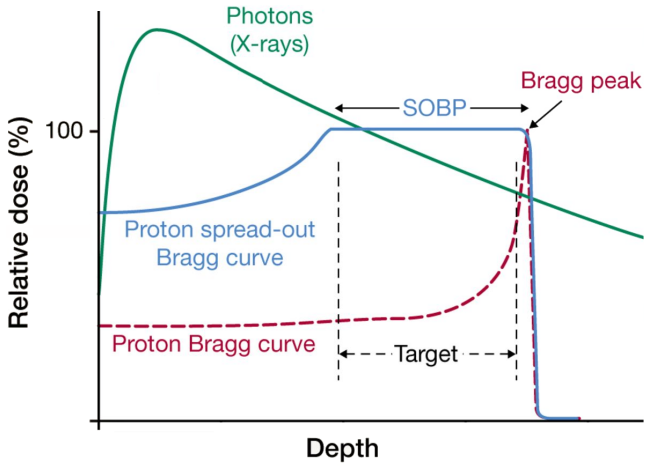


FIGURE 1.1: Depth-dose profiles for photon and proton beams (based on Mohan et al. [4]).

Ideally, the tumor should receive 100% of the prescribed dose and no dose shall be given to healthy tissues [5]. Practically, this optimal solution is not achievable and trade-offs have to be made in order to create an acceptable treatment plan. This patient specific process relies, usually, on planning Computed Tomography (CT) scans to map the patient anatomy into stopping power ratios (SPRs). These SPRs aim to describe the range traveled by a beam in a given medium with respect to the range in water. On the basis of SPRs estimation for a given patient anatomy, dedicated optimization tools approximate the dose absorbed by the tissues (dosimetry study) and find the best treatment parameters.

Particle range evaluation is a critical step in radiotherapy to assess the quality of the dosimetry study [3]. Since range uncertainties can lead to discrepancies between planned and delivered dose, they should be integrated in the planning phase in order to ensure robustness of the treatment.
Many factors cause uncertainties in range evaluation. For example, some range uncertainties already appear during the treatment planning step due to the incapacity of the imaging process to capture exactly the patient anatomy. Other uncertainties occur during the conversion of the CT pixels intensities expressed in Honsfied Units (HUs) to SPRs. By widening the irradiated zone, which corresponds to the Planning Target Volume (PTV), these uncertainties can be mitigated. PTV is a zone embracing the Clinical Target Volume (CTV) defined as the Gross Tumour Volume (GTV), i.e. volume emcompassing the tumor cells, extended by some margins incorporating surrounding tissues that could potientially contains tumor cells. GTV margins are visible on imaging while CTV are imaginary boundaries.

Range uncertainties affect all types of radiotherapy but are particularly crucial in proton therapy. Due to the proton beam physics, proton range is particularly dependent on the material the beam is traveling trough. A bad evaluation of the anatomy crossing the beam path leads to a bad evaluation of proton range, thus affecting directly the robustness of the treatment. Even an uncertainty a few millimeters can lead to overdosage of surrounding tissues, which can include an Organ At Risk (OAR), and underdosage of the tumor [5]. Note that for lung tumors, where the density of tissues is heterogeneous, range uncertainties are all the more harmful.

Motion mitigation strategies are of outmost importance in proton therapy where there is a clear interest in reducing the PTV margins as much as possible to benefit from the optimized local energy deposition achieved with protons. Therefore, a simple solution such as the enlargement of PTV margins to compensate for tumor motions negates benefits of proton therapy. Both interfractional and intrafractional motions, which are respectively motions between distinct fractions and within the same fraction, account for these uncertainties. Interfractional motions are well handled thanks to patient immobilization and image-guided registration. Intrafractional motions mitigation is more challenging and depends on the proton delivery mode [6].
With passive beam shaping, a flat dose profile can be obtained irradiating the whole tumor volume quasi instantaneously. In the one hand, the process involves range modulator that allows the beam to vary its energy in a short period of time. This has an effect on the longitudinal direction of the beam. On the other hand, collimator is used to scatter the narrow beam produced by the particle accelerator device. This has an effect on the lateral axis of the beam. Finally, specific range compensator device modifies the shape of the beam to adjust it to the tumor. If a moving tumor is considered as stationary within the fraction, lateral geometrical misses occur when the tumor is not at the predicted position.
With active beam scanning approach, a narrow beam irradiates the tumor point by point. Advantage of this approach, compared to passive beam shaping, is that it does not require any patient specific device. But if a moving tumor is considered as stationary within the fraction, important interplay effect, described in Figure 1.2, is added to range uncertainties due to anatomical variations [7]. In this figure, we observe that the scanning is performed through an isoenergetic slice in the stationary coordinates denoted by the initial tumor location (orange). During radiation delivery, a narrow dose is deposited at a single spot denoted by a red dot. These red dots are represented in moving coordinates. Color of the border of the red dots corresponds to the moving coordinates at which the dose was delivered (blue, orange and green) that depends on the respiratory phase shown in the diagram in the bottom of the figure. Right picture shows the final result in the stationary coordinates where irradiated spots are denoted by red dots.
Motion has then a bigger impact on active beam scanning than on passive beam shaping [8] because the interaction between the beam delivery dynamics and tumor motion need to handled.

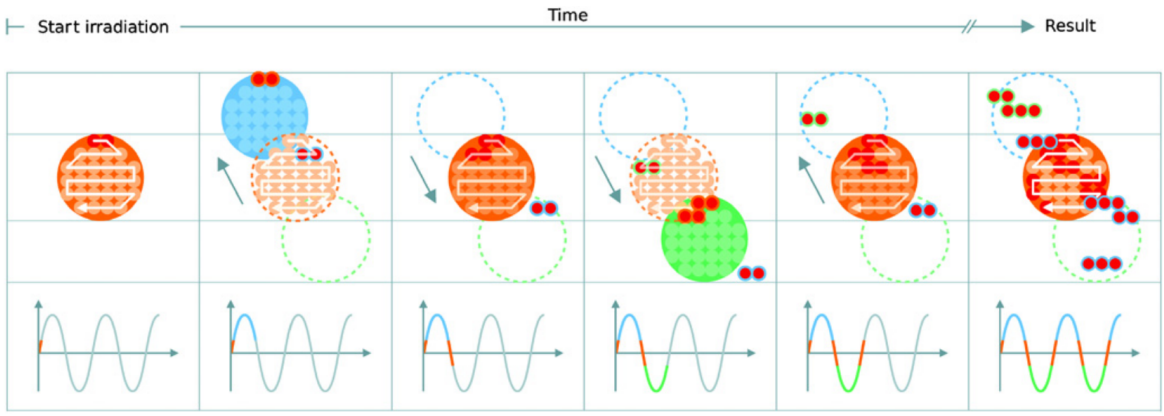


FIGURE 1.2: Interplay effect for a moving tumor treated with pencil beam scanning (from Bert, Grözinger and Rietzel [8]).

For both passive and active beam delivery modes, motion mitigation techniques are essential to deal with intrafractional motions and ensure conformity in the delivered dose distribution profile. This phenomenon is of outmost importance for active beam scanning. In this case, interplay effect has to be added to uncertainties due to anatomical variations. Solutions addressing uncertainties for active beam scanning but are then also applicable for passive beam shaping.
In the specific case of lung tumors, which is one of the most common cancers and accounts for 12.3%, breathing induced intrafractional motions can be substantial. Therefore, specific methods have to be designed to offer the possibility of delivering proton therapy. Up to now, three major approaches allow to mitigate lung tumor motions:

Rescanning. This first approach does not mitigate motion directly. Instead, it relies on the dependency between over- and underdosage and motion factors such as its period or phase. The idea is to average, from a statistical point of view, these over- and underdosage by scanning over the entire area covered by the tumor motions, i.e. Internal Target Volume (ITV). Providing that tumor motion and scanning period are temporally uncorrelated, the over- and underdosage areas are averaged over the whole tumor volume. If it is not the case, over- and underdosage averaging is biased.

Beam gating. This second method aims at mitigating tumor motions induced by breathing by synchronizing the beam-on/beam-off periods with respiratory motion, in breath hold conditions or not. The choice of the beam-on/beam-off periods is not trivial and should be carefully taken based on informative morphological parameters. For example, tumor-volume to lung-volume ratio and tumor-to-spinal cord distance were used in [9] to determine optimal beam-on/beam-off periods. However, beam-on period often includes several motion states leading to a residual interplay effect.

Beam tracking. This third approach consists in real-time tracking of the tumor allowing continuous irradiation of the tumor during the fraction. This method is technically more challenging compared to the two others. It relies on an accurate and flexible control on the beam energy and delivery and and requires a very fast processing time from image acquisition to the update of the beam delivery system.

Rescanning differs from the two other methods in that it requires no synchronization with tumor motion. Since the entire area covered by the tumor (ITV) can be defined offline, no tumor motion characterization or breathing monitoring are needed during the treatment delivery phase. In contrast, both beam gating and beam tracking are synchronized with tumor motion and require an accurate knowledge of the tumor position and morphology at each moment of the respiratory cycle. Tumor position and shape assessed with imaging techniques are important parameters to quantify how the tumor deviates with respect to the planning position, and hence optimize the dose delivered.

Used alone or in combination, the choice of the motion mitigation strategy is patient specific and should be carefully taken. For lung tumors exhibiting small motion range (< 5mm), rescanning is a good option. This solution has the advantage of being the simplest in nature and not extending the duration of the treatment session. Therefore, rescanning is an acceptable trade-off between the enhancement of dose conformity and the relative complexity carried by the implementation of a motion mitigation technique. Also, this could be the preferred option for patient showing irregular breathing patterns because it does not rely on the real-time monitoring of the tumor motions.

For tumor with larger motion range ($>$ 5mm), the gain of using motion mitigation techniques, such as beam gating or tracking, is all the more substantial. In these cases, the volume of surrounding healthy tissues that can be spared is more important justifying the implementation of a more complex motion mitigation technique.

Both for the choice of motion mitigation strategy and the implementation of beam gating and tracking, a characterization of the tumor motion acquired in a daily setup scheme is necessary. This thesis investigates strategies for tumor delineation and localization in fluoroscopic images, comprising an image contrast improvement for better visualization of the tumor and tumor shape delineation.

1.1 Organization of the Thesis

The remaining of the thesis is divided into six chapters that highlight different aspects of lung tumor detection in fluoroscopic images :

Chapter 2. This chapter describes the context of the problem. First, the equipment required for fluoroscopy-guided proton therapy is presented. Then, the core problem of the thesis, the segmentation and the tracking of lung tumors, is introduced. Nature of fluoroscopic images is discussed in Section 2.2 which highlights the need of some preprocessing of the images. Finally, materials used for this work are presented at the end of the chapter.

Chapter 3. In this third chapter, a review about ongoing methods referenced in the literature is presented. The structure of this chapter is the same as the thesis. First, methods about fluoroscopic image quality enhancement used as preprocessing are developed. Second, a general overview about segmentation methods in all types of images is proposed. Thirdly, this chapters ends up with a presentation of tracking algorithms applied specifically in fluoroscopic images.

Chapter 4. Purpose of this chapter is to present a fully automated fluoroscopic image enhancement algorithm used as preprocessing. This method combines a difference of Gaussian filtering followed by the application of the contrast limited adaptive histogram equalization technique. By reducing noise and increasing contrast, the proposed method aims to enhance the lung tumor visibility in the processed images. This step should bring some improvement for the application of segmentation and tracking algorithms in the next two chapters.

Chapter 5. This chapter addresses the difficult problem of segmenting lung tumors in fluoroscopic images. The approach chosen to achieve this task is the graph cut method. Our contribution is the modification of the initialization procedure involving an expert into the integration of a binary mask drawn with some prior knowledge about the tumor position and shape. Parameters of the method and results obtained are discussed.

Chapter 6. This chapter proposes the kernel mean-shift algorithm to track lung tumors in fluoroscopic data. This method is tested on our data set and result hence achieved are compared with the literature and discussed.

Chapter 7. This last chapter summarizes main results obtained in the field of lung tumor detection in fluroscopic images. Directions for future work are presented at the end of the present document.

Chapter 2

Problem Formulation

2.1 Technical Requirements for Image-Guided Proton Therapy

As explained in the introduction, tumor mitigation techniques rely on a characterization of tumor motions to choose the optimal motion mitigation strategy for a given patient. Moreover, real-time acquisition and processing of images is also required to mitigate breathing motion with either beam gating or beam tracking approach. For this purpose, imaging devices need to be integrated into the proton therapy gantry, which is the rotating device placed around the patient that delivers the proton beam.

Recent studies [10, 11] strongly suggest the feasibility of combining real-time Magnetic Resonance Imaging (MRI) technique with proton therapy. The current clinical applications of MRI-guided photon therapy push forward the demand of such techniques for proton therapy.
Magnetic interaction between the MRI equipment and the proton beam represents a huge technical challenge. Some software and hardware aspects have to be considered. On the one hand, effects of magnetic field induced by the MRI device on the beam propagation need to be taken into account during the treatment planning and the beam delivery phases. On the other hand, taking into account the MRI magnet design and the rotating proton gantry is a complex task requiring eventually some change of paradigms. A mobile couch in favor of a rotating gantry is an example. Due to all the technical challenges mentioned previously, MRI guided proton therapy still requires some years of research and development before being clinically available.

In the context of this thesis, focus is on another image modality: the X-ray fluoroscopy. In this technique, one or more fluoroscopic flat panel detectors and X-ray tubes are mounted in the treatment gantry and enable to continuously capture 2-dimensional projections of the patient anatomy. Such installation is represented in Figure 2.1. By the use of two orthogonal X-ray sources placed in front of two detectors, the tumor can be located in a 3-dimensional space. The system can also be used to generate Cone-Beam Computed Tomography (CBCT), which are useful to reduce interfractional errors.
Acquisition of fluoroscopic images requires the generation of X-rays which constitute an additional dose delivered to the patient. However, even for real-time applications, this additional is relatively low making the system appropriate for image-guided proton therapy use [12]. This technology has already been implemented and produces promising outcome [13, 14].

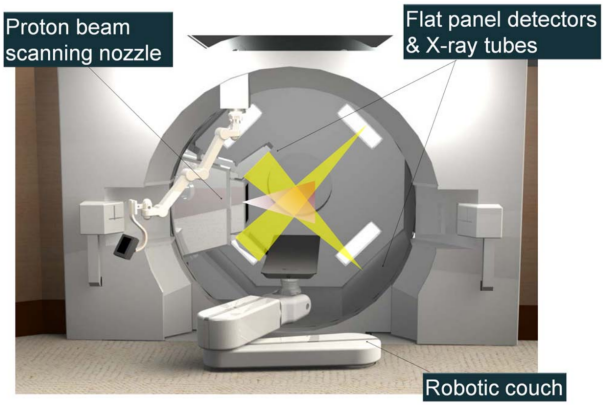


FIGURE 2.1: Proton therapy system with integrated X-ray fluoroscopy device developed at Hokkaido University (from Shimizu et al. [12]).

2.2 Purpose

Objective of the thesis is to address the problem of lung tumor detection in fluoroscopic images. Here, detection has to be taken in the broad sense of the term. Both segmentation and tracking algorithms are under the scope of the problem. On the one hand, segmentation methods aim to delineate boundary region of the tumor. On the other hand, tracking methods enable to locate the tumor in the image based on some a priori information about its appearance. In both cases, automated approaches, i.e. that do not require any user interaction, are preferred. However, before addressing these problems, quality of the fluoroscopic images is enhanced through a preprocessing step.

Quality of an image is usually determined by its resolution, its contrast and its Signal-to-Noise Ratio (SNR). Resolution is determined by the detectors and is not subject to any enhancement in the context of this work. Contrast between two regions is determined by the absolute value of the difference between mean intensities of the two regions. SNR is the ratio between the effective signal content of an image and the amount of noise.
In fluoroscopic imaging, contrast depends mainly on the voltage applied across the X-ray tube. With low voltage, difference between tumor and surrounding tissues photon attenuation is higher. This results in higher absolute difference between pixels intensities. But the drawback of lower voltage is more photons being absorbed by the patient increasing the radiation dose.
The number of photons reaching the detector has an impact on the SNR. For equal number of photons emitted per surface crossing an homogeneous material, the number of photons reaching the detector is subject to random distribution. This random distribution leads to noisy appearance of the image. Noise decreases quadratically with the number of emitted photons. One way to increase the number of emitted photons is to increase the current or the exposure time. For this purpose, exposure time can be extended but risk of artifacts due to motions is also higher. The other solution is to increase the tube current augmenting also the dose delivered to the patient.
Both voltage and current influence contrast and SNR of fluoroscopic images. Quality of the image is a trade-off with the dose delivered to the patient. Fluoroscopic images often exhibit very poor contrast and relatively high noise content. Therefore, a preprocessing of the image is required in order to increase tumor visibility and accuracy of segmentation/tracking methods.

2.3 Materials

Owing to the lack of data, we were not able to process fluoroscopic images acquired on patients with lung tumors. Instead, we simulated fluoroscopic acquisition out of 4-Dimensional Computed Tomography scans (4DCTs). By integrating over the pixels intensities (in HU) along given trajectories, the acquisition of an X-ray image can be simulated out of a 4DCT. This technique, named forward projection and illustrated in Figure 2.2, produces Digitally Reconstructed Radiographs (DRRs). Each DRR is a 2-dimensional image approximating an X-ray acquisition along a given angle. As presented in Section 2.1, two orthogonal sets of X-ray fluoroscopes are used in order localize the tumor in a 3-dimensional volume. In the context of this thesis, only DRRs simulated along the coronal plane are processed but the methods are applicable for any gantry angle.

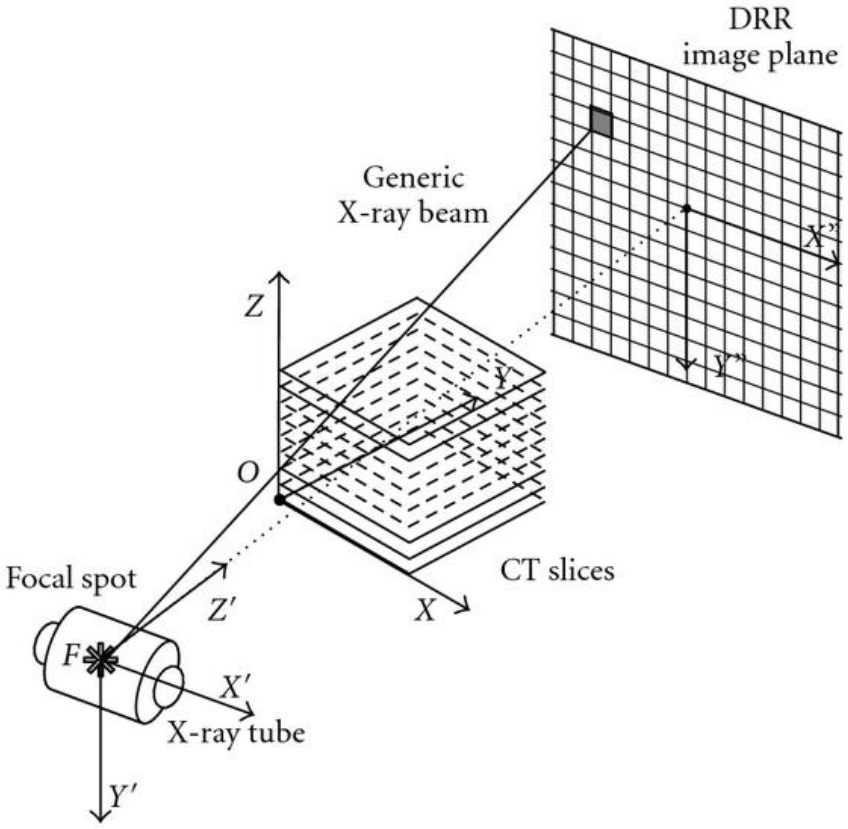


FIGURE 2.2: Integration of the pixel intensities along a simulated x-ray beam path in order to obtain the projection of the volume into the DRR image plane (from Bifulco et al. [15]).

DRRs are constructed with the `rtkforwardprojections` function of the `RTK` library [16] after converting 4DCTs and corresponding tumor contours from DICOM to RAW format with `OpenREGGUI` (`https://openreggui.org/`). All DRRs have a size of 3000x3000 pixels with gray values encoded in 16 bits.
From hereon in this thesis, whenever we refer to fluoroscopic data used to test the developed methods we are indeed talking about DRRs.

Data were provided for three patients. Each patient data set consists of four series of four 4DCTs with the 3-dimensional binary mask corresponding to the tumor contour on each of the ten scans. Last two series are acquired with one week interval with respect to the two first. Series 1 and 3 are 4DCTs acquired in normal breathing condition and series 2 and 4 with Continuous Positive Airway Pressure (CPAP). Summary of the data set is provided in Table 2.1.

TABLE 2.1: Data set constitution for one patient.

Week	Series Number	Breathing Condition
Week 1	Series 1	Normal Breathing
	Series 2	CPAP
Week 2	Series 3	Normal Breathing
	Series 4	CPAP

All the rest of the implementation was written in `PYTHON 3` and executed on an `INTEL` i5-4200M CPU 2.50 GHzx4 with 8 GB of RAM. Depending on the library, GPU execution of the code is sometimes available, which can speed up processes. Execution times provided further are given for information purposes only and should not be taken as definitive.

Chapter 3

State of the Art on Dedicated Image Processing Tools

Since the discovery of X-rays by Röntgen in 1895, medical imaging techniques play a prominent role in medicine. By allowing a representation of the anatomical or functional components inside the body in a non-invasive way, medical imaging is used for both diagnosis and treatment. Medical image processing involves image formation, visualization, analysis and management [17].

Firstly, Section 3.1 addresses methods enhancing fluoroscopic image visualization. It is important to keep in mind that improvement of visual appearance of an image is application dependent. For example, methods used to improve X-rays visualization are not the same as the those used to improve images captured by the Hubble telescope. The focus is maintained on clinical applications and especially on methods enhancing fluoroscopic or X-ray images.
Secondly, segmentation methods for medical purpose are discussed in Section 3.2. Image segmentation is one of the hotspots in computer vision. It refers to techniques allowing to partition an image into regions where all pixels share similarities such as color, gray level or texture [18]. Interest here is to distinguish pixels belonging to the tumor region from the rest of the DRR considered as background. Section 3.2 reviews multiple approaches to perform medical images segmentation. Nowadays, there is no research available on segmentation performed in fluoroscopic images. Several other image modalities are then concerned by this literature review.
Thirdly, Section 3.3 addresses methods to track moving tumors in fluoroscopic images. Usually, a tracking task is defined as the problem of estimating the trajectory of an object within the images of a scene, as while the object moves around the scene. Unlike a segmentation problem, the result does not provide the contour of the tumor but instead its relative position in the image. Tumor tracking in fluoroscopic images has been widely addressed in the literature. A review of the different recent methods are presented in this chapter.

3.1 Fluroscopic Image Quality Improvement

At equal resolution, contrast and noise are two major factors influencing image quality. In the context of fluoroscopic data acquisition, there is a trade-off between image quality and data acquisition parameters such as X-ray dose delivered to the patient, exposition time, resolution of the detector, etc. Low-dose X-ray or fluoroscopic images are known to exhibit relatively poor contrast, especially in high-voltage mode, and high noise level due to the statistical distribution of particles on the detector.

Even if the dynamic range of medical images is large (12, 16 or 32 bits) compared to natural images, many low-contrast details may not be perceptible. Therefore, preprocessing methods to improve contrast and reduce noise content are valuable. Indeed, the outcome of the segmentation and tracking phases dependents on the input image quality. In particular, segmentation tools are known to perform better with salient edges than with blurred ones. As sharp transitions lies in the higher frequency range of the spectrum and noise to most higher frequencies, trade-offs between noise removal and edge sharpness preservation need to be handled. Several approaches are available in the literature to address this problem. Such methods aim to improve both visual inspection of soft tissues and tumor (qualitative improvement) and automated detection of the tumor (quantitative improvement).

In [19], effort is made to improve quality of medical images. Their contribution resides in the use of guided filtering for medical images, and especially X-ray scans. Guided filtering refers to a methodology inspired from bilateral filtering introduced in [20]. As bilateral filtering, it makes use of a guided image, which can be the input image or another one, to adapt locally weights of the smoothing kernel depending on the gray level value of corresponding pixels in the guided image. Compared to classical filtering, which applies the same kernel over the entire image, guided filtering shows better edge preservation property. This means that homogeneous areas are smoothed to remove noise while edges are preserved. Computationally efficient, guided filtering suffers somewhat from the halos artifact, i.e. unwanted smoothing of edges, but less than other filtering methods. Unfortunately, the result is influenced by tiny textures from the guidance image. The global method of [19] combines by fusion, the output from guided filtering, a morphological filtering and a contrast stretching step. The overall method produces very good enhancement of the images but unfortunately its complexity makes it unappropriated for real-time processing.

Based on the hypothesis that the content of an image is likely to repeat itself within some neighborhood, a new patch based approach has been proposed by [21]. As for guided filtering, a guiding image is used to compute weights of the smoothing kernel. But the difference is that the weights of the smoothing kernel are computed non locally. They also integrated a noise model describing noise distribution in the specific case of X-ray images to determine some parameters of the kernel giving rise to the X-rays Non Local Means filter method. Finally, this non local kernel is used in combination with downsampling as operator to perform a multiscale decomposition [22]. The nature of the Laplacian Pyramid decomposition in coarser and finer scales makes the noise content larger in finer scales. With a weighted-wise reconstruction, great improvement of global image quality can be achieved. Unfortunately, in addition to being extremely computationally expensive, the method relies on user-defined parameters subject to intra-patient variability.

While the first two methods focused more on noise reduction, the approach described by Qiu et al. in [23] proposes a fully automated image enhancement method focused on the contrast enhancement. The first step consists in an enhancement of bony structure visibility thanks to a high pass filter implemented by the subtraction of the image and a smoothed version of itself. The second step is an application of the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm.
Histogram Equalization (HE) is probably the most popular method improving image contrast. It stretches the histogram, thus increasing the dynamic range of the image. During

this process, each initial pixel intensity is mapped into a new value depending on the intensity of the initial pixel only. The mapping can also be adaptive. An adaptive mapping is a non-stationary transformation for which the new intensity value does not depend only on the initial value but also on the intensities of its neighborhood or contextual region. In the fully adaptive approach, one mapping is computed independently for each pixel making the method time consuming. A compromise can be made between the global HE and the fully AHE by dividing the image into several sub-regions on which we apply HE. Without any correction, blocking artifacts would be observed in the border of each sub-region. This can be corrected with bilinear interpolation of the mappings of each sub-region.
In homogeneous areas, corresponding to large peaks in the histogram, both HE and AHE amplifies noise by mapping these peaks into a wide region in the resulting histogram. Therefore, this range over which the peak was spread can not be used to enhance contrast within other regions of the image. To overcome this limitation, Contrast Limited Adaptive Histogram Equalization (CLAHE) has been introduced in [24] and further summarized by Zuiderveld in [25]. By introducing a clip limit defining the maximum number of pixels that can belong to the same bin of the input histogram, large peaks are redistributed over the entire range of gray values. In this way, tall peaks are not mapped into a wide range in the output histogram and the spared range can be used to improve contrast within homogeneous regions. CLAHE can be efficiently implemented to suit to real-time tasks as discussed in [26]. Parameters of CLAHE can be automatically determined by entropy maximization of the result as described in [27, 23] or based on multi-objective metrics [28].

3.2 Segmentation Algorithms

A large variety of algorithms from standard image processing field can be applied to medical images in order to perform segmentation tasks [18, 29]. This comparative overview divides segmentation algorithms into several categories.

Both threshold-based and region-based methods aggregate pixels sharing similar properties such as gray level. Threshold-based methods process the histogram, instead of the image directly, loosing then pixel localization information. Region-based segmentation also relies on a threshold but with considering local relationships between pixels in order to build contiguous regions.
Boundary detection segmentation relies on differential operators to estimate image gradients and then locate edges. Watershed is an hybrid method between region-based and boundary detection approaches.
Instead of working directly in the spatial domain of the pixels, clustering-based segmentation allows to segment the image by clustering pixels projected in another feature space. All pixels that are close in this space are likely to be assigned to the same cluster. Each cluster denotes for a segmented zone of the image.
Atlas-based segmentation relies on the correspondence between a labeled reference image and a deformed version of the image to segment the image in a supervised scheme.
Three different model-based approaches are then presented. Both deformable models and graph-based models segment the image by minimizing an energy function. On the one hand, deformable models adapt a parametric curve whose evolution is driven by the energy function. On the other hand, graph-based approach computes minimal cut of a graph describing the image. Deep Learning segmentation makes use of a fully trainable layered model. Supervised training of the parameters relies on a large labeled database.

Threshold-based Segmentation

Pixels can be classified with respect to their intensity. Each class corresponds to pixels with similar gray levels. The decision relies on one or multiple threshold(s), that can be global or adaptive. For this category of segmentation methods, Otsu thresholding is probably one of the most famous. Based on the same idea as discriminant analysis, Otsu thresholding method classify pixels into clusters maximizing interclass variance or, equivalently, minimizing intraclass variance.
Simple and fast, this method can classify pixels based on their intensity provided that each cluster has uniform brightness. This method fails if some gray values overlap between clusters (if they are n clusters, histogram should have a n-mode) and is very sensitive to noise. However, combined with morphological postprocessing, thresholding methods have been applied to perform lung segmentation [30, 31].

Region-based Segmentation

As for threshold-based approach, region-based segmentation forms clusters of pixels sharing similar properties. On the one hand, the approach can follow either a bottom-up (region merging) or a top-down (region slitting) scheme. The idea is to recursively merge/split parts of the image that share similar/dissimilar properties. The method can rely on an initialization. In this case, some initial pixels are labeled as foreground. Progressively, the foreground region is extended by combining neighboring pixels of the current foreground region that are considered as similar.
These approaches are more robust to noise than threshold- based methods but they also rely on the homogeneity of regions criterion. As for threshold-based segmentation, a pixel is assigned to one class if it is considered as similar to the class model below a given threshold. The threshold value is not easy to determine and its impact on the result is substantial. In order to get rid of this parameter, a fuzzy approach can be used. Instead of classifying each pixel into one unique cluster, a membership probability to each class can be computed. This approach has been used in [32] to segment pulmonary nodules in CT scans.

Boundary-based Segmentation

Unlike the first two methods which use similarity between pixels to classify them, segmentation methods based on boundary detection use dissimilarity between neighboring regions to perform the segmentation task. More precisely, a discontinuity between the gray values of two distinct regions is called an edge. In a 2-dimensional image, an edge or a boundary can also be defined as the set of points maximizing the first-order derivative of the pixels intensities. Therefore, first-order differential operators such as Sobel, Prewittt or Roberts or second order operators like the Laplacian operator can be used to detect edges along one direction.
Another standard technique to detect boundaries is known as Canny edge detection. After a Gaussian filtering to reduce noise, a first order operator is applied on the image to estimate gradients. All maxima which do not include other neighborhood maxima in the orthogonal direction of their gradient are discarded. Finally, all edges with intensity gradient lower than a threshold are also discarded. In [33], a Canny edge detector is used to improve segmentation performed with region growing in brain MRI images.

Watershed

Watershed is also a popular segmentation algorithm which combines both aspects of region-based and boundary-based segmentation. It can therefore be considered as an hybrid technique. The method is usually explained thanks to the increasing water level metaphor illustrated in Figure 3.1. In this figure, Black curve, determined by pixel intensities, forms an uneven landscape progressively filled with water then forming lakes. As water rise, two lakes can be joined in the water junction area. Edges correspond to all lake sides and water junctions.
Usually, results obtained with watershed are oversegmented due to noise and irregular textures sensitivity of the method. In [34], watershed was used in order to segment breast tumors in 2-dimensional sonography. Authors were able to reduce sightly the oversegmention by correcting the result with some prior information about the gray level of the tumor. However, the enhancement is highly dependent on the quality of the prior information.

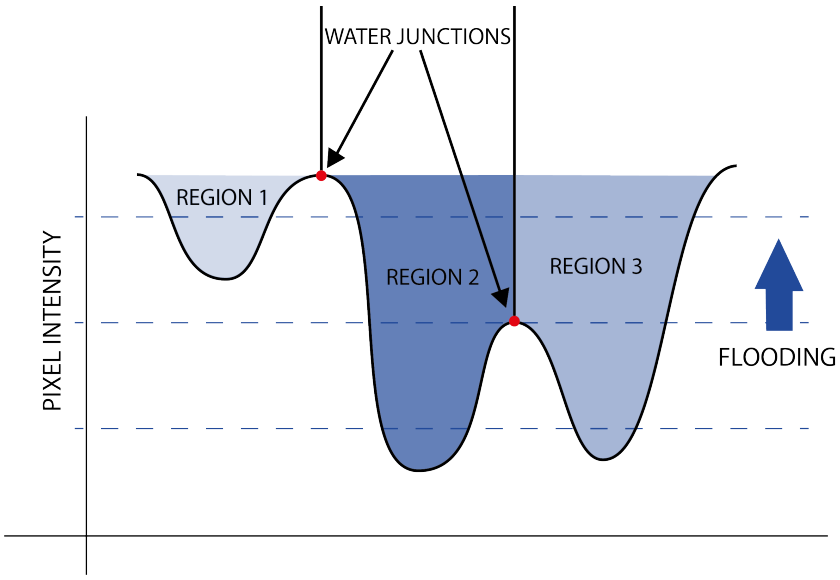


FIGURE 3.1: Water rising metaphor illustration of the Watershed algorithm.

Clustering-based Segmentation

Clustering methods allow to group observations in a way that they are more similar to other observations belonging to the same cluster than any other one. One of the most famous clustering tool is the k-means algorithm. The similarity between two pixels is computed thanks to a distance measure evaluated in a feature space on which data are projected. In the case of images, a simple feature space is the Red Green Blue (RGB) color model in which each axis is the pixel intensity along one of the 3 RGB colors. For gray scale images, usually the feature space 1-dimensional and determined by pixel intensity. However, other features such as mean gray value or SNR of neighborhood can be considered to extend the dimensionality of the clustering space. As for Threshold-based method, the localization of the pixel is not taken into account. Moreover, the number of clusters should be known in advance and the design of a relevant feature space is not straightforward. Despite these limitations, the method is fast and easy to implement [35].

In order to relax the prior knowledge about the number of clusters required by k-means algorithm, a fuzzy approach can be used. In [36], a fuzzy c-means clustering approach has been applied to CT scans to segment lung tumors. It is known that results of k-means and c-means clustering are dependent on the initialization of the centroids. In the article, authors used the Firefly algorithm to improve initialization of the centroids.

Atlas-based Segmentation

The core principle of atlas-based segmentation relies on the registration of one reference image in which contours have been drawn, i.e. the atlas, and the image in which segmentation has to be performed. After registration, labels manually defined on the reference image can be transferred to the unlabeled image. This principle is outlined in Figure 3.2.
Quality of the atlas-based segmentation relies both on the methods used for the registration and the atlas building strategy. Concerning the registration step, any registration method can be applied. In its simplest form, the atlas can be composed of one patient-specific single image. More generally, the atlas can be composed of labeled images collected on multiple patients. Since all images require to be manually labeled, data set construction represents a substantial amount of work. However, segmentation accuracy can be compromised if inter-patient variability is too high. Multiple atlases can then be built depending on key parameters influencing variability (age, gender, etc.) and selected with a majority vote. Atlas-based segmentation is therefore used in practice in body regions showing low level of anatomical variations such as the head and neck region. However, some efforts have been made recently to use atlas-based segmentation in more variable regions. A review on atlas-based recent techniques performed on medical images are presented in [37].

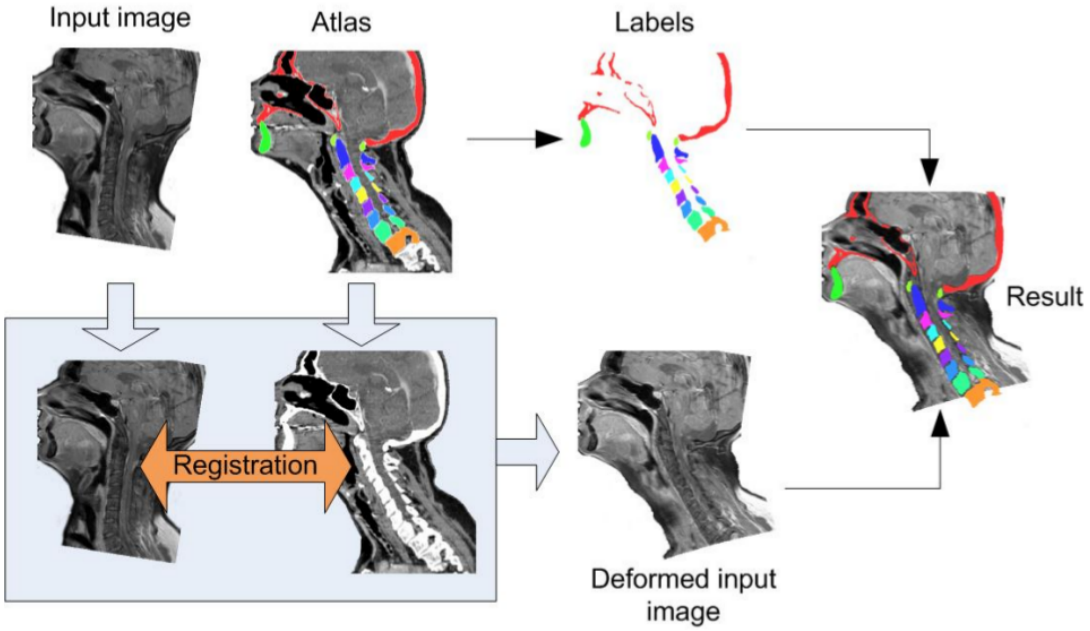


FIGURE 3.2: Core principle of atlas-based segmentation (from Erdt, Steger and Sakas [38]). Labeled reference image or atlas is registered with input image. Labels are then transferred from the atlas to the deformed input image.

Deformable Models Segmentation

Deformable models segmentation aims to find the equations of the curve or surface corresponding to the contour of a foreground object. Starting from an initial shape containing the foreground, the curve is deformed under the action of two forces. On the one hand, the internal force, which is a function of the curvature, specifies the tension and the smoothness of the contour. This term promotes minimal length curves. On the other hand, the external force, which is function of the gradient content in the image, attracts the curve toward the foreground object boundaries. This term promotes maximal length curves [39].
Depending on the geometry defining the curve, deformable models are divided into parametric and geometric models. One famous approach is the active contours, a.k.a. Snakes. This parametric method defines the curve as a discrete number of points. The approach has the advantage of being relatively fast. Unfortunately, the formulation involves parameters that are not easy to determine. Moreover, its noise sensitivity requires an initialization close to the boundary in order to avoid local minimima. In [40], active contours has been used to perform thyroid segmentation in ultrasound images. In order to compensate for the lack of gradients in the original images, a preprocessing involving HE has been applied (see Section 4.1.2). A postprocessing involving some metric of interest based on distribution of pixel intensities inside and outside the border evaluates to need of expanding further the contour. This postprocessing brings some improvement on the quality of the result.

Graph-based Segmentation

In graph-based segmentation, a.k.a. graph cut [41, 42], image is represented by a graph in which vertices represent the image pixels. The aim of the method is to find a cut of the graph by assigning vertices to distinct classes. Best segmentation result is assumed to correspond to the cut minimizing a cost function. The cost of a cut is obtained by summing weights of all edges joining vertices belonging to distinct classes. This problem can be solved thanks to min-cut/max-flow algorithm [43].
Of course, there are multiple ways to define edges of the the graph. Usually, the first type of links connects pixels with its 4 or 8 neighborhood (for a 2-dimensional image). The larger the similarity between pixel intensities, the larger the weight. These links will encourage same labeling of pixels with similar intensities in a region. A second type of links, based on an initialization performed by a user, allows to consider initialization. In the context of binary classification, all pixels marked as foreground are constrained to belong to the foreground region of the resulting image and reversely for the background region.
The initialization performed by the user constitutes the major drawback of the method. Manual labeling of pixels belonging to foreground and background regions has to be performed via an interactive window. However, this initialization usually only consist of two lines (one crossing the background and the other crossing the foreground). Advantage of the method is the versatility of the formulation that makes it able to produce application-specific results. For example, in [44], the formulation has been extended to consider 3-dimensional images varying in time in the purpose of segmenting lung tumor on 4DCTs. In this case, pixels were connected with their 6 neighbors. Additional links joined corresponding pixels from one 3DCT frame to the next one to preserve continuity of the segmentation across the frames.

Deep Learning Segmentation

Artificial intelligence methods and in particular Convolutional Neural Networks (CNNs) are increasingly used for solving wide range of image processing problems such as image classification, pattern detection or segmentation. Neural networks are multi-layered models able to capture complex nonlinear mappings between input and output. CNNs are a specific architecture of networks designed to handle input data in the form of multidimensional arrays. They can be 2-dimensional (gray scale image), 3-dimensional (RGB or 3-dimensional image) or n-dimensional in general. Usually, CNNs architecture follows a classical scheme. First layers are convolutional and pooling layers. Pooling layers are used to subsample the image and convolutional layers act as feature extractors by identifying some patterns in the image like edges. High-level or complex features can then be constructed by assembling lower-level ones in the last layers of the network.
As for any supervised machine learning problem, parameters of the model have to be trained over a labeled training set. In the case of CNNs, where millions of parameters can be tuned, the amount of data available for training have to be substantial to guaranty robustness of the network against new images, i.e. regularization property. However, the model is flexible and easily trainable on any input images. In order to segment child brain tissues, authors of [45] used multi-modality MR images as input of a CNN.

Comparison of the Methods

Table 3.1 on page 22 summarizes the main strengths and weaknesses of each of the presented algorithms.

3.3 Tracking Algorithms in Fluoroscopic Images

Object tracking is another core problem in computer vision. It can be used for numerous applications such as surveillance, human computer interaction, traffic monitoring, augmented reality, etc. Typical factors influencing robustness of tracking are illumination changes, partial or global occlusion of the target, complex or changing backgrounds. Due to the large variety of applications with their own characteristics, multiple solutions are available to handle these variations [46].

Tracking task is addressed by two main approaches: detect-to-track and track-to-detect methods [47]. On the one hand, detect-to-track method relies on two consecutive steps. First, objects are segmented in the current frame. The core problem is then to assign labels to all objects in the consecutive frames. Providing that good detectors exist, the method is robust and well suited for multiple objects tracking. These kinds of methods are also relatively robust to sporadic appearance cues. Unfortunately, label propagation is usually performed simultaneously on multiple frames involving a delayed result which is not suited for real-time processing.
On the other hand, track-to-detect approach propagates the detection in the first frame to all consecutive frames. Assuming constant appearance of the target, a model based on the initial detection, eventually updated with further detections, is used to find the part of the new frame which is most likely to contain the target. These methods are sensitive to occlusions. Bad detection at a given frame can cause bad detection in all further ones. Moreover, the problem formulation is not well suited for multiple objects tracking. However, track-to-detect approach allows real-time processing.

In the context of lung tumor tracking, track-to-detect approach is preferred over detect-to-rack due to the above-mentioned strengths and weaknesses. Note that for this specific context, the camera is assumed to be still.
In order to perform tumor tracking in fluoroscopic images, three main approaches are presented in the literature: tracking based on fiducial marker detection, deriving tumor position based on external surrogates and markerless tracking.

Tracking Based on Fiducial Marker Detection

This kind of methods provides a solution to track tumor directly but rely on an implanted fiducial marker. This marker, implanted inside or near the tumor, gives a reference point of high visibility making the tracking problem easier. Note that this approach offers a real advantage when tumor is hardly distinguishable from surrounding tissues. In the case of lung tumors, the air filling of the lungs makes lung tumors more visible than in other organs. For example, fiducial markers are particularly convenient to treat hepatic tumors with particle therapy [48].
However, the placement of the marker in the body is a hazardous task that can lead to a large pneumothorax, potentially fatal for the patient. Different methods of placement can be considered but although none of them are free of pneumothorax risk [49].

Deriving Tumor Position Based on External Surrogates

This kind of methods relies on a regression model to estimate target position based on other anatomical features easier to compute. The main weaknesses are the need on labeled data to train the regression model and the indirect aspect of the tracking. Such models assume that the relationship between external surrogates, i.e. relevant anatomical features in the image, and tumor position is constant. This restrictive assumption is not valid for the majority of the cases. However, these methods could work even if the tumor is hardly visible on acquired images.

In [50], a method is proposed to extract the lung tumor location in a 2-dimensional space based on a regression model learned over 3825 fluoroscopic images acquired over 15 patients. First, three rectangular Regions of Interest (ROIs) with fixed location are selected. One large ROI contains the diaphragm, one medium ROI captures anatomical region near the tumor and one small ROI is likely to contain the tumor itself. The 3 ROIs are processed in order to reduce their dimensionality while preserving relevant information for the regression. For this purpose, Principal Component Analysis (PCA) [51] is applied and the 3 new extracted features corresponding to the 3 largest eigenvalues are kept. A total of 3x3=9 features are then used to perform the regression. An example of projection is given in Figure 3.3. In the article, a comparison is made between four regression models: a linear and a second order polynomial regression, a neural network with one hidden layer and a Support Vector Machine (SVM) [51]. With the neural network regression model, which performed slightly better than the others, comparable results to markerless tracking can be achieved. However, the method needs to be trained over a large number of labeled data. The authors also highlighted the sensitivity of the approach faced to the ROIs selection and the number of PCA features kept. All these factors could potentially show a lack of robustness of the method especially over new patients not well represented by the training set.

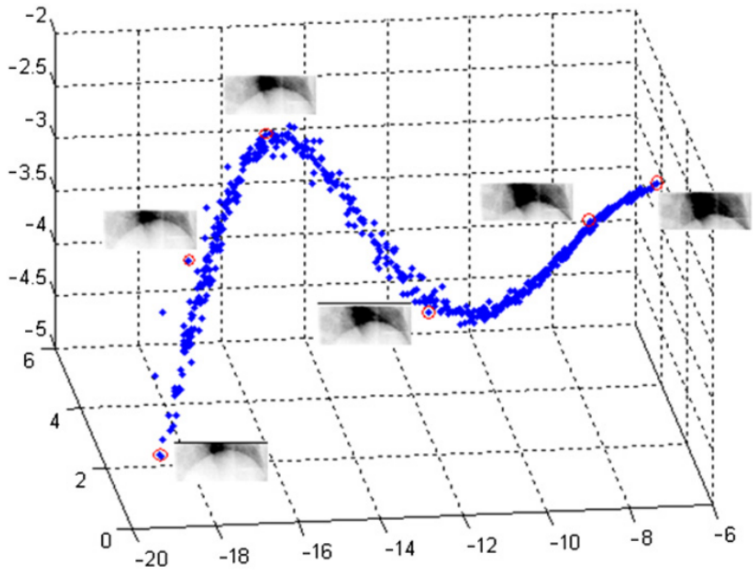


FIGURE 3.3: Projection of the ROI embedding the diaphragm in the features space corresponding to the 3 axis with largest corresponding eigenvalues (from Lin et al [50]). Corresponding Images are provided for observations denoted by a red circle.

Markerless Tracking

Unlike the previous approach that formulates the indirect tumor tracking task as a regression problem, the markerless tracking approach performs a direct tracking of the tumor. Two main approaches can be used: Template matching and Kernel-based tracking.

Template matching. Template matching uses a rectangular window encompassing the target as tumor model. By working in the original pixel space, the template represents well the tumor in all subsequent frames on the condition that tumor aspect does not vary. Therefore, the template is a bad representation of the tumor in cases where overall brightness of the image is not constant or in cases where the tumor shape varies. Template matching involves then a search mechanism, exhaustive or not, to find the rectangular region in the new frame which maximizes a similarity measure with the template. This search mechanism starts from the previous tumor location and compare the template with rectangular regions shifted in both vertical and horizontal axis. This approach is computationally expensive as the search region grows. Therefore, only a small region is considered as search space.

In [52], a multiple template-based approach is proposed in order to make the tracking more robust against illumination and shape changes. For lung tumor tracking, these changes are assumed to be mainly caused by breathing. Therefore, authors divided the breathing cycle into 12 equal time bins, each represented by a single template. Position of the tumor is determined by a voting mechanism between the location of all of the 12 rectangular regions matching best each template. The voting mechanism is a weighted sum of all final locations found if the similarity measure is above a given threshold. Given that all configurations (combination of shape and illumination) of the tumor are captured by the different templates, this template-base approach, illustrated in Figure 3.4, produces a more robust tracking than single template tracking.

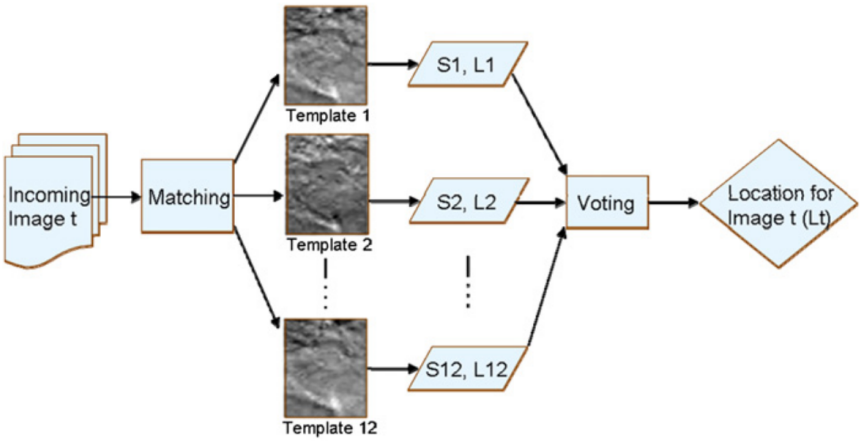


FIGURE 3.4: Principle of multiple template-based approach (from Cui et al [52]). All templates are matched with a new image acquired at time t. Each matching produces a similarity score S and a potential tumor localization L. If the score is above a given threshold, the localization is considered in the weighted average of all localization producing the final tumor localization estimation.

Kernel-based tracking. In kernel-based tracking, the template representing the target in the first frame is usually projected on a feature space of lower dimensionality. First advantage of this projection is being more robust against illumination and tumor shape change. For example, if the template is the histogram of the tumor region, the tracking will be robust to shape changes. If the template is an Histogram of Oriented Gradients (HOG), the tracking will be robust to illumination variations. A second advantage is allowing to compute similarity measure scores faster. Integrated in an iterative searching scheme, i.e. mean-shift algorithm, target searching can be achieved in less than ten iterations. Compared to template matching search mechanism, the computational complexity gain is huge. However, as for gradient descent, mean-shift is likely to fall into local maxima and a relative smooth profile of the function to maximize (determined by the similarity measure) is required for the convergence.

The authors of [53] applied kernel-based tracking to track lung tumors in fluoroscopic images previously enhanced with HE (see Section 4.1.2). Tumor template is represented by a weighted histogram of the tumor region in the first frame. The closer is a pixel from the center of the tumor region, the larger is its weight in the histogram. Therefore, the impact of a small shift in any direction results in a lower impact on the weighted histogram than on the unweighted one. Similarity function to maximize is then less sensitive to noise and therefore smoother. The similarity function used in the article is the Bhattacharyya coefficient. This measure defines a cosine distance between two vectors. First order derivative of the Bhattacharyya coefficient can be easily approximated and integrated in the mean-shift algorithm. All details of the procedure will be given in Section 6.1. Result shows as high accuracy as multiple template matching of [52]. However, search mechanism is faster and only one labeled frame is required.

TABLE 3.1: Comparison of main approaches to perform segmentation tasks.

Approach	Algorithms	Strengths	Weaknesses
Threshold based	Otsu, etc.	- Fast (process the histogram) - Simple	- Fail for overlapping gray values between clusters - Sensitive to noise - Ignore Pixel location
Region based	Region merging/splitting, Region growing, etc.	- Simple - Pixel location taken into account	- Require homogeneity of intensities within regions - Often lead to over-/under-segmentation
Boundary based	Canny, etc.	- Edges are good features to separate regions	- Often lead to over-segmentation - Sensitive to noise and edge occlusion
Watershed		- Gives connected components	- Often lead to over-segmentation - Sensitive to noise
Clustering based	k-means, Fuzzy c-means, etc.	- Fast - Simple	- Ignore Pixel location - Depend on centroids initialization (number and location) - Need to project the image in another space
Atlas based		- Efficient - Undefined number of regions	- Require labeled images - Very sensitive to variation between images
Deformable models	Active contours, etc.	- Efficient - Robust to noise - Relatively fast	- Require salient edges - Require an encapsulated initialization of the contour close to the boundary region - Need to fix parameters
Graph based	Graph cut, etc.	- Efficient - Robust to noise - Relatively fast	- Require an initialization of the foreground region - Need to fix parameters
Deep Learning		- Efficient - Able to solve very complex problems	- Require a substantial labeled data set - Training process can be time and resource consuming

Chapter 4

Image Quality Enhancement of DDRs

Due to their nature, quality of fluoroscopic images is relatively poor making, thus, the observation of the tumor difficult if not impossible. This issue has been discussed in Section 2.2. Therefore, preprocessing of these images is essential in order to use the information carried by these scans.
Preprocessing requires particular attention on contrast enhancement to make the tumor visible regarding to the surrounding tissues. Moreover, edge enhancement is also important to improve the perception of details and make the bony structures distinctly visible from the background. These two factors contribute to the improvement of automated tumor detection, performed with either segmentation or tracking methods.

The image quality enhancement step is integrated in a fully automated method chosen for the quality of its results compared to its execution time. The procedure, inspired by the work of Qiu et al. [23], involves two stages: an enhancement of edge visibility implemented through a Difference of Gaussian (DoG) filter followed by a Contrast Limited Adaptive Histogram Equalization (CLAHE) filter. The first step consists in a contrast enhancement in the frequency domain while the second operates in the spatial domain of the image.
First, we filter the image in the frequency domain with a DoG operator. The purpose of this treatment is to increase visibility of edges while avoiding amplification of noise. This issue is mitigated by removing higher frequencies of the spectrum. Implemented in the Fourier domain, DoG filter is applied on the image in a shorter period of time than guided-image filtering which require to convolve the image with local filters.
Second, we process the image in the spatial domain through a CLAHE filter. The aim is to enhance the contrast of the whole image by finding a mapping function that assigns a new value for each pixel depending on its neighborhood. the mapping function is designed in order to make the histogram as close as possible to a uniform distribution. By this way, contrast in the image should be enhanced.

As suggested by Qiu et al. in [23], parameters of the CLAHE method can be determined by entropy maximization. Intuitively, the entropy of an image whose histogram is distributed over a narrow intensity region is low. On the contrary, entropy increases when histogram tends to a uniform distribution. Maximizing the entropy leads to a set of parameters generating a resulting image with a well-spread histogram and therefore highly contrasted aspect.
Optimal values of the parameters will be determined independently for each patient. For this purpose, a single optimization is made on one DRR scan. Values found will then be used to enhance all other DRR scans of the same patient.

4.1 Theoretical Background

4.1.1 Image Enhancement in Frequency Domain

Edge detection is a well-known problem in the field of image processing. Boundary regions or edges correspond to sharp intensity transitions which are included in high frequencies region of the spectrum of an image. On the one hand, smoothing is often valuable to decrease high frequency components standing for noise content. On the other hand, smoothing has a significant blurring effect of transitions. Finding an appropriate trade-off between noise attenuation and preservation of edges location is a difficult task [54].

DoG Filter

In order to increase visibility of edges, a Difference of Gaussian (DoG) filter can be used. In a Cartesian coordinate system where each pixel is represented by a pair of values (x,y), the DoG filter is defined by Equation 4.1.

$$\begin{aligned} DoG(x,y) &= G_{\sigma_1}(x,y) - G_{\sigma_2}(x,y) \\ &= \frac{1}{2\pi\sigma_1^2} e^{-(x^2+y^2)/(2\sigma_1^2)} - \frac{1}{2\pi\sigma_2^2} e^{-(x^2+y^2)/(2\sigma_2^2)} \\ &= \frac{1}{2\pi}\left(\frac{1}{\sigma_1^2} e^{-(x^2+y^2)/(2\sigma_1^2)} - \frac{1}{\sigma_2^2} e^{-(x^2+y^2)/(2\sigma_2^2)}\right) \end{aligned} \tag{4.1}$$

where σ_1 and σ_2 are two standard deviations linked by the relation $\sigma_1 < \sigma_2$.

The filter computes the subtraction of one blurred version of the image from one less blurred one as illustrated in Figure 4.1. Each blurred image is obtained by the convolution of the image with a Gaussian kernel of specific standard deviation. The DoG filter is equivalent to a band-pass filter. By adjusting the σ_1/σ_2 ratio, different frequency bands of the image can be attenuated or exacerbated.

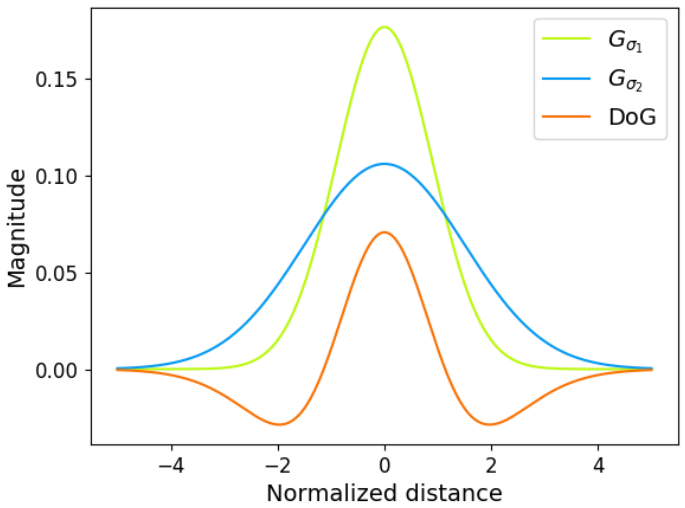


FIGURE 4.1: DoG filter in one dimension with $\sigma_1 = 0.9$ and $\sigma_2 = 1.5$.

4.1.2 Image Enhancement in Spatial Domain

In contrast to frequency domain processing techniques whose aim is to modify the spectrum of the image, we focused on image enhancement in the spatial domain. These kinds of methods operate directly on pixels by assigning a new value for each of them. This new value can be determined by a look up table or a gray-level mapping function. Such functions can be defined by the relation in Equation 4.2.

$$s = T(r) \tag{4.2}$$

where r is the original pixel value and s the new one after transformation T.

A mapping function, as described in Equation 4.2, assigns a gray level s for each original pixel value r in the image. Since we work with digital images, s and r are discrete values in the range [0, L-1]. For 16 bit encoding, L is equal to $2^{16} = 65536$.

Histogram Processing

The normalized histogram of a gray level image is defined by Equation 4.3.

$$p_r(k) = \frac{n_k}{N} \tag{4.3}$$

where n_k is the the number of pixels with gray level k with $k = 0, 1, ...L-1$ and N is the total number of pixels. $P_r(k)$ can be interpreted as the probability of occurrence of gray level k. A plot of n^k versus k is known as a histogram.

Gray level transformations applied to the histogram, as introduced by the relation depicted in Equation 4.2, are among the simplest of image enhancement techniques. Such transformations can be linear (identity or negative), logarithmic (log or inverse log) or power-law transforms (nth power or nth root). All these examples are represented in Figure 4.2.

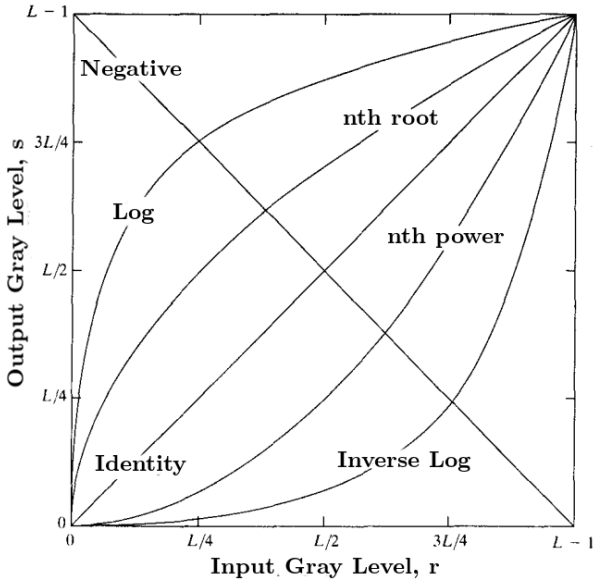


FIGURE 4.2: Basic gray-level transforms (from the book of Gonzalez and Woods [55]).

These transforms have a wide range of applications despite their simplicity. For example, logarithmic transform can be used to expand darker regions of an image and compress brighter regions at the same time. Power law transforms, depicted by the relation $s = cr^{\gamma}$ with c and γ being positive constants, can be used for gamma correction. The name is due to the γ symbol referring to the exponent. Gamma correction has been used for decades to compensate for the non-linearity of intensity-to-voltage response of Cathode Ray Tubes (CRT).

Histogram Equalization (HE)

HE is a kind of transformation depicted by the Equation 4.4.

$$s_k = T(r_k) = \sum_{j=0}^{k} p_r(r_j) = \sum_{j=0}^{k} \frac{n_k}{N} \tag{4.4}$$

where k = 0, 1, 2, ..., L-1, r_k is the original value of a pixel with gray level k, s_k the new value for the corresponding pixel, $p_r(r_j)$ the probability of occurrence of gray level j (defined by Equation 4.3), n_k the number of pixels with gray level k and N the total number of pixels.

HE aims at finding a mapping function that stretches the histogram to approximate it to a uniform distribution, i.e. n_k is a constant equal to N/L. After mapping, the number of pixels in each bin of the histogram should be equal.
The transform respects the 2 conditions :

1. T(r) is a single-valued monotonically increasing function in the interval $0 \leq r \leq 1$.
2. T(r) maps the range [0,1] to [0,1].

Holding these two conditions ensures to find the inverse mapping $r = T^{-1}(s)$.

Even if the method shows improvement in terms of contrast, some undesirable outcomes often result from the mapping. Noise can be over amplified in homogeneous regions such as background. In this case, large peaks from the original histogram are mapped in a much spread region in the new histogram resulting in noise amplification.

Adaptive Histogram Equalization (AHE)

HE is based on the assumption that image quality and noise level are constant over the entire image. Therefore, applying the same mapping function over all sections of the image will produce similar enhancements. When grayscale distribution of pixels changes from one region to another, the assumption is not valid. For such cases, AHE approach can bring significant improvement compared to HE [26].

The principle of AHE is to consider the neighborhood of pixels when computing the mapping function. The function is then dependent of the intensity the pixel itself and of the ones in the surrounding region. Compared to HE that needs only the histogram to compute mapping function, the number of operations required to compute each new gray value of each pixel is equal to the number of pixels in the image ($\mathcal{O}(N)$). This fully adaptive approach is computationally expensive and a block-based processing of histogram is often preferred. In block-based processing, the image is divided into subregions within which HE is performed. A bilinear interpolation scheme depicted in Figure 4.3 is then applied on the

result to avoid blocking artifacts. In this figure, new gray value of the selected pixel (denoted by a white dot) is derived from assignment of pixels belonging to the center of contextual region (A, B , C and D blocks whose centers are denoted by black dots). Assuming that s is the value of the pixel after projection and r the original value, the assigned gray value for the selected pixel is given by $s = (1-y)\left((1-x)T_A(r) + xT_B(r)\right) + y\left((1-x)T_C(r) + xT_D(r)\right)$

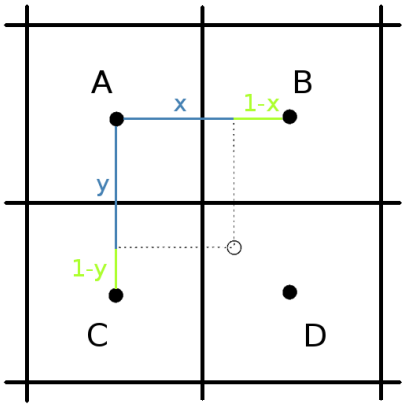


FIGURE 4.3: Interpolation scheme of Adaptive filtering (based on Zuiderveld [25]).

Contrast Limited Adaptive Histogram Equalization (CLAHE)

The issue of noise amplification in homogeneous regions can be mitigated by limiting level of contrast enhancements in such areas [56]. An homogeneous region is characterized by a large peak in the histogram. With traditional HE, this peak would have been spread over a wide range of gray values. The specificity of CLAHE is limiting the slope of the mapping function. This can be done by clipping the histogram, i.e. enforcing a maximum on the count in the histogram as seen in Figure 4.4.

Clipping the histogram is equivalent to limiting the number of pixels in each of the bins of the histogram. All pixels above this threshold are redistributed over the whole histogram. The threshold called clip limit, whose value is related to a clip factor α expressed in percent, is defined by Equation 4.5.

$$\beta = \frac{N_{\text{block}}}{L} \underbrace{\left(1 + \frac{\alpha}{100}\left(S_{max} - 1\right)\right)}_{\text{multiplicative factor}} \tag{4.5}$$

where N_{block} is the number of pixels in each block, L is the number of gray levels and S_{max} the maximum allowed slope of the mapping function determined experimentally (4 is a good values for X-ray images [26]).

The clip limit β is interpreted as a multiple of the average histogram content (N_{block}/L). A low value of this factor limits the contrast enhancement related to the slope of the mapping function. A multiplicative factor of 1, which corresponds to a clip factor of 0, prohibits any contrast enhancement since the mapping function is then the identity transform. On the contrary, a very large multiplicative factor (higher than one thousand) produces similar results than AHE technique.

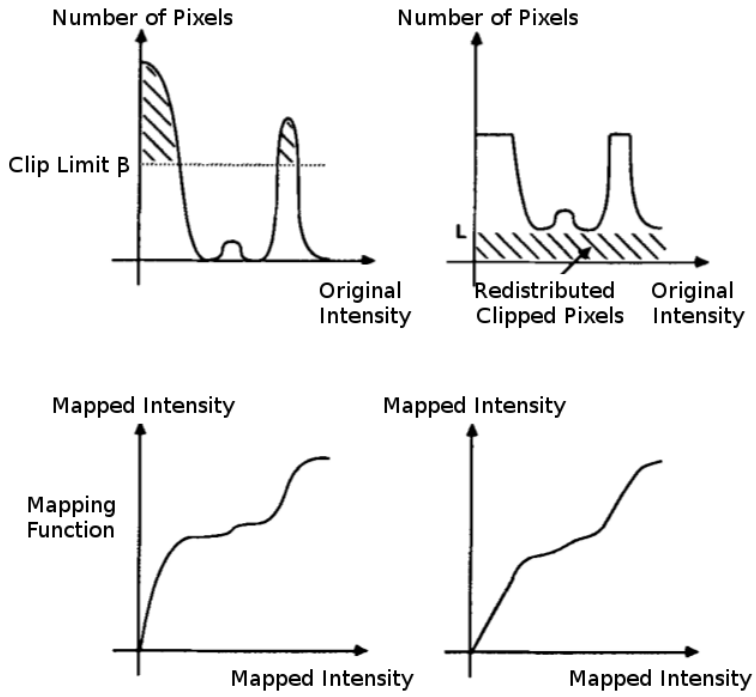


FIGURE 4.4: Non-clipped and clipped histograms of an image and their corresponding mapping functions to equalize the original histograms (from Zuiderveld [25]).

4.2 Proposed Method

The proposed method to enhance quality of the DRRs is implemented thanks a two-stage procedure relying on DoG filtering and CLAHE. The method involves 2 steps and 3 parameters :

Step 1 - A filtering in the frequency domain implemented with a DoG filter

Parameter 1: p_1 **-** A multiplicative factor to balance the level of edge visibility enhancement (see Equation 4.6). p_1 is greater than or equal to 0.

Step 2 - A filtering in the spatial domain implemented with a CLAHE filter

Parameter 2: p_2 **-** Number of divisions along one axis for the CLAHE filtering. The total number of blocks is then equal to p_2^2. p_2 is a positive integer greater than or equal to 2.

Parameter 3: p_3 **-** The multiplicative factor used to compute the clip limit of the CLAHE filtering (see Equation 4.5). p_3 is greater than or equal to 1.

The value of these three parameters are automatically determined for each patient through a maximization of the entropy. Two formulations of the loss function are studied. In the first case, the loss function to maximize is the entropy of the whole image. In the second case, the loss function to maximize is the entropy of a neighborhood region defined around the tumor.

4.2.1 Step 1: Enhance Edge Visibility

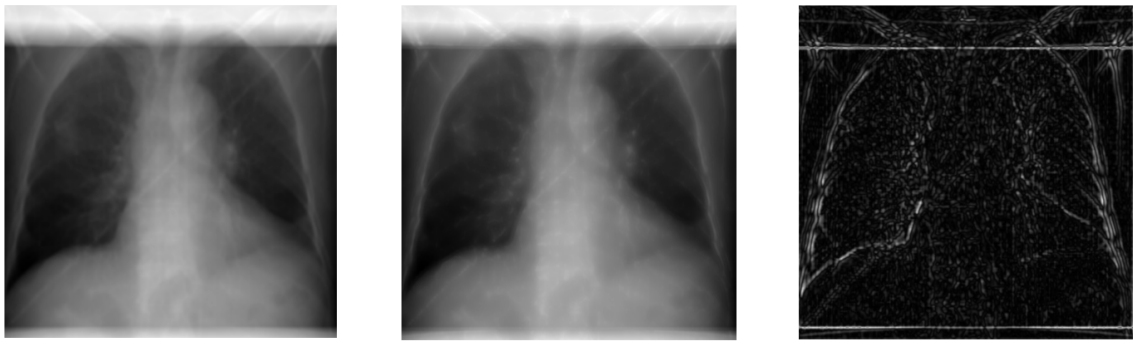
The first step of the method produces an image I_{DoG} by the transformation described in Equation 4.6.

$$\begin{aligned} I_{DoG} &= I \circledast G_{\sigma_1} + p_1 \times I \circledast \underbrace{(G_{\sigma_1} - G_{\sigma_2})}_{\text{DoG filter}} \\ &= I \circledast G_{\sigma_1} + p_1 \times I \circledast G_{\sigma_1} - p_1 \times I \circledast G_{\sigma_2} \\ &= (p_1 + 1) \times I \circledast G_{\sigma_1} - p_1 \times I \circledast G_{\sigma_2} \end{aligned} \tag{4.6}$$

where I is the original image, and G_{σ_1} and G_{σ_2} are two Gaussian filters of strandard deviations σ_1 and σ_2 respectively (with $\sigma_1 < \sigma_2$ as defined in the Equation 4.1).

The first step of the method consists in adding edges ($I \circledast (G_{\sigma_1} - G_{\sigma_2})$) to the original image convolved with a high pass filter ($I \circledast G_1(\sigma_1)$). Edges concerned by the enhancement correspond to sharp transitions between gray level of adjacent regions whose frequencies are contained in the pass band range of the DoG filter. As mentioned previously, the σ_1/σ_2 ratio affects the scale at which edges are detected. In this context, we decided to fix this ratio to $3/5$ by visual inspection of the result. This specific ration allows the filter to detect edges of bony structures and diaphragm. Note that the tumor boundaries are not detected by this filter.

An example is shown in Figure 4.5. In this figure, we observe that the DoG filter detected bony structures as well as the diaphragm boundaries. Even if the edge visibility enhancement is relatively low at this stage, the next step of the procedure consists in applying CLAHE algorithm on the image. This method will increase the contrast present in the DRRs. By this way, a small increase of edge visibility will be exacerbated by CLAHE during the next step of the quality enhancement method.



(a) Original image

(b) Image with edge visibility enhancements

(c) Difference between image filtered with G_{σ_1} and G_{σ_2} with $\sigma_1 = 0.9$ and $\sigma_2 = 1.5$

FIGURE 4.5: Edge enhancement step performed on patient 2.

DoG filtered image is multiplied by a weighting factor p_1 adjusting level of edge visibility increase. The value of this factor is the only parameter of the first step optimized through entropy maximization.

4.2.2 Step 2: Enhance Contrast

The second step of the method consists in applying CLAHE (described in Section 4.1.2) on I_{DoG}. As mentioned above, two parameters have to be fixed: p_2 the number of divisions along one axis of the image (the image is then divided into p_2^2 distinct blocks) and p_3 the multiplicative factor used to clip the histograms of each block.

4.2.3 Optimization of the Parameters

As mentioned in Section 4.2, the method requires 3 parameters whose value is automatically determined for each patient with entropy guided optimization scheme. The entropy of an image is a measure of randomness of the image defined by Equation 4.7.

$$\mathcal{H}(x_j) = -\sum_{j=1}^{L} p(x_j)\, log_2\left(p(x_j)\right) \tag{4.7}$$

where $p(x_j)$ is the normalized probability of the gray level x_j.

The entropy of an image is directly linked to the usage of available gray levels in it. The wider the histogram over its range of gray values, the higher the entropy. Entropy maximization tends to parameters promoting well spread of the histogram. In this context two loss functions based on entropy have been studied :

Loss 1 - Global entropy. The function to maximize is the entropy over the entire image as defined by equation 4.7.

Loss 2 - Local entropy. The function to maximize is the entropy of a contextual region around the tumor. Given all the points defining the contour of the tumor, the contextual region is defined as all the points belonging in the rectangle of coordinates (x_{min}, y_{min}), (x_{min}, y_{max}), (x_{max}, y_{max}), (x_{max}, y_{min}) :

$$\begin{cases} x_{min} &= max(0, & X_{tumorMIN} & - 0.5 * (X_{tumorMAX} - X_{tumorMIN})) \\ x_{max} &= min(X_{imageMAX}, & X_{tumorMAX} & + 0.5 * (X_{tumorMAX} - X_{tumorMIN})) \\ y_{min} &= max(0, & Y_{tumorMIN} & - 0.5 * (Y_{tumorMAX} - Y_{tumorMIN})) \\ y_{max} &= min(Y_{imageMAX}, & Y_{tumorMAX} & + 0.5 * (Y_{tumorMAX} - Y_{tumorMIN})) \end{cases}$$

where $X_{imageMAX}$ and $Y_{imageMAX}$ are the size of the image relative to X and Y axes respectively, $X_{tumorMIN}$ and $X_{tumorMAX}$ are the maximum index of tumor contour relative to X and Y axes respectively. The contextual region is represented in Figure 4.6.

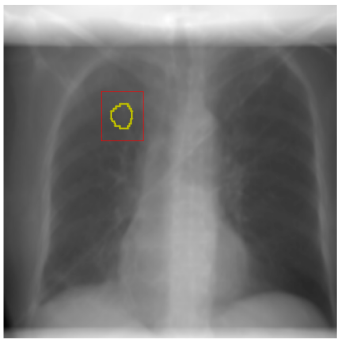


FIGURE 4.6: Contextual region and Gross Tumor Volume (GTV) for patient 1.

4.3 Results

Optimal CLAHE parameters have been determined for each patient on the first scan of the first series used as reference image. Results of constrained optimization with both global and local entropy loss functions are shown in Table 4.1 and 4.2 respectively. These tables report optimal values obtained for the 3 parameters, the increase of entropy, in either the whole image (Loss 1) or in the contextual region (Loss 2), expressed in percent and the time required by the optimization algorithm to converge.
Values of p_1, which quantify the level of edge visibility enhancement, are higher when the optimization is driven by the entropy of the contextual region. Gain of entropy is also more substantial with local entropy loss function. Parameter p_2 is equal to 2 except in one case. The multiplicative factor p_3 varies between 8.7 and 17.18 and computational time between 10 and 21 minutes. Both do not show any particular trend.

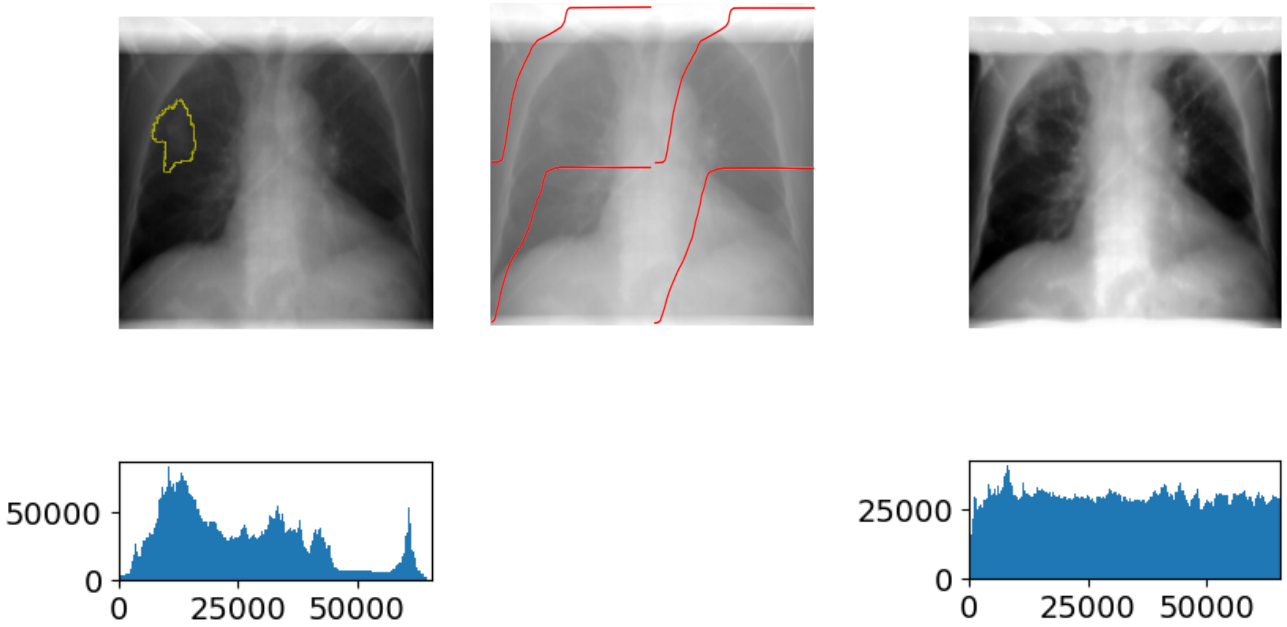
TABLE 4.1: Optimized parameters values according to global entropy maximization with initial values: p_1 = 0.5, p_2=2 and p_3=4.

Loss1 : Global entropy					
Patient ID	p_1	p_2	p_3	Global Entropy Increase (in %)	Convergence Time
patient1	1.81	2	14.8	4.28	14 mins 48 secs
patient2	2.1	2	7.68	2.07	11 mins 22 secs
patient3	0.45	2	9.36	2.09	10 mins 29 secs

TABLE 4.2: Optimized parameters values according to local entropy maximization with initial values: p_1 = 0.5, p_2=2 and p_3=4.

Loss2 : Local entropy					
Patient ID	p_1	p_2	p_3	Local Entropy Increase (in %)	Convergence Time
patient1	2.7	2	11.01	14.7	14 mins 03 secs
patient2	5.5	3	17.18	9.97	20 mins 58 secs
patient3	4.53	2	8.7	4.33	11 mins 47 secs

Application of the proposed method shows significant contrast enhancement of the image for all three patients. Dynamic range of histograms has been increased in order to make use of available gray levels. Depending on the patient, results obtained with global and local entropy maximization lead to more or less different images. Visibility of tumor edges has significantly increased with parameters determined with local entropy maximization for patient 2 as we can see in Figure 4.8(c) versus 4.7(c).
In order to quantify this enhancement, Table 4.3 gives the increase percentage of the sum of the norm of the gradients in X and Y directions in the contextual region of the tumor. Values are compared with gradients from the original image. For the three patients, optimal parameters maximizing Loss 2 lead to higher gradients in contextual region than those maximizing Loss 1. However, this increase is more important for patient 2.

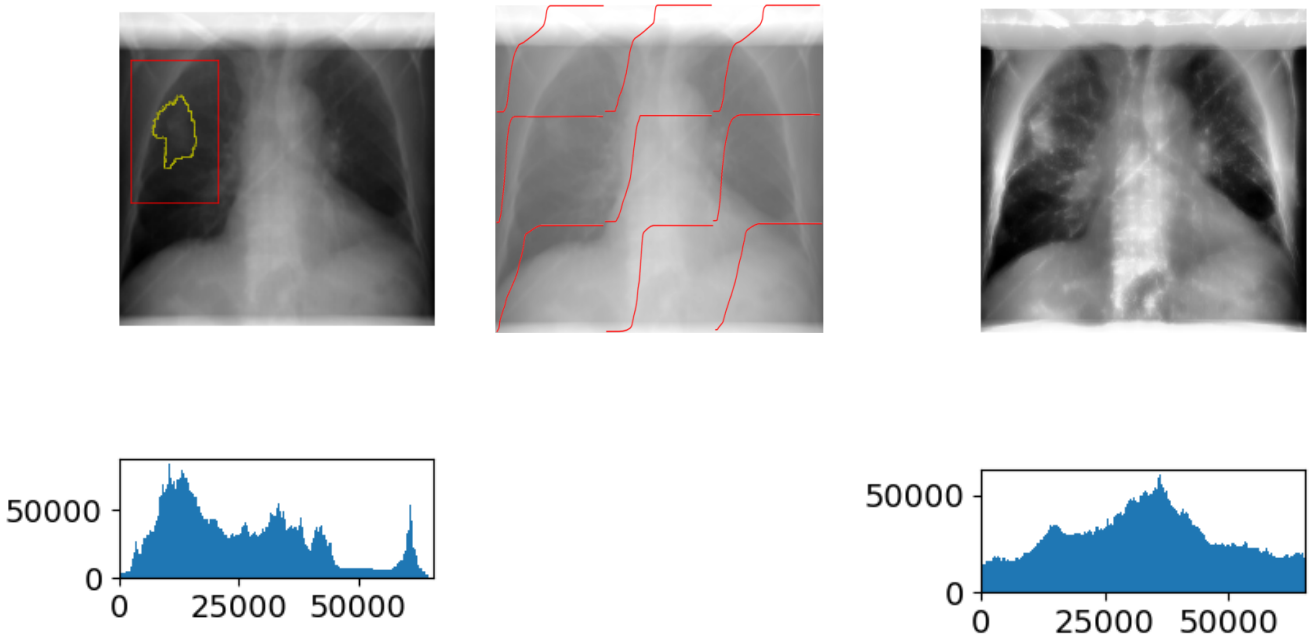


(a) Original image

(b) Representation of the mapping functions of each subregion of the image

(c) Contrast enhanced image with optimal parameters determined by global entropy maximization

FIGURE 4.7: Results of contrast enhancement method for patient 2 with optimal parameters determined by global entropy maximization (Loss 1). Each image is presented with its histogram (computed on the entire image).



(a) Original image

(b) Representation of the mapping functions of each subregion of the image

(c) Contrast enhanced image with optimal parameters determined by global entropy maximization

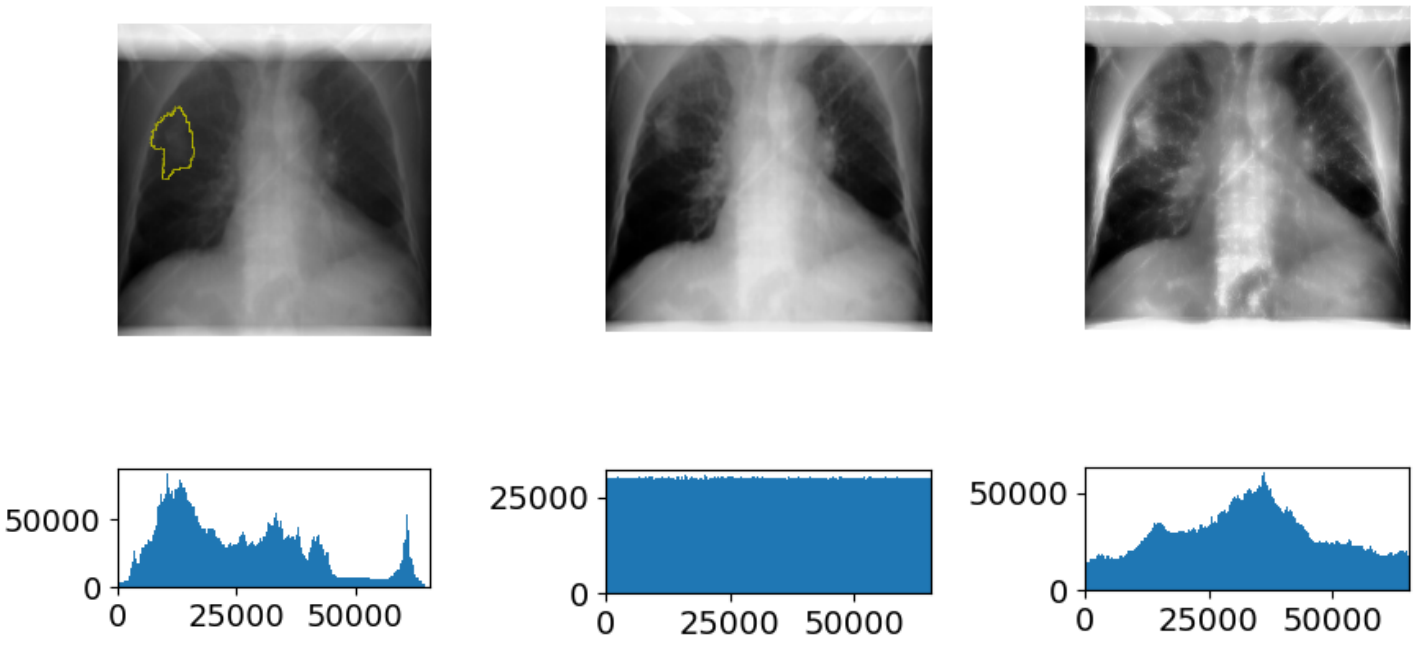
FIGURE 4.8: Results of contrast enhancement method for patient 2 with optimal parameters determined by local entropy maximization (Loss 2). Each image is presented with its histogram (computed on the entire image).

TABLE 4.3: Percentage of increase in the sum of the norm of the gradients in X and Y direction in the contextual region around the tumor.

Percentage of increase in the sum of norms of the X and Y gradients in tumor's contextual region		
Patient ID	Proposed method with optimal parameters found by global entropy maximization (Loss1)	Proposed method with optimal parameters found by local entropy maximization (Loss2)
patient1	199.9	219.35
patient2	106.7	276.72
patient3	46.48	70.18

4.4 Discussion

Performances of CLAHE versus HE. Compared to CLAHE, HE has the advantage of being a non-parametric method. Since parameters optimization can be relatively time consuming as presented in the previous section, a comparison of the results obtained with the two methods is valuable to assess the need of a parametric method. Figure 4.9 shows enhanced images obtained with HE (see Figure 4.9(b)) and with CLAHE (see Figure 4.9(c)). Note that edge enhancement step has been performed in the two cases.

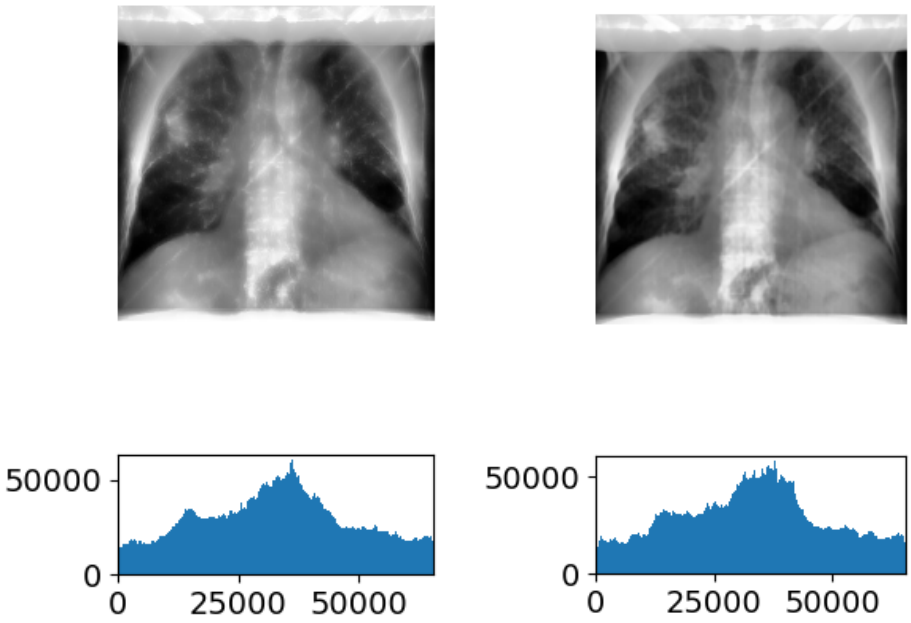


(a) Orignial image (b) With HE (non parametric) (c) With CLAHE (optimal parameters determined by local entropy maximization)

FIGURE 4.9: Comparison of contrast enhancement performed with HE and CLAHE for patient 2. Both results are performed with enhancement of edges visibility via DoG filtering.

Visual inspection of the result unequivocally indicates that HE method is inadequate to perform contrast enhancement of the DRRs. With HE, contrast within soft tissues is not enhanced and bony structures, like vertebrae, are less visible. Moreover, the tumor is much more noticeable with CLAHE than with HE. Results obtained with CLAHE outperform those of HE.

Influence of parameters. The first parameter of the proposed method that is discussed is p_1. Effect of p_1 is to quantify level of edges enhancement of the resulting image. Figure 4.10 compares result obtained with (Figure 4.10(a)) and without (4.10(b)) edge enhancement. Small differences can be observed near to bony structures but they are relatively low. Considering this argument, utility of this filtering can be argued. However, in the context of this work, we will keep this filtering in order to perform segmentation and tracking on best possible images.

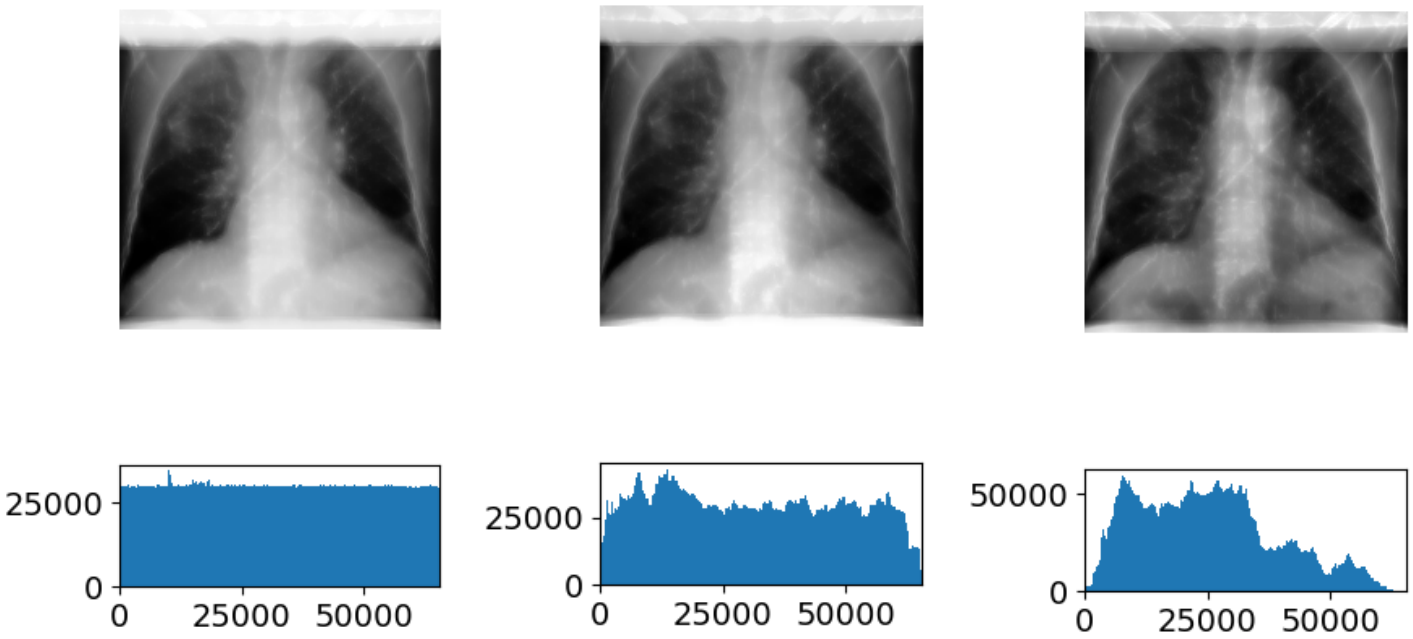


(a) p_1 maximizing local entropy (b) $p_1 = 0$

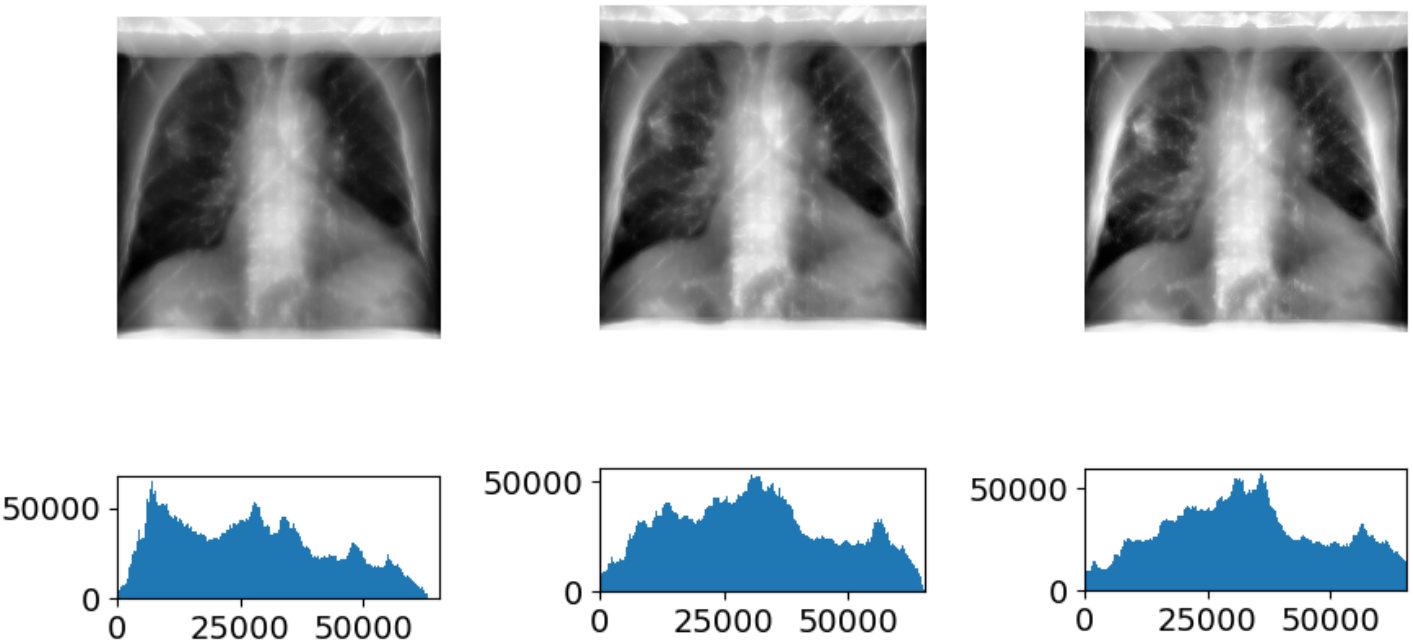
FIGURE 4.10: Comparison of CLAHE contrast enhancement with p_1 determined by local entropy maximization and $p_1 = 0$ for patient 2.

Influence of p_2 and p_3 can also be studied in order to get some insight about the CLAHE method. Figure 4.11 presents some results obtained by varying p_2 and p_3.
First, the variation of the number of blocks dividing the image (p_2^2) can be seen in Figures 4.11(a), 4.11(b), 4.11(d) and 4.11(c). In a general way, the larger is p_2, the less uneven is the histogram and larger is the contrast bony structures. Regarding the tumor, the effect of p_2 is relatively low to increase tumor visibility. Note that with $p_2 = 1$ and p_3 sufficiently large, result is identical as HE.
Second, the effect of the multiplicative factor determining the clip limit is quite significant on the tumor's visibility as seen in Figures 4.11(d), 4.11(e) and 4.11(f). Within each block, a large clip factor encourages the spreading of large peaks into wide regions in the output histogram. While it can be unfavorable over the entire image because of the non-uniformity of

gray level distribution, this can increase contrast within a block in which gray level content is more or less narrow. The combination of the division into blocks and a large clip factor produces large enhancement of the tumor visibility at the expense of the natural look of the result.



(a) $p_2 = 1$ and $p_3 = 5$ (b) $p_2 = 2$ and $p_3 = 5$ (c) $p_2 = 6$ and $p_3 = 5$



(d) $p_2 = 4$ and $p_3 = 5$ (e) $p_2 = 4$ and $p_3 = 10$ (f) $p_2 = 4$ and $p_3 = 20$

FIGURE 4.11: Influence of p_2 and p_3 parameters on the results obtained with CLAHE for patient 2. p_1 is equal to the optimal value determined by local entropy maximization.

Chapter 5

Lung Tumor Segmentation

In this chapter, lung tumor segmentation is performed on the DRRs preprocessed with the method described in Chapter 4. In order to perform the binary segmentation in foreground/background regions, the graph cut approach has been chosen due to a good trade-off between the expected quality of results (as demonstrated in the literature), the execution time, the versatility of its formulation and the information required as initialization.

More specifically, as discussed in Section 3.2, threshold- and region-based approaches were discarded due to their poor performances when homogeneity of gray values is not well established within regions.
Boundary-based detection and Watershed were also discarded due to the oversegmentation issue. Several corrections of the result in a feedback loop are usually required to enhance the segmentation making the approach time consuming.
Clustering was not eligible due to the loss of localization information and its sensitivity to the initialization. A more robust approach that produces constant high quality result is preferred.
Atlas-based and deep learning methods, despite their promising results, are discarded due to the limited amount of data collected. On the one hand, even if atlas based methods can be applied with one single atlas, single-atlas approaches are not competitive compared to other segmentation methods in body regions subject to high anatomical displacement. On the other hand, the supervised learning of deep neural networks is simply not feasible with such amount of data.
Active contours and graph cut are the two strongest candidates to achieve the segmentation problem. On the one hand, active contours method is initialized with a curve encompassing the tumor. This curve needs absolutely to encapsulate the whole target and should be as close as possible to the boundaries in order to avoid delineations corresponding to local minima of the energy function. Such an initialization is not easy to provide with a high degree of certainty. On the other hand, the proposed method takes advantage of the versatility of the graph cut problem formulation by initializing the method with a binary mask of the tumor acquired in the same respiratory phase. We hope that, with proper parameters, the method could find the true delineation even if the binary mask is not at the exact position of the target. This would make the initialization less complex than the one of active contours.

Assuming that the tumor location and shape are similar within a given respiratory phase, the contour of the first series are used as initialization to perform graph cut on each corresponding scan on next three series (as referred previously in Table 2.1). Parameters of the graph cut algorithm will also be discussed in this chapter.

5.1 Theoretical Background

The aim of binary segmentation is to assign a label to each pixel of the image. The label set is $L = \{0, 1\}$. All pixels belonging to the "object" or "foreground" region, denoted by the subset $\mathcal{O}$, are assigned the value 1. On the contrary, all pixels belonging to the "background" region, denoted by the subset $\mathcal{B}$, are assigned the value 0.

5.1.1 Graph Construction

In a graph cut segmentation approach, the image is modeled with an undirected weighted graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V}$ is the set of nodes (the vertices) and $\mathcal{E}$ is the set of edges connecting them. In the context of computer vision, each pixel corresponds to one node. Additional nodes can also be defined. Usually, two nodes, referred as terminals, are added: the source s and the sink t. Two types of edges link nodes between them:

n-links. These links connect a non-terminal node with its other non-terminal neighbors nodes. A four neighborhood region $\mathcal{N}$ is used in this context for 2-dimensional image segmentation. The non-negative weight associated with these links is proportional to the similarity between intensities of the two pixels. For a node q belonging to the neighboring region $\mathcal{N}$ of p, this weight is given by the function $B(I_p, I_q)$ where I_p and I_q are intensities of p and q nodes respectively.

t-links. They connect each non-terminal node to the source s and sink t nodes by a terminal link. Initial foreground and background regions need to be manually defined by the user. For this purpose, the user "marks" several pixels as belonging to the foreground region $\mathcal{O}$ and other as belonging to to the background $\mathcal{B}$. Costs associated with these links are very high in order to classify them in the region marked by the user.
All other "non-marked" nodes receive a weight proportional to the probability that a pixel belongs to the foreground (link with s node) or background (link with t node) regions. This probability is computed with function $R(p)$ where p is the non-terminal node for which probabilities to belong to foreground and background region are computed.

Table 5.1 summarizes costs attributed to edges depending on the link between the two concerned pixels.

TABLE 5.1: Weight of edges between a non-terminal node p and other nodes depending of the type of link. The λ parameter is a weighting factor balancing the relative weights of *n-links* and *t-links*.

Link	Weight	Initialization
n-link: $\{p, q\}$	$\lambda \times B(I_p, I_q)$	$\forall p$ if $q \in \mathcal{N}_p$
t-link: $\{p, s\}$	$R(I_p, \mathcal{O})$	$p \notin \mathcal{O} \cup \mathcal{B}$
	∞	$p \in \mathcal{O}$
	0	$p \in \mathcal{B}$
t-link: $\{p, t\}$	$R(I_p, \mathcal{B})$	$p \notin \mathcal{O} \cup \mathcal{B}$
	0	$p \in \mathcal{O}$
	∞	$p \in \mathcal{B}$

An example of graph construction for a simple 3x3 pixels image is represented in Figure 5.1. In this figure, the upper left picture represents the image in the pixel space with two nodes "marked" as background/foreground. These two nodes are the two "marked" nodes provided by the initialization process. The down left picture shows the corresponding graph for which the edge thicknesses are proportional to the weight of the edge (bold link between t node and pixel labeled as part of the background/ bold link between s node and pixel labeled as part of the foreground). The minimal cut of the graph is represented on the graph showed in the down right part of the figure. The image in the upper right corner is the binary segmentation result determined by the minimal cut. Light gray pixels belong to the background region and black pixels to the foreground region.

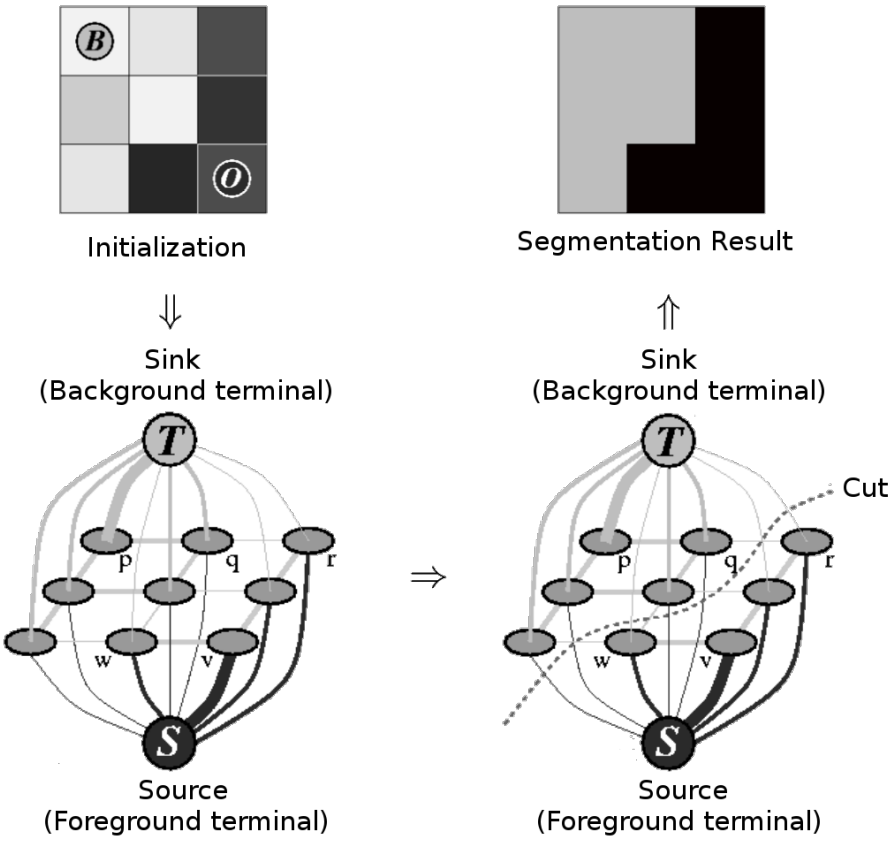


FIGURE 5.1: Graph cut segmentation of a 3x3 pixels image (based on Boykov and Jolly [41]).

A segmentation of the image corresponds to an s-t cut of the graph. A cut $\mathcal{C}$ is a subset of $\mathcal{E}$ that divides all nodes into two sets when $\mathcal{C}$ is removed from $\mathcal{G}$: $\mathcal{G}(\mathcal{C}) = \langle \mathcal{V}, \mathcal{E} \backslash \mathcal{C} \rangle$. The first set $\mathcal{S}$ embraces all nodes that are likely to belong to the foreground region. The second set $\mathcal{T} = \mathcal{V} - \mathcal{S}$ embraces all nodes that are likely to belong to the background region. According to these definitions, the two sets are disjoints : $\mathcal{S} \cap \mathcal{T} = \emptyset$. Note also that $s \in \mathcal{S}$ and $t \in \mathcal{T}$.

Each cut $\mathcal{C}$ is associated with a cost equal to the sum of the edge weights belonging to the cut. In other words, this cost is $|\mathcal{C}| = \sum_{e \in \mathcal{C}} w_e$.

5.1.2 Energy Minimization Problem

The segmentation problem, which aims to find minimal cost cut of a graph, can be solved through minimization of the energy function described in Equation 5.1. In this formula, $\mathcal{P}$ is the set of all non-terminal nodes ($\mathcal{P} = \mathcal{V} \backslash \{s, t\}$) and $\mathcal{A}$ is a binary vector of size $|\mathcal{P}|$ assigning a label from $L = \{0, 1\}$ to each node of $\mathcal{P}$.

$$E(A) = \underbrace{\sum_{p \in \mathcal{P}} R(I_p, A_p)}_{\text{Regional term (\textit{t-links})}} + \lambda \underbrace{\sum_{\{p,q\} \in \mathcal{N}} B(I_p, I_q) \cdot \delta(A_p, A_q)}_{\text{Boundary term (\textit{n-links})}} \tag{5.1}$$

$$\text{with } \delta(A_p, A_q) = \begin{cases} 1 & \text{if } A_p \neq A_q \\ 0 & \text{otherwise.} \end{cases}$$

Regional term $R(I_p, A_p)$ reflects how well a pixel p of intensity I_p fits into a known foreground and background model. Value returned by $R(I_p, A_p)$ is a penalty for assigning label A_p to pixel p depending on prior information about foreground and background appearance. Boundary term $B(I_p, I_q)$ can be interpreted as a penalty when different labels are assigned to neighboring pixels. This term preserves continuity.
The λ parameter is a positive factor balancing importance of the regional term versus the boundary term. Smaller values of λ give large importance to the prior information given during initialization. Bigger values of λ increase the impact of the intensities of the pixels on the result.

Minimization of the energy function, i.e. segmentation the image, can be achieved using max-flow/min-cut algorithm [41]. Several algorithms like augmenting paths-based methods or Dinic algorithm solve this problem. However, in the context of computer vision, Boykov-Kolmogorov algorithm introduced in [42] outperforms standard algorithms accuracy even if its complexity is a little bit much higher (in $\mathcal{O}\left(mn^2|\mathcal{C}|\right)$ with n the number of nodes and m the number of edges).

5.2 Proposed Method

In its classical form, graph cut method requires the user to mark pixels as belonging to foreground and background regions. Usually, an interactive window allows the user to draw two lines. All pixels crossing foreground line are classified as foreground and all pixels crossing background line are classified as background. Weights of *t-links* associated with these "marked" nodes have either an infinite or a null weight (see Table 5.1). Since the objective is to find a minimal cut, these pixels need to be well classified to avoid infinite cost cut. Therefore, assigning them an infinite weight is a way to define a hard constraint on the image segmentation.

The proposed approach differs in the sense that the initialization is performed automatically reducing by this way the initialization time which is necessary for real-time applications. Also constraints on "marked" pixels are relaxed. The initialization is given by a binary mask that should be representative of the tumor shape and location in the image in which the segmentation task is preformed.

5.2.1 Initialization

Instead of asking a user to draw manually foreground and background regions, we use the tumor binary masks of the first series to initialize the foreground region on all scans of the three remaining series that correspond to the same respiratory phase. This Initialization procedure is represented in Figure 5.2. In order to avoid misalignment between the tumor binary mask and the tumor location in the new scan, first scans of the three last series are aligned with the first scan of the first series with a rigid registration process. The affine transform matrix found for the first scan is applied on all other scans of the same series in order to align tumor motion with the one of the first series.

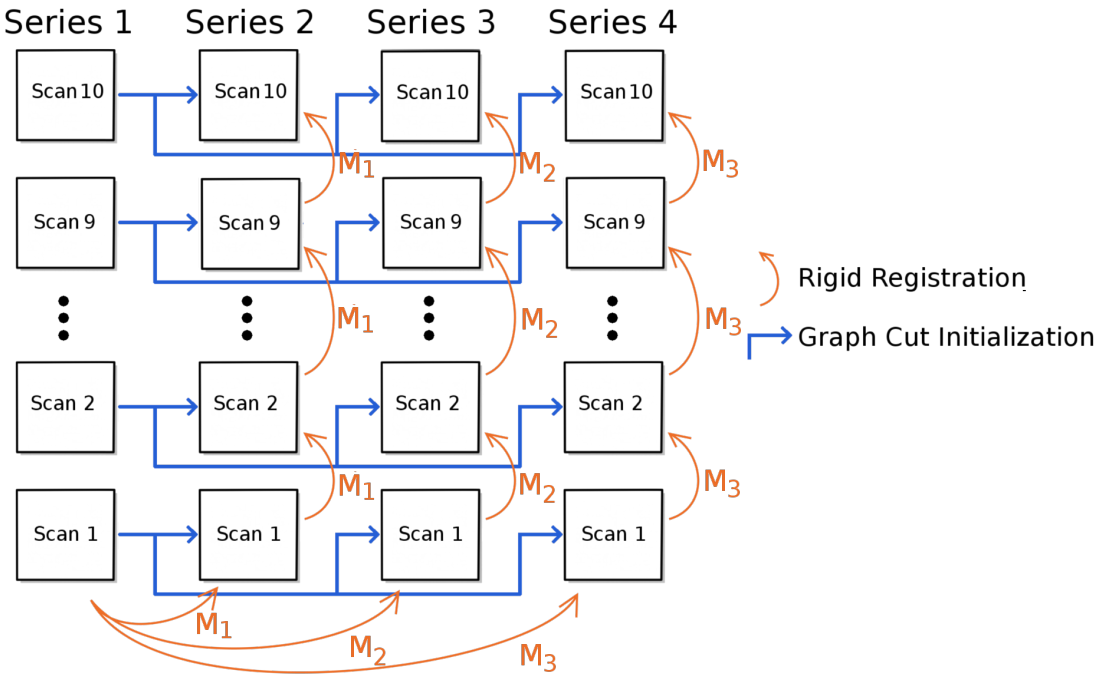


FIGURE 5.2: Initialization of the graph cut segmentation algorithm with our data base.

Rigid registration aims to find a spatial transformation of a "moving" image so that the match between this transformed image is the most similar to a "fixed" reference image [57]. More precisely, an affine transform consists of translations, rotations, zooms and shears preserving line parallelism property in the image. All these operations can be combined into a single linear transformation denoted by a matrix $\boldsymbol{M}$. Each point in the original 2-dimensional image (x,y,1) is mapped to a new coordinated system (x',y',1) as presented by Equation 5.2. Note that the points are expressed in homogeneous coordinates which add an extra coordinate in order to denote all these transforms by a single 3×3 matrix.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \boldsymbol{M} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & t_x \\ m_{21} & m_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Leftrightarrow \begin{cases} x' = & m_{11}x + m_{12}y + t_x \\ y' = & m_{21}x + m_{22}y + t_y \end{cases} \tag{5.2}$$

where t_x and t_y are the 2 degrees of freedom of the translation and m_{11}, m_{12}, m_{21}, m_{22} the 4 degrees of freedom of the rotation, zoom and shear operations. A 2-dimensional affine registration is then characterized by 6 degrees of freedom.

This initialization relies on a crucial hypothesis: it assumes that the tumor location and morphology in a given respiratory phase are approximately the same for all series.
On the one hand, it assumes that the tumor trajectories of all series are aligned. This assumption should be fulfilled by the application of the affine registration step which compensates for interfractional motions caused by the change of patient position on the couch. On the other hand, the tumor should be approximately at the same location and should have the same shape for a given respiratory phase. This assumption is stronger the previous one and rely on a constant tumor motion range induced by breathing.

However, assuming that the tumor location and shape in a given respiratory phase are similar for all series, binary masks drawn in the first series are good initialization of background/foreground regions for the graph cut algorithm. As mentioned previously, the complexity of the graph cut depends on the square of the number of nodes in the graph. Therefore, the entire image is not considered as search space. Instead, only the pixels belonging to the contextual region of the tumor (as defined in Figure 4.6) are considered during the graph construction. The limitation of the search space should not be problematic since we assumed that the tumor location is similar than the one on the initialization image.

The graph cut problem is solved by the max-flow/min-cut algorithm of the `maxflow` package wrapping the C++ maxflow library of Kolmogorov [41].

5.2.2 Regional and Boundary Terms

The new initialization through the binary mask of the first series makes the traditional *t-links* weights (as defied in Table 5.1) inappropriate. In the original formulation, all "marked" pixels must absolutely be contained in the region (foreground or background) in which they were drawn by the user. In our formulation, pixels "marked" as foreground are likely to belong to the tumor region, without being strictly labeled as part of the tumor. Since this is not a hard constraint, we need to adapt weights associated to *t-links*.

For this purpose, new weights assignment are presented in Table 5.2. Compared to Table 5.1, weights of "marked" pixels are given by the function R while other weights are set to 0. Intuitively, labeling a pixel initialized as foreground as part of the new foreground region is encouraged and equivalently for background region.

TABLE 5.2: New weight of edges between a non-terminal node p and other nodes depending of the type of link. All pixels belonging to the initial tumor binary mask are labeled as part of the foreground $\mathcal{O}$ while other are labeled as part of the background $\mathcal{B}$. In this case, $\mathcal{P} = \mathcal{O} \cup \mathcal{B}$ and $\mathcal{O} \cap \mathcal{B} = \emptyset$.

Link	Weight	Initialization
n-link: $\{p, q\}$	$\lambda \times B(I_p, I_q)$	$\forall p$ if $q \in \mathcal{N}_p$
t-link: $\{p, s\}$	$R(I_p, A_p)$ 0	$p \in \mathcal{O}$ $p \notin \mathcal{O} \Leftrightarrow p \in \mathcal{B}$
t-link: $\{p, t\}$	$R(I_p, A_p)$ 0	$p \in \mathcal{B}$ $p \notin \mathcal{B} \Leftrightarrow p \in \mathcal{O}$

Regional Term

The regional term definition follows the one of the traditional approach used to perform graph cut in computer vision and is given by Equation 5.3.

$$B(I_p, I_q) = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\text{dist}(p,q)} \tag{5.3}$$

where σ is a positive factor that adjusts the penalty of assigning different labels to pixels considered as similar (similarity notion depends on σ) and $\text{dist}(p,q)$ the distance between pixels. In a 4 neighborhood system for 2-dimensional segmentation, this distance is always equal to 1 (a node is not linked to itself).

Boundary Term

Boundary term computation relies on the definition of function R introduced above. Traditionally, this function is based on negative log-likelyhoods as formulated in Equation 5.4.

$$R(I_p, A_p) = \begin{cases} -\ln\left(\Pr\left(I_p|\mathcal{O}\right)\right) & A_p = 1 \Leftrightarrow p \in \mathcal{O} \\ -\ln\left(\Pr\left(I_p|\mathcal{B}\right)\right) & A_p = 0 \Leftrightarrow p \in \mathcal{B} \end{cases} \tag{5.4}$$

where for a given pixel p, I_p is its intensity, A_p its label given by the initialization and $\Pr\left(I_p|\mathcal{O}\right)$ and $\Pr\left(I_p|\mathcal{B}\right)$ the probabilities that this pixel belongs to foreground and background models, respectively.

Foreground and background models are computed for each scan with the labeling of the pixels of the corresponding scan of the first series within the contextual region of the tumor. The histogram of the foreground and background regions are both approximated by a Gaussian Mixture Model (GMM) with 3 components as shown in Figure 5.3. Probabilities in Equation 5.4 are a posteriori probabilities of an intensity value knowing foreground and background models. GMMs are implemented thanks to the `GMM` class of the `scikit-learn` package that uses `fit` function to estimate model parameters with expectation-maximization algorithm and the function `score` to compute the a posteriori log probabilities.

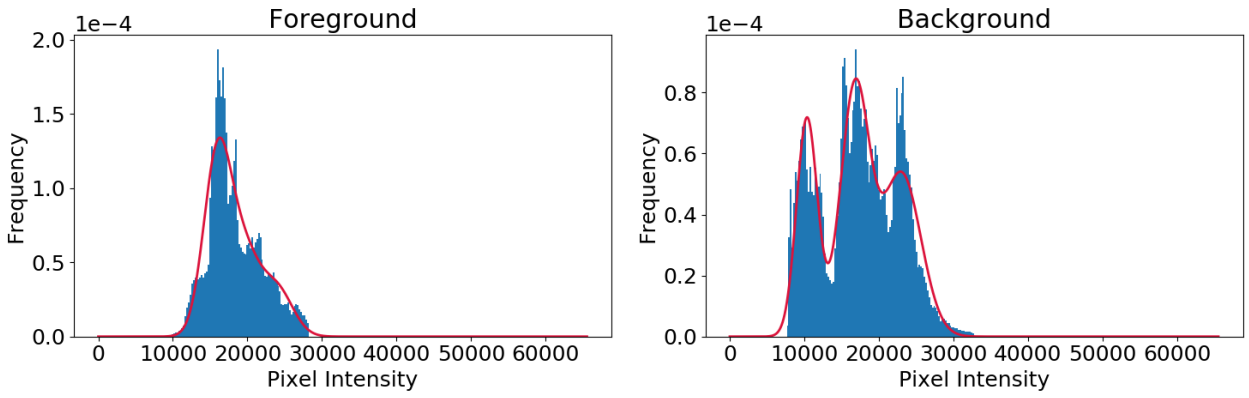


FIGURE 5.3: GMMs (red) of foreground and background regions for the computation of *t-links* weights in the first scan of the first series for patient 1. Foreground and background regions are represented by their histogram (blue).

5.2.3 Parameters

Finally, the graph cut formulation proposed here relies on two parameters, whose optimal values will be empirically studied in the following section :

Parameter 1: λ **-** A positive factor balancing the regional and boundary penalties. It adjusts the level of dependence to the initialization compared to the pixels intensities observed in the new image. The larger the value of λ, the less the result relies on the binary mask provided as initialization. In such cases, the result is driven mainly driven by the intensities in the image.

Parameter 2: σ **-** A positive factor adjusting the notion of how well are two pixels similar or dissimilar when assigning boundary constraints. The larger σ, the larger the penalty associated to labeling two neighbor "similar" pixels into distinct regions.

5.3 Results

The objective of this analysis is to assess whether a given set of values of the λ and σ parameters can produce a better tumor delineation than the binary mask given as initialization. As mentioned previously, we suppose that the tumor location and morphology in a given respiratory phase are approximately the same for all series. Therefore, the initial binary mask should give good information about the tumor delineation while not being perfect. The aim of graph cut, is thus to correct for misalignment of this mask and for shape changes of the tumor compared to the initial knowledge.
In order to quantify how good the match between evaluated tumor contour and ground truth is, we need first to define an evaluation metric. Note that tumor delineation was provided for the whole data set and serves as ground truth.

5.3.1 Evaluation of the Segmentation Result

In addition to visual inspection of the result, several metrics assess quantitatively the accuracy of a segmentation compared to the ground truth [58]. Some of them rely on computation of overlapping regions, eventually normalized by a factor. Most famous overlap-based metrics are Dice and Jaccard similarity coefficients. Other are based on distances between the contours like Hausdorff distance or Average/maximum/minimum symmetric surface distances. The choice of the metric is application-dependent since information carried by each of them is different.

In the context of this thesis, the Dice similarity coefficient (Dice) is the one that has be chosen. This is a typical value used to asses segmentation results in the medical field. For two sets, Dice value is given by the Equation 5.5.

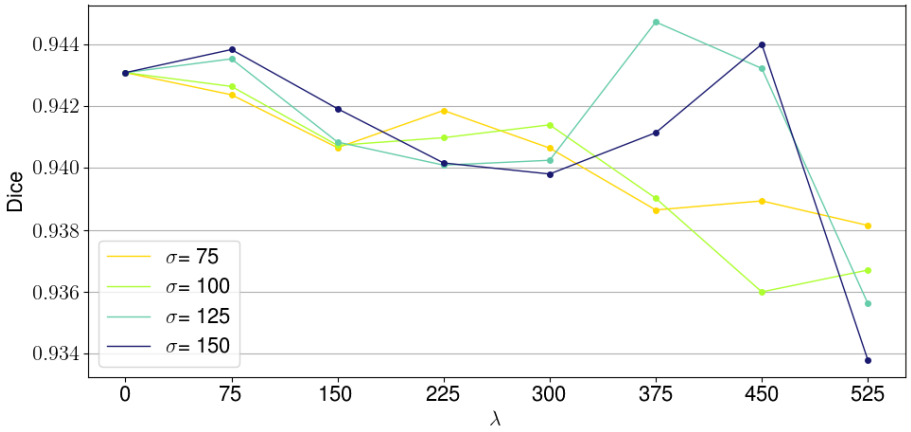
$$\text{Dice} = \frac{|A \cap B|}{1/2 * (|A| + |B|)} = \frac{2|A \cap B|}{(|A| + |B|)} \tag{5.5}$$

where A and B and the two sets that correspond to all pixels belonging to segmented regions A and B respectively. Dice quantifies how well the two regions A and B are overlapping by associating a value between 0 and 1 with the result. A value of 0 is attributed to disjoints sets while a value 1 corresponds to two perfectly overlapping sets.

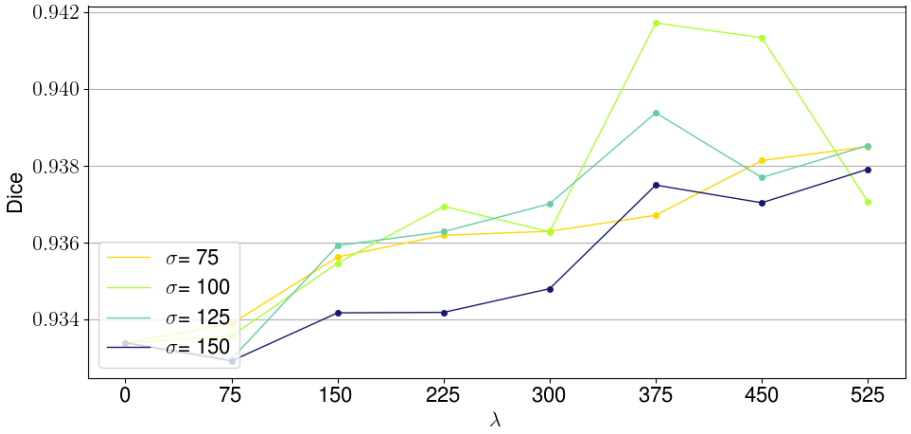
5.3.2 Presentation of the Results Depending on the Parameters

In order to assess the impact of the parameters, graph cut segmentation has been performed for different values of λ and σ. On the one hand, λ varies between 0 and 575 using steps of 75. The result given with $\lambda = 0$ consists in the Dice score obtained when the estimated tumor delineation is the one provided by initialization, i.e. the binary mask of the first series for the same respiratory phase. This is our benchmark case to assess if some values of the parameter could enhance this initial segmentation. On the other hand, σ varies between 75 and 150 using steps of 25.

The objective of the analysis is to highlight patterns in the results depending on the two parameters and in particular, identify a set of parameters that produces a better delineation than the benchmark case. Out of the 90 scans (3 patients $\times$ 3 series $\times$ 10 scans), Dice results obtained two of them are shown in Figure 5.4. For the first scan (see Figure 5.4(a)) Dice seems to decrease as λ increase. The opposite trend can be observed in the second scan (see Figure 5.4(b)). At the first glance, no global pattern seems to show up out of these results.



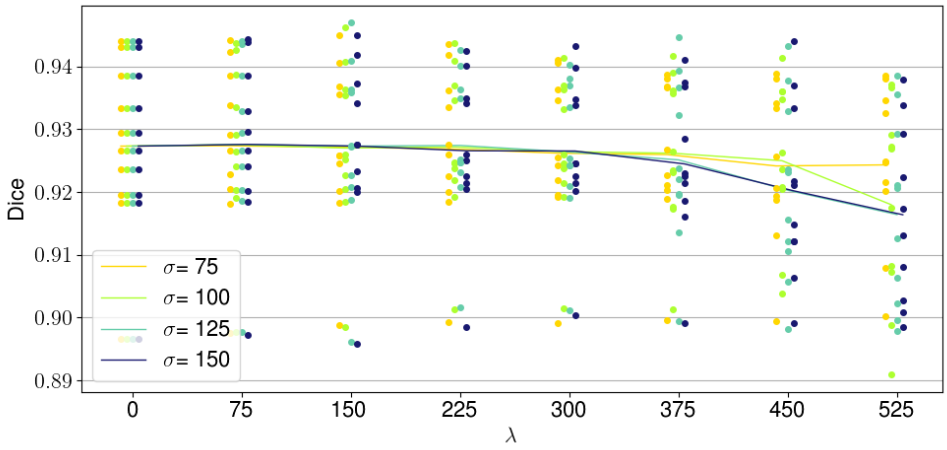
(a) Scan 1 of series 2.



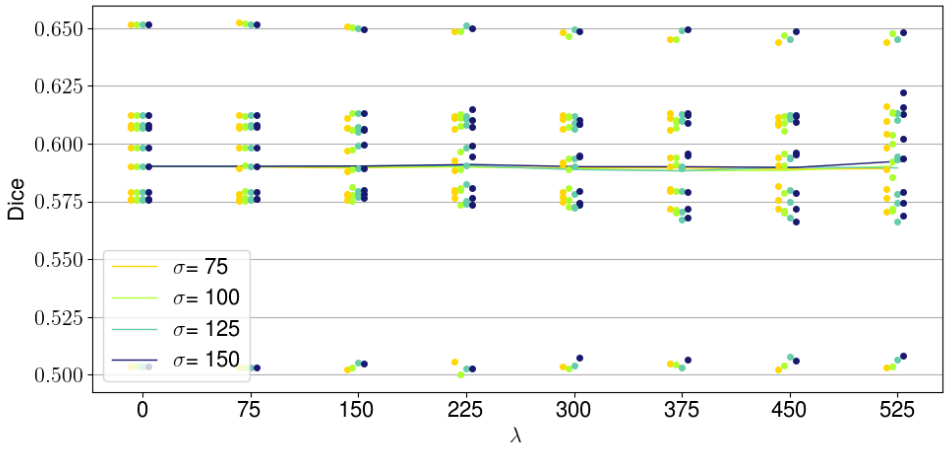
(b) Scan 2 of series 2.

FIGURE 5.4: Dice scores obtained on two scans for different values of λ and σ on the patient 2.

In order to facilitate the interpretation, results of all 10 scans of a given series have been grouped on a single graph presented in Figure 5.5. In these graphs, each dot denotes for a single experiment and solid lines are the mean values of all experiments made with the same set of parameters.
First, the impact of the parameters seems to be averaged across the repetitions of the experiments. For series 2 (see Figure 5.5(a)), Dice score slightly decreases as λ increases. This tendency is less pronounced for smaller values of σ. For the series 3 (see Figure 5.5(b)), the Dice value is relatively constant despite the variation of λ and σ.
Second, the Dice scores obtained on series 2 (around 0.925) are much higher than those obtained on series 3 (around 0.578). Since this is also valid for the benchmark case ($\lambda = 0$), it shows that the tumor contour initialization is closer to the ground truth for all scans of series 2 than series 3. At this stage of the analysis, no particular set of parameters seems to obtain higher Dice score than the benchmark cases.



(a) Series 2.



(b) Series 3.

FIGURE 5.5: Summary of Dice scores obtained for one series on the patient 2. Each dot is the Dice score for a single experiment and solid lines are the mean values of all experiments made with the same σ depending on the λ parameter.

Dice scores obtained for all scans of all three series for a given patient are represented in Figure 5.6. A quick glance at the mean value of the Dice score (solid lines) is enough to get a hint of the inability of the graph cut method to enhance the segmentation given as initialization ($\lambda = 0$). The mean score of the segmentation is about 0.7 and none of the set of parameters tested could improve this score globally.

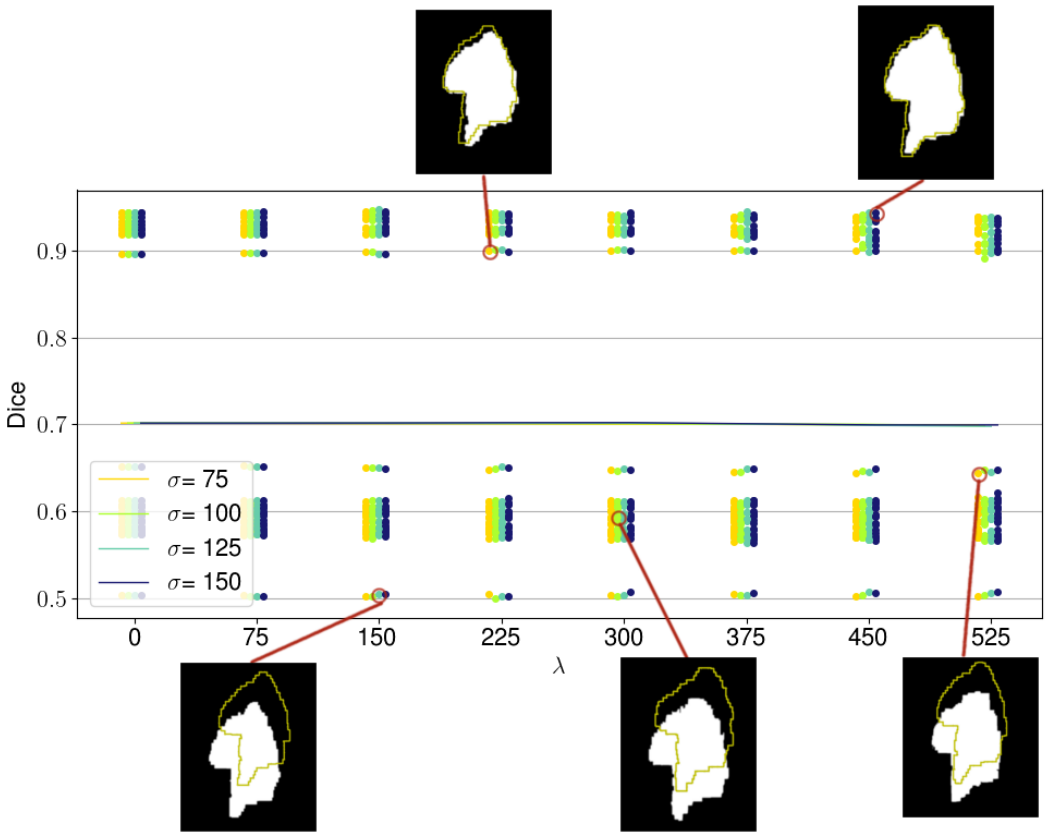


FIGURE 5.6: Summary of Dice score obtained on the patient 2. Each dot is the Dice score for a single experiment and solid lines are the mean values of all experiments made with the same σ depending on the λ parameter. Some images showing the graph cut segmentation into background (black) and foreground (white) regions are given as example. On these images, the yellow contour corresponds to the ground truth tumor delineation.

Note that all dots located in the same horizontal line correspond to the same scan. For example, samples whose Dice score is around 0.5 correspond to the scan 7 of the series 4, samples around 0.67 to scan 3 of series 3 and samples around 0.89 to scan 7 of series 2. All samples around 0.9 and above belongs to the series 2. This indicates that the average Dice score achieved on one scan depends mainly on the match between the initial tumor delineation and the ground truth.

Results obtained on the patient 2 have been presented. Similar conclusions can be taken out the segmentation results obtained for the two other patients. However, mean Dice score are different. Outcome of graph cut segmentation is a mean Dice score of 0.8 for patient 1 and 0.35 for patient 3.

5.4 Discussion

Outcome of the experiments. In its original formulation presented in Section 5.1, the graph cut algorithm requires initial labeling of some pixels belonging to foreground and background regions on which hard classification constraints are applied. This time consuming initialization, involving an interaction with an expert, constitutes a major disincentive to use of graph cut method for real-time segmentation tasks.
In this thesis, a novel formulation of the problem addresses this issue by substituting the manual drawing of the user with a binary mask defining foreground and background regions based on some prior knowledge (approximate location and morphology of the tumor) acquired previously in another series of scans. In the purpose of integrating correctly this initialization, the computation of the weights associated to *t-links* has been adapted as presented in Section 5.2. We hoped that this formulation would enable the graph cut method to enhance the initial segmentation of the binary mask by both relying on this prior knowledge and the actual image in which the result has to be computed.

The results showed that the graph cut method was not able to provide more information on the true delineation of the tumor than the one carried by the prior information. It has been shown that parameters have an influence on the result (see Figure 5.4) but no "optimal" parameter values could bring a global enhancement of the segmentation performed on all scans of a given patient (see Figure 5.6). In light of this disappointing outcome, the proposed changes to the initial formulation of the graph cut are not sufficient to accomplish lung tumor segmentation in fluoroscopic images.

However, these poor results should be nuanced as the segmentation task in fluoroscopic images is an extremely difficult problem. Due to the nature of fluoroscopic images, the superposition of tissues and organs results in smooth and blurred transitions between neighboring regions. This has a direct impact on the weights of *n-links*. As neighboring regions are not characterized by large and abrupt changes in gray values of the pixels, the information carried by *n-links* is less crucial for the segmentation than for an image with high contrast and sharp transitions. Since all pixels within a transition have only slightly different intensities, costs associated to these *n-links* are not so low compared to *n-links* weights within a homogeneous region. The costs associated with the breaking these links located in the transition regions between two tissues are then too high with respect to the costs of assigning two distinct labels to pixels within a same tissue.
Based on this observation, hoping that information of *n-links* (and then gray level of pixels in the image) could rectify the initial segmentation given by the prior binary mask was maybe too ambitious. But applied in another type of images with higher contrast and sharper transitions, this assumption could potentially be valid.

Time Complexity. Although, graph cut segmentation remains a powerful method. As mentioned previously, the result is obtained with a time complexity in $\mathcal{O}\left(mn^2|\mathcal{C}|\right)$, with n the number of nodes, m the number of edges and $|\mathcal{C}|$ the size of the cut.
Average execution time of graph cut segmentation for each patient are listed in the Table 5.3. These values, that should not be taken as absolute, indicate the possible application of graph cut in a real-time context. Note that the time needed to initialize the graph is not covered by the execution times provided below. The initialization lasts about 6 seconds for patient 1 up to 47 seconds for patient 3, but is subject to many improvements.

TABLE 5.3: Average execution time of graph cut segmentation algorithm for the three patients depending on the size of the search space.

Patient	Search space size (pixels)	Average execution time (ms)
Patient 1	378×442	120
Patient 2	840×1364	650
Patient 3	584×546	240

Possibilities of improvement. According to the results presented previously, some changes have to be made in order to improve the lung tumor segmentation performed by the graph cut method in fluoroscopic images. A promising way of improvement is presented in [59]. The idea is to integrate the prior information about the tumor shape and location to restrict the search space containing the solution within two bands as illustrated in Figure 5.7. In our context, the search space could be, for example, the tumor delineation given by the initial binary mask extended by some margins in a such way that the tumor contour in the current scan is contained in this search space.

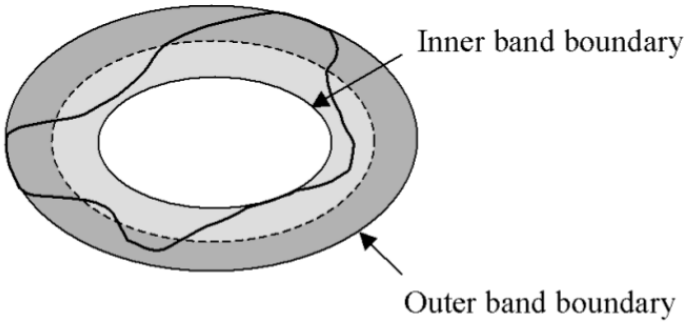


FIGURE 5.7: Restriction of the search space containing the segmentation into a elliptical shape (from Slabaugh and Unal ([59]).

As with our method, results of such an approach are highly dependent on the binary mask given as initialization. The assumption that the tumor location and shape in a given respiratory phase were approximately the same for all series is all the more crucial with this new approach since the search space would be restricted. Unfortunately, this assumption is not valid in practical cases. Even if we could predict perfectly the current respiratory phase in fluoroscopic videos, the tumor position within this phase depends on breathing characteristics, such as amplitude, that can vary from session to session.
With our method, we expected that the graph cut could recover from a sightly shifted initial mask, making this assumption less critical. With this potential approach, the design of the search space could encompass for these uncertainties while providing a search space as small as possible. In the context on real clinical applications, this trade-off represents a huge challenge.

Chapter 6

Lung Tumor Tracking

This chapter develops a solution to perform lung tumor tracking in DRRs. As for the segmentation task presented in Chapter 5, the process is performed in the DRRs obtained with the image quality enhancement method detailed in Chapter 4. As already mentioned in Chapter 3, segmentation and tracking are two different tasks despite the fact that they both aim to localize an object. On the one hand, segmentation methods partition an image into multiple components by defining all pixels belonging to each of these components. On the other hand, tracking methods identify the image region which is the most likely to contain an object for which prior appearance clues have been provided.

Probably the easiest way to perform tracking is to select a rectangular window containing the object of interest in the first frame and find, in all subsequent frames, the location minimizing the difference between the reference rectangular window, i.e. the template. This simple strategy is known as template matching.
However, such a simple method is not robust to appearance changes of the target. In the case of lung tumor tracking in fluoroscopic sequences, the global intensity of the image varies during breathing. The images are brighter during the inhale phase than during the exhale phase. Beside intensity variations, tumor shape changes also affect the robustness of template matching. These can occur between fractions due to a reduction of the tumor mass induced by the treatment or within a fraction due the breathing motion. This issue has been addressed in [52] (developed in Section 3.3) by the use of multiple templates to represent the tumor in different breathing phases and account for the appearance changes. With this technique, the authors performed lung tumor tracking in fluoroscopic images with an accuracy within 3 millimeters. However, the templates construction requires a set of fluoroscopic scans with tumor contour drawn by an expert.
Moreover, testing all possible locations of the template in a new image is time consuming. Without any improvement, the search phase can be too slow for real-time applications.

Kernel mean-shift tracking [60] brings some modifications to template tracking in order to improve robustness against targets with sightly varying appearance while accelerating considerably the search mechanism. First, the template is represented by its histogram instead of the initial representation in the pixel space. Second, the search mechanism is based on the mean-shift algorithm that converges in a few iterations. For these reasons, this approach has been chosen to perform the tracking of lung tumors in fluoroscopic images. The method is presented more extensively in the next sections.

6.1 Theoretical Background

Kernel mean-shift tracking, introduced by Comaniciu et al. in [60] and applied in [53] to lung tumor tracking in fluoroscopic images, provides a framework to perform robust tracking of non-rigid objects. This task can be divided into two steps :

Target modeling. Generally, object tracking relies on prior clues about target appearance. This step requires a manual contouring of the target region drawn by an expert in either images acquired before the experiment or in the first frame of the video in which the target is tracked. In the context of kernel tracking, the model of the target is not an image but a projection of this image on a feature space of lower dimensionality. Usually, an histogram with a restricted number of bins is a good representation of the target in a low dimensional space for tracking tasks.

Target tracking. This iterative process consists in finding the best region in all subsequent frames matching the target model. This step constitutes the main contribution of the work of Comaniciu et al. by the use of the mean-shift algorithm [61]. The mean-shift algorithm describes a framework to perform gradient optimization in order to solve the search problem. Compared to an exhaustive search, the time required to find the new location of the target in subsequent frames is considerably reduced. However, the gradient optimization is likely to converge toward a local maximum, if any.

6.1.1 Target Modeling

The aim of target modeling is to represent the target as a 1-dimensional feature vector. Originally, the target is expressed in the image space of dimension $n_x \times n_y$ where n_x and n_y are the number of pixels along x and y axes, respectively. This target is then mapped into a 1-dimensional feature space of size m. In this new space, the feature vector modeling the target is named $\hat{\boldsymbol{q}}$.
During the tracking process, this target model is compared to target candidates models located at positions $\boldsymbol{y}$, i.e. any $n_x \times n_y$ window centered in $\boldsymbol{y}$ that could potentially contain the target. All these target candidates are also mapped in the m-dimensional space in which they are denoted by $\hat{\boldsymbol{p}}(\boldsymbol{y})$. These notations are summarized in Equation 6.1.

$$\begin{aligned} \text{Target Model} \quad & \hat{\boldsymbol{q}} = \{\hat{q}_u\}_{u=1\ldots m} \\ \text{Target Candidate Model} \quad & \hat{\boldsymbol{p}}(\boldsymbol{y}) = \{\hat{p}_u(y)\}_{u=1\ldots m} \end{aligned} \tag{6.1}$$

Target Model : A Weighted Histogram

Multiple mappings exist, but the one that has been used by [60, 53] is a m-bin weighted histogram of the original $n_x \times n_y$ window. The computation of this histogram relies on the function $b : \mathcal{R}^2 \rightarrow \{1\ldots m\}$, which associates to a pixel at normalized location $\boldsymbol{x}_i^*$ (with the center of the rectangle at the origin $\mathbf{0}$) the index of the bin in the m-histogram. For a m-bin histogram, the definition of the function b is given in Equation 6.2.

$$b(\boldsymbol{x}_i^*) = \text{floor}\left(\frac{m\ I(\boldsymbol{x}_i^*)}{2^L}\right) + 1 \tag{6.2}$$

where $I(\boldsymbol{x}_i^*)$ is the intensity of the pixel at location $\boldsymbol{x}_i^*$, floor a function that rounds a fraction to the smallest or equal integer and L the number of bits used to encode the image (here $L = 16$ bits).

An unweighted histogram $\hat{\boldsymbol{q}}_{\text{unweighted}}$ is then a normalized sum of the values returned by b over all pixels as depicted by Equation 6.3. First, all pixels are assigned to a bin of the histogram and added to the actual number of pixels already belonging to the considered bin. Then, the count of pixels falling in each bin is normalized by the total number of pixels considered.

$$\hat{\boldsymbol{q}}_{\text{unweighted},u} = \frac{1}{n_x \times n_y} \sum_{i=1}^{n_x \times n_y} \delta\left[b(\boldsymbol{x}_i^*) - u\right] \quad \text{with } u = 1...m \tag{6.3}$$

where δ is the Kronecker delta function defined by $\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases}$

However, the feature vector used in this context is not an unweighted histogram but a weighted one. Formula of the weighted histogram feature vector is given by Equation 6.4.

$$\hat{q}_u = C \sum_{i=1}^{n_x \times n_y} k\left(\left\|\frac{\boldsymbol{x}_i^*}{\boldsymbol{h}}\right\|^2\right) \delta\left[b(\boldsymbol{x}_i^*) - u\right] \quad \text{with } u = 1...m \tag{6.4}$$

where k is a kernel profile, $\left\|\frac{\boldsymbol{x}_i^*}{\boldsymbol{h}}\right\|$ the normalized distance between $\boldsymbol{x}_i^*$ and the center of the rectangular region in the original space, δ the Kronecker delta function as defined previously and C is a normalization constant that imposes $\sum_{u=1}^{m} \hat{q}_u = 1$ given in Equation 6.5.

$$C = \frac{1}{\sum_{i=1}^{n_x \times n_y} k\left(\left\|\frac{\boldsymbol{x}_i^*}{\boldsymbol{h}}\right\|^2\right)} \tag{6.5}$$

The weighting process is performed through a multiplication of the kernel profile k and the target region in the original space of pixels (see Figure 6.1(a)). Note that the origin $\mathbf{0}$ of the rectangular region is located at the center of the image and not in the upper left corner. All pixel locations ($\boldsymbol{x}_i^*$) and distances are then expressed with respect to this center.
The kernel is a 2-dimensional function centered in $\mathbf{0}$ that is convex and monotonically decreasing. As suggested by Comaniciu [53], the kernel used is a Gaussian whose formula is given in Equation 6.6 and represented figure 6.1(b).

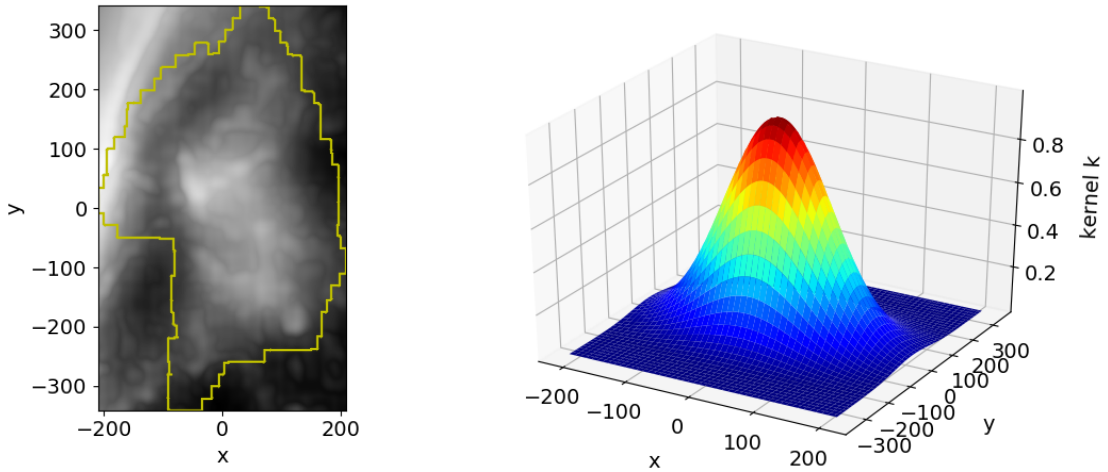
$$k(x) = \exp\left(-\frac{x}{2}\right) \tag{6.6}$$

where x is a normalized distance scaled by the $\boldsymbol{h}$ parameter introduced in Equation 6.4.

Note that other kernels can also be used, as the Epanechnikov kernel profile for example. This kernel is defined in Equation 6.7.

$$k(x) = \begin{cases} 1 - x & \text{if } x < 1 \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

where x is a normalized distance scaled by the $\boldsymbol{h}$ parameter.
Thanks to its linear derivative, this kernel is often used in kernel mean-shift tracking, which is based on a gradient optimization scheme. More precisely, the g term in the computation of the mean-shift vector presented further in Equation 6.11 becomes a constant.



(a) Intial rectangular region drawn around the tumor whose countour is given by the yellow curve.

(b) Gaussian kernel used to weight histogram of the rectangular region. Pixels located near the center have a bigger impact on the histogram than those near the borders. The parameter $\boldsymbol{h}$ is fixed so that each corner is at a normalized distance 5 from the center.

FIGURE 6.1: Example of rectangular region and associated kernel for the patient 2 (first scan of the series 1).

The aim of this kernel is to emphasize the importance of pixels located near the center of the rectangular region compared to those close to the borders. Intuitively, pixels close to the center are more likely to represent the tumor than the background. They should then be more representative in the feature vector.
Behind this first intuitive reason, this kernel brings some smoothness during the computation of the gradient of the similarity between target and target candidates models. Since outer pixels are less taken into account, a small shift of the target candidate in the original space results in a smaller difference of the similarity value (defined in Section 6.1.2) between the target and candidate models. This phenomenon is represented in Figure 6.2. By this way, two weighted histograms computed around sightly shifted centers will be almost similar. This "quasi invariance to translation" property, induced by the decreasing kernel profile, makes possible the application of gradient optimization in order to maximize a similarity metric between two histograms.

Target Candidate Model : Also A Weighted Histogram

The target candidates, i.e. any $n_x \times n_y$ window centered in $\boldsymbol{y}$, need also to be projected in the m-dimensional feature space in which they are denoted by $\hat{\boldsymbol{p}}(\boldsymbol{y})$. The projection is the same as the one applied to the target model and is defined in Equation 6.8.

$$\hat{p}_u(\boldsymbol{y}) = C \sum_{i=1}^{n_x \times n_y} k\left(\left\|\frac{\boldsymbol{y} - \boldsymbol{x}_i}{\boldsymbol{h}}\right\|^2\right) \delta\left[b(\boldsymbol{x}_i) - u\right] \quad \text{with } u = 1...m \tag{6.8}$$

where all definitions of k, C and δ have the same meaning as in Equation 6.4. The only difference is that pixels of coordinates $\boldsymbol{x}_i$, which belong to the rectangular region representing the target, are centered around the location $\boldsymbol{y}$ and not $\boldsymbol{0}$ as for the target model.

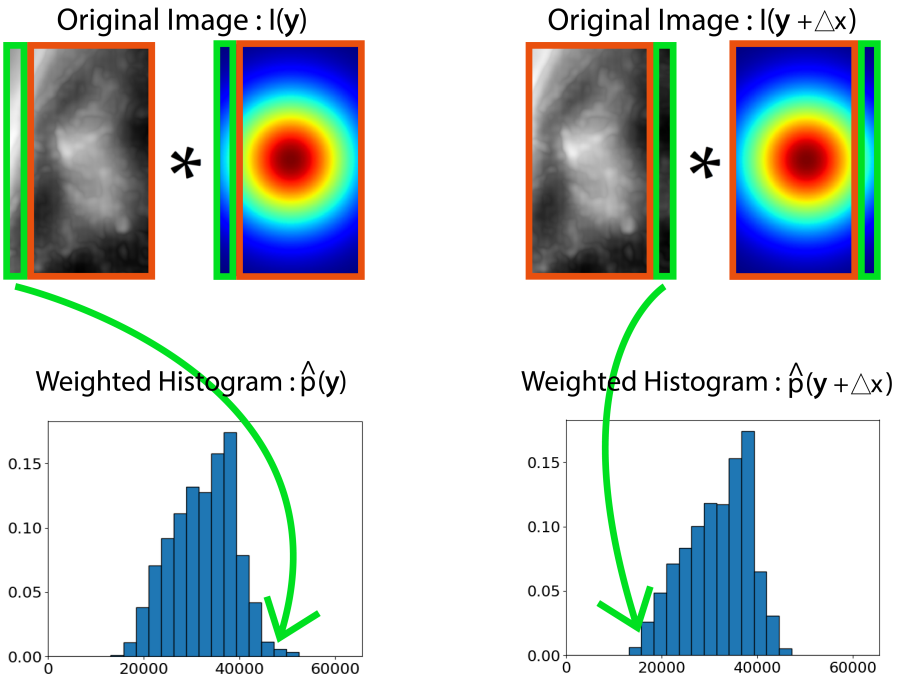


FIGURE 6.2: Representation of the "quasi invariance to translation" property of the kernel. Pixels in the orange regions are common to the two images (original and horizontally shifted version of it) and pixels in green rectangles contribute to the difference between the two weighted histograms. Weights associated to these pixels (given by the kernel k) are low compared to those associated to other pixels.

6.1.2 Target Tracking

The tracking task can be formulated as the problem of finding the location of the target candidate whose feature vector $\hat{\boldsymbol{p}}(\boldsymbol{y})$ is the most similar to the one of the target $\hat{\boldsymbol{q}}$. For this purpose, a similarity metric has to be defined in order to compare target model with target candidates models. A search mechanism is also needed to select which target candidates will be compared to the target. The aim is to find the best target candidate, i.e. the one maximizing the similarity metric with respect to the target.

Similarity Metric : The Bhattacharyya Coefficient

The similarity between the two feature vectors $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}(\boldsymbol{y})$, i.e. the similarity between the two corresponding weighted histograms, can be measured with the Bhattacharyya coefficient defined in Equation 6.9. This metric is denoted by $\rho[\hat{\boldsymbol{p}}(\boldsymbol{y}), \hat{\boldsymbol{q}}]$.

$$\rho[\hat{\boldsymbol{p}}(\boldsymbol{y}), \hat{\boldsymbol{q}}] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\boldsymbol{y})\, \hat{q}_u} \tag{6.9}$$

Geometric interpretation of this coefficient is *the cosine distance of the angle between the m-dimensional unit vectors* $\left(\sqrt{\hat{p}_1}, \ldots \sqrt{\hat{p}_m}\right)^T$ *and* $\left(\sqrt{\hat{q}_1}, \ldots \sqrt{\hat{q}_m}\right)^T$ [60]. The Bhattacharyya coefficient has the important property of being a smooth function of $\boldsymbol{y}$. This will have an impact on the search mechanism which is a gradient-based optimization process.

Search Mechanism : The Mean-Shift Algorithm

Given an initial target location $\hat{\boldsymbol{y}}_0$, the Bhattacharyya coefficient at location $\hat{\boldsymbol{y}}$ can be approximated linearly by a Taylor expansion around the values $\hat{p}_u(\hat{\boldsymbol{y}}_0)$ as expressed in Equation 6.10.

$$\begin{aligned}\rho[\hat{\boldsymbol{p}}(\boldsymbol{y}),\hat{\boldsymbol{q}}] &\approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{\boldsymbol{y}}_0)\,\hat{q}_u} + \frac{1}{2}\hat{p}_u(\hat{\boldsymbol{y}})\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\boldsymbol{y}}_0)}} \\ &\approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{p}_u(\hat{\boldsymbol{y}}_0)\,\hat{q}_u} + \frac{C}{2}\sum_{u=1}^{m} w_i k\left(\left\|\frac{\boldsymbol{y}-\boldsymbol{x}_i}{\boldsymbol{h}}\right\|^2\right)\end{aligned} \tag{6.10}$$

where $w_i = \sum_{u=1}^{m}\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\boldsymbol{y}}_0)}}\delta\left[b(\boldsymbol{x}_i)-u\right]$.

The first term of the Equation 6.10 is independent on $\boldsymbol{y}$. The second term represents a density estimation computed with the kernel profile at location $\boldsymbol{y}$. This function needs to be maximized in order to find the location $\boldsymbol{y}$ of the target candidate corresponding to a (local) maximum of the Bhattacharyya coefficient.
This maximization is achieved by fixed point iteration algorithm. Starting from an initial point, the trajectory of this point toward the local maximum is given in the direction maximizing the density and stops when two consecutive iterations converge toward the same point. The iterative procedure is given by Equation 6.11. The value $\hat{\boldsymbol{y}}_{j+1} - \hat{\boldsymbol{y}}_j$ is called the mean-shift vector and is represented in Figure 6.3.

$$\hat{\boldsymbol{y}}_{j+1} - \hat{\boldsymbol{y}}_j = \frac{\sum_{i=1}^{n_x\times n_y}\boldsymbol{x}_i w_i g\left(\left\|\frac{\boldsymbol{y}_0-\boldsymbol{x}_i}{\boldsymbol{h}}\right\|^2\right)}{\sum_{i=1}^{n_x\times n_y} w_i g\left(\left\|\frac{\boldsymbol{y}_0-\boldsymbol{x}_i}{\boldsymbol{h}}\right\|^2\right)} - \hat{\boldsymbol{y}}_j \tag{6.11}$$

where $g(x) = k'(x) = \frac{1}{2}\exp\left(-x/2\right)$.

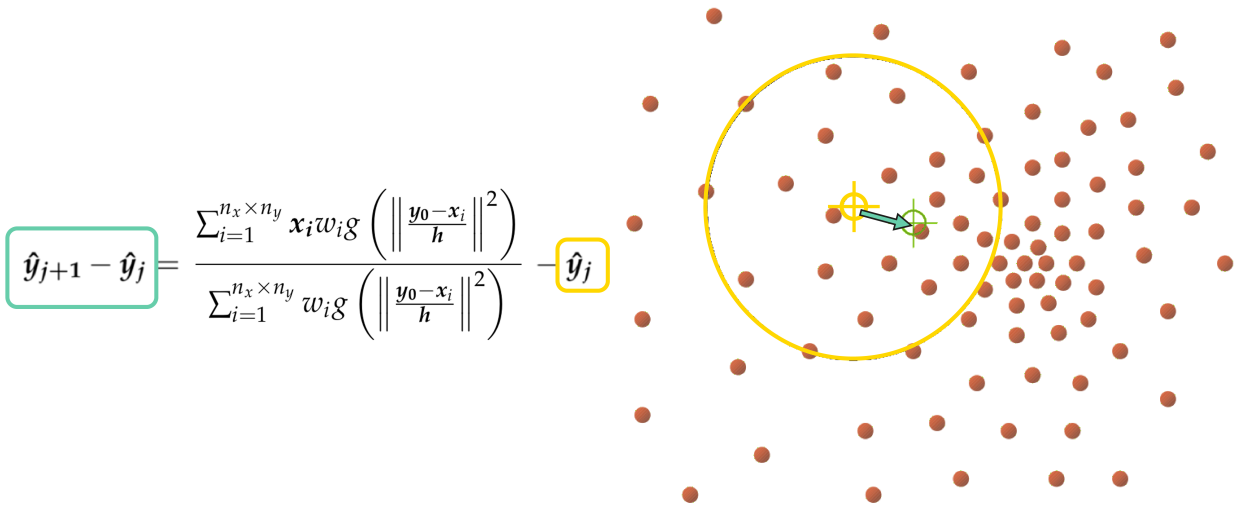


FIGURE 6.3: Illustration of the geometric meaning of the mean-shift vector defined in Equation 6.11 (based on the slides of Collins [62]).

According to Comaniciu et al. [60], the Bhattacharyya maximization problem can be solved by the mean-shift algorithm [61]. The aim of the method is to integrate the computation of a vector pointing toward the direction of maximum increase of the density in a recursive framework (see algorithm in Section 6.2). However, the expression of the Bhattacharyya coefficient maximization as a density maximization problem is not trivial and generally not explained in the literature.
The intuition behind the transformation of the Bhattacharyya coefficient maximization into a density maximization task is represented in Figure 6.4. This figure illustrates the estimation of a probability density function (pdf) generated out of data samples. In this context, we can consider that the two axes of the 2-dimensional space containing the data samples (right part) are feature 1: $\sqrt{\hat{p}_{u1}(\boldsymbol{y})\hat{q}_{u1}}$ and feature 2: $\sqrt{\hat{p}_{u2}(\boldsymbol{y})\hat{q}_{u2}}$. If the feature space of $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}(\boldsymbol{y})$ is a 2-bin weighted histogram, the sum of the two features is the Bhattacharyya coefficient for a target candidate located in $\boldsymbol{y}$. By computing these two features for multiple values of $\boldsymbol{y}$, we can sample the space represented in the right part of the figure. In this space, dense regions correspond to target candidates similar to the target.

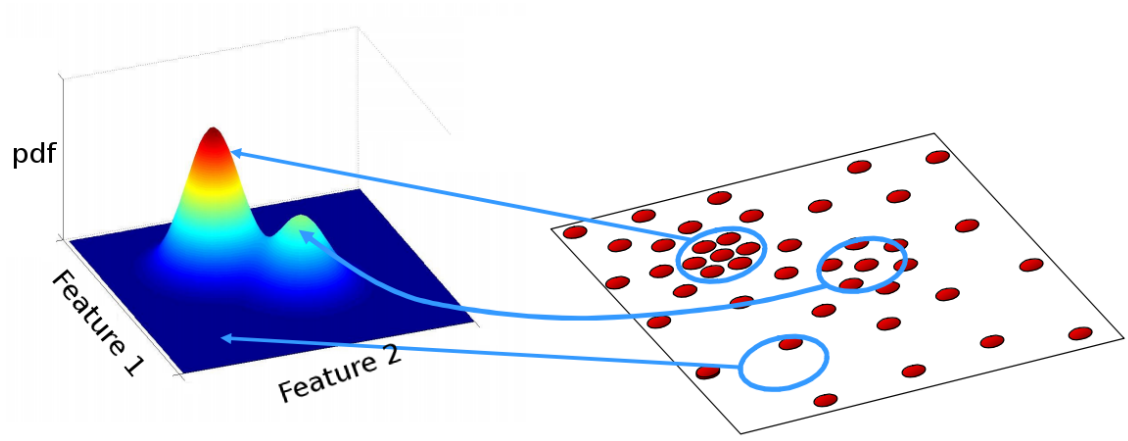


FIGURE 6.4: Generation of a pdf sampled from real data (based on the slides of Collins [62]).

Thanks to the kernel operator weighting the histograms, the Bhattacharyya coefficient is a smooth function of $\boldsymbol{y}$. Therefore a small shift of $\boldsymbol{y}$ results in close points in the feature space. In this context, finding the region maximizing the density of the cloud of points or maximizing the pdf resulting of this data sampling are two equivalent tasks. Starting from an intitial location of the target candidate $\boldsymbol{y}_0$, the mean shift vector (as evaluated by Equation 6.11) points toward the direction of higher density. Equivalently, the mean-shift vector points toward the direction maximizing the Bhattacharyya coefficient and can be interpreted as a gradient estimation of the pdf.

The mean-shift algorithm is guaranteed to converge to a local maximum if kernel k has a convex and monotonically decreasing profile [61]. These properties of the kernel profile makes it smooth relative to $\boldsymbol{y}$. According to the computation of the target and target candidates models by Equations 6.4 and 6.8, the Bhattacharyya coefficient is also smooth relative to $\boldsymbol{y}$. Then, a differentiable kernel yield to a differentiable Bhattacharyya coefficient and an efficient gradient optimization process.

6.2 Proposed Method

The method used to track lung tumors in fluoroscopic images is a kernel-based approach obtained from the work of Zhang et al. [53]. The major difference is that the images processed in this thesis are enhanced via the method described in Section 4.2, whereas Zhang et al. used a classical histogram equalization (see Section 4.1.2). Moreover, the discussion will be pushed a little bit further. The impact of the scaling factor $\boldsymbol{h}$ (see Equation 6.4) and the dimension of the feature space m will be addressed.

6.2.1 Pseudo-code

The method used to perform lung tumor tracking in DRRs, whose quality has been enhanced with our method, is described as follow:

Step 1 - Initialize location of the fist target candidate in $\hat{\boldsymbol{y}}_{\mathbf{0}}$ and compute Bhattacharyya coefficient $\rho[\hat{\boldsymbol{p}}(\boldsymbol{y}_0), \hat{\boldsymbol{q}}]$ (see Equation 6.9).

Step 2 - Compute the weights w_i (see Equation 6.10).

Step 3 - Compute the new location of the target $\hat{\boldsymbol{y}}_{\boldsymbol{j+1}}$ based on the kernel density gradient estimation developed in Equation 6.12, which can be obtained directly out of Equation 6.11.

$$\hat{\boldsymbol{y}}_{\boldsymbol{j+1}} = \frac{\sum_{i=1}^{n_x \times n_y} \boldsymbol{x_i} w_i g\left(\left\|\frac{\boldsymbol{y_j} - \boldsymbol{x_i}}{\boldsymbol{h}}\right\|^2\right)}{\sum_{i=1}^{n_x \times n_y} w_i g\left(\left\|\frac{\boldsymbol{y_j} - \boldsymbol{x_i}}{\boldsymbol{h}}\right\|^2\right)} \tag{6.12}$$

Following the reasoning developed in Section 6.1.2, the new target candidate location is found in the direction maximizing density gradient estimation.

Step 4 - If $\rho[\hat{\boldsymbol{p}}(\boldsymbol{y_{j+1}}), \hat{\boldsymbol{q}}] < \rho[\hat{\boldsymbol{p}}(\boldsymbol{y_j}), \hat{\boldsymbol{q}}]$ then Stop.

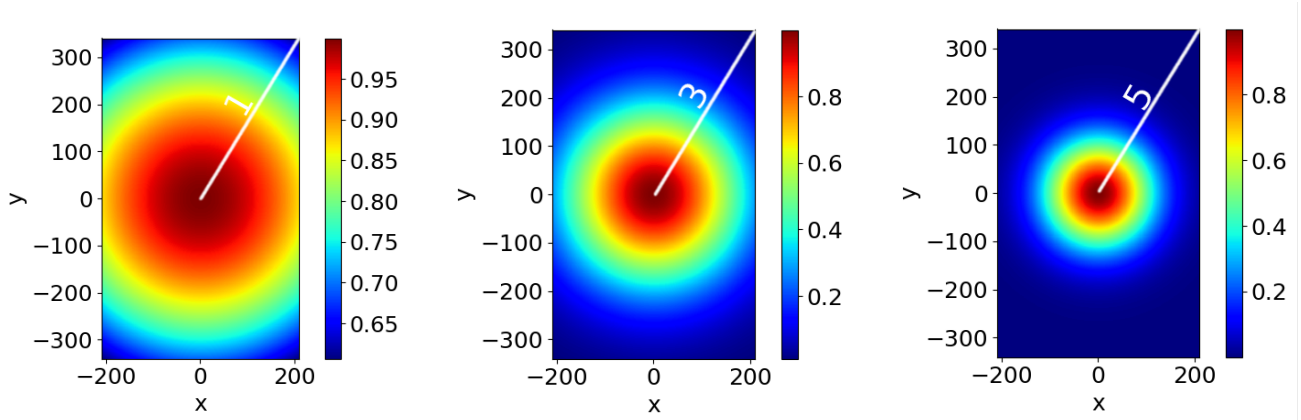
Step 5 - Iterate over steps 2 to 4 until reaching convergence: $\left\|\hat{\boldsymbol{y}}_{\boldsymbol{j+1}} - \hat{\boldsymbol{y}}_{\boldsymbol{j}}\right\| < \|1 \text{ pixel}\|$ or a maximum number of iterations: $j > 20$.

6.2.2 Parameters

This kernel tracking algorithm has two main parameters whose impact will be presented in the next sections. They are :

Parameter 1: m **-** The size of the 1-dimensional feature vector modeling the target and the target candidates (see Equation 6.1). This corresponds to the number of bins constituting the weighted histograms of the target and target candidates images.
A 25-bins weighted histogram has been used in [53].

Parameter 2: $\boldsymbol{h}$ **-** The bandwidth of the kernel profile. This parameter can be interpreted as the distance between the center of the image and each corner. The value of $\boldsymbol{h}$ is dependent on the original size of the image but we consider that $\boldsymbol{h}_d$ denotes for the vector that normalizes the distance from the center to any corner at a distance d. This constitutes an abuse of notation which simplify the interpretation of the parameter $\boldsymbol{h}$. The kernel k is represented in Figure 6.5 for patient 2 with different values of the parameter $\boldsymbol{h}$.



(a) Gaussian kernel profile k normalized with $\boldsymbol{h}_1$.

(b) Gaussian kernel profile k normalized with $\boldsymbol{h}_3$.

(c) Gaussian kernel profile k normalized with $\boldsymbol{h}_5$.

FIGURE 6.5: Representation of the Gaussian kernel profile k depending on the bandwidth parameter $\boldsymbol{h}$ for patient 2. Note that the x and y axes correspond to the normalized pixel location in the image space (as for Figure 6.1(a)).

6.3 Results

The metric used to present the results is the mean of the absolute value of the difference between the target predicted center and the ground-truth center for all scans (4 series $\times$ 9 scans) of a given patients. This distance is expressed either in pixels or in millimeters. The resolution of the tested images is 0.15 mm/pixel.

The kernel mean-shift tracking have been tested over the whole population, i.e. 9 frames $\times$ 4 series $\times$ 3 patients (the first frame being used as initialization) for multiple values of the parameters m and $\boldsymbol{h}$.
On the one hand, the evolution of the number of bins constituting the weighted histograms m does not follow a given pattern. Values have been chosen "on the fly" in order to compare the standard value used in the literature ($m = 25$) to other values in other orders of magnitude (10, 50, 100, 500, 1000). Note that DRRs are encoded in 16 bits, making $m = 500$ and $m = 1000$ acceptable values for the m parameter. With small values of the m parameter, the kernel representation of the target is not sensitive to little intensity changes. Therefore, the effect of noise is mitigated but distinct regions with quite similar intensities may not be discriminated making the kernel less informative. On the contrary, larger values of m result in histograms affected by small changes in intensity. The kernel is then more informative but also likely to suffer from a lack of regularization property. However, large values of m slow down the computation of the Bhattacharyya coefficient which is undesirable for real-time applications.
On the other hand, five kernel bandwidths applied on the target or target candidates in the image space have been tested. Their definition follows explanation given in Section 6.2. Therefore, $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, $\boldsymbol{h}_3$, $\boldsymbol{h}_4$, $\boldsymbol{h}_5$ correspond to normalized distance between the center of the rectangular target region to any corner of 1, 2, 3, 4, 5 respectively. Kernel obtained with $\boldsymbol{h}_1$, $\boldsymbol{h}_3$ and $\boldsymbol{h}_5$ are represented in Figure 6.5. This value is considered as difficult to adapt and is often application dependent [60].

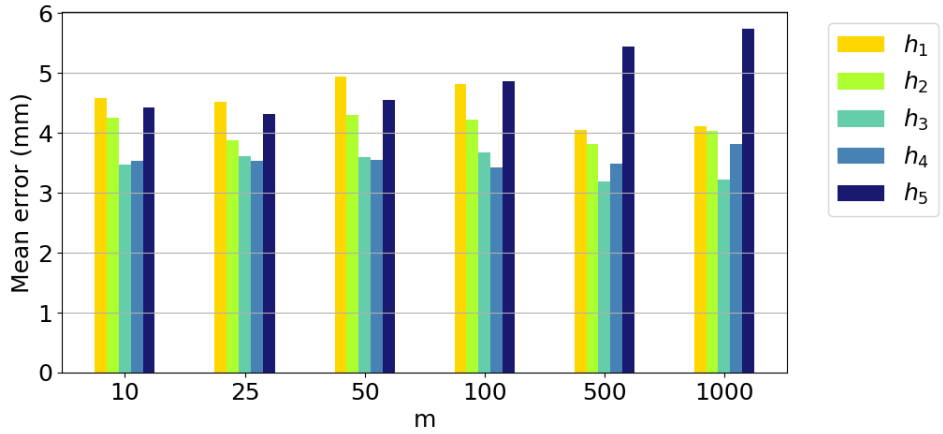
First of all, we will comment the results obtained with best and worst sets of m and $\boldsymbol{h}$ parameters. For each patient the mean value of the absolute error between the predicted and ground truth target location has been computed. Mean absolute errors along horizontal and vertical axes have also been studied. Best and worst absolute, horizontal and vertical mean errors for each patient are summarized in Table 6.1.
In this table, we can read that lowest absolute mean error rate achieved for patient 1 is equal to 3.8 mm ($m = 1000$ and $\boldsymbol{h}_1$). For this patient the motion range of the tumor motion was 6.8 mm. Patient 2 shows a minimum absolute mean error rate of 3.2 mm ($m = 500$ and $\boldsymbol{h}_3$) for a 6.2 mm motion range. For patient 3, lowest absolute mean error rate is of 3.1 mm ($m = 25$ and $\boldsymbol{h}_5$) with a motion range of 4.9 mm.

TABLE 6.1: Summary of absolute, horizontal and vertical mean errors (in millimeters) obtained on all three patients. These errors are computed by averaging all errors obtained on each scans of a given patient. Range of tumor motions are also provided as indication.

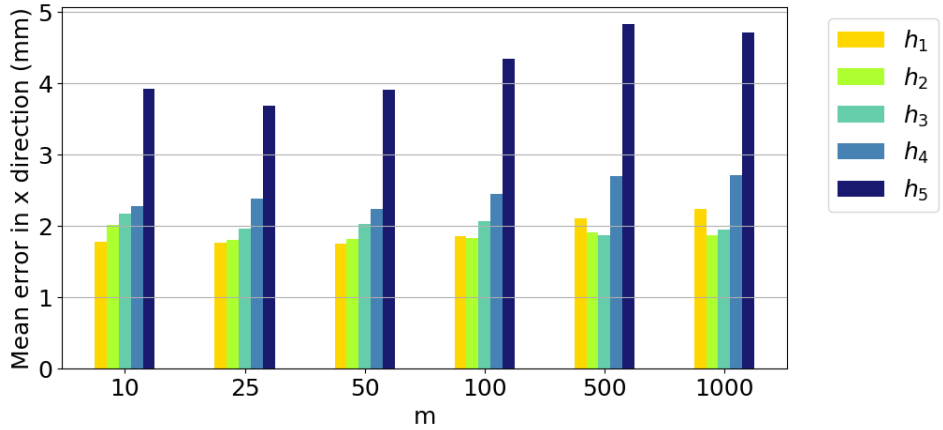
Absolute mean error		Horizontal mean error (x axis)		Vertical mean error (y axis)	
Patient 1					
Min	Max	Min	Max	Min	Max
3.8	6.1	0.4	1.9	3.7	5.9
Motion range : 6.8		Motion range : 0.2		Motion range : 6.8	
Patient 2					
Min	Max	Min	Max	Min	Max
3.2	5.8	1.8	4.8	2.1	4.8
Motion range : 6.2		Motion range : 2.8		Motion range : 5.5	
Patient 3					
Min	Max	Min	Max	Min	Max
3.1	4.9	2.7	4.7	1.1	1.9
Motion range : 4.9		Motion range : 4.8		Motion range : 1.2	

At the first glance, minimum absolute mean error achieved is higher for patient 1 than the two others but motion range is also higher. Regarding the minimum absolute error rate obtained relatively to motion range, results obtained on patient 2 are better than those obtained on patient 3 (accuracy of 3.2 mm for a motion range of 6.2 mm versus an accuracy of 3.1 mm for a motion range of 4.9 mm).

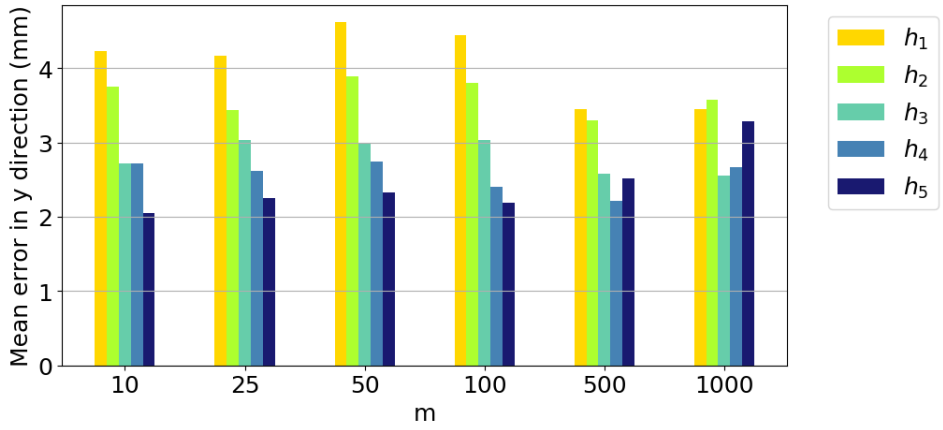
This table also suggests the importance of the parameters on the tracking results. Depending on their values, the error in the target location prediction can be almost the double (6.1 mm versus 3.8 mm for patient 1, 5.8 mm versus 3.2 mm for patient 2 and 4.9 mm versus 3.1 mm for patient 3). This indicates that the effect of m and $\boldsymbol{h}$ on the result is substantial. In order to go deeper in this analysis, absolute, horizontal and vertical mean errors depending on the parameters are presented for patient 2. This analysis is proposed in Figure 6.6.



(a) Absolute mean error (motion range around 6.2 mm).



(b) Horizontal mean error (motion range around 2.8 mm).



(c) Vertical mean error (motion range around 5.5 mm).

FIGURE 6.6: Absolute, horizontal and vertical mean errors obtained with the kernel mean-shift tracking on the patient 2.

For patient 2, the mean errors between the center referenced as ground truth and predicted tumor location found by kernel mean-shift tracking is presented in Figure 6.6(a) for multiple values of m and $\boldsymbol{h}$. The errors of prediction along the x axis (horizontal) and the y axis (vertical) are also presented in Figures 6.6(b) and 6.6(c), respectively.
As observed in Figure 6.6(a), the absolute mean error varies approximately between 3 and 6 mm. Lower errors rates are obtained with $\boldsymbol{h_3}$ and $\boldsymbol{h_4}$. The number of bins in the weighted histograms (m) seems not to have a predominant influence on the result.
In Figures 6.6(b) and 6.6(c), we observe that the mean horizontal error increases as the kernel bandwidth ($\boldsymbol{h}$) decreases and inversely for the mean vertical error. Such tendencies are not observed on patients 1 and 3.

The effect of the $\boldsymbol{h}$ parameter on the results is not easy to interpret. This parameter can be somewhat considered as a scaling factor. As presented in [60], if the target takes away from the camera, the kernel bandwidth should decrease in order to be scale invariant. In our context where the tumor is not affected by scale changes, we need to find another explanation to determine the effect of $\boldsymbol{h}$ on the results. However, since the 3 patients do not exhibit the same variation pattern depending on $\boldsymbol{h}$, we were not able to figure out why $\boldsymbol{h}$ parameter has inverse effect on horizontal and vertical target location for patient 2.

6.4 Discussion

Impact of the Low-Frame-Rate (LFR). As the method employed to produce results presented above is very similar to the one used by Zhang et al. [53], a comparison of the two results could determine if we reached expected performances. However, the comparison of the results performed under different conditions is a delicate task. In [53], the kernel mean-shift algorithm is applied to fluoroscopic videos whose 300×300 pixels images have a 0.26 mm/pixel resolution. These images have been acquired at a 15 frame per second (fps) rate. The comparison of the results obtained with our method and those obtained in the aforementioned article suggests that the accuracy here obtained is lower than expected.

A plausible explanation of these higher distances between the ground truths and the predicted centers lies on the low fps rate of our data. Instead of working with a real fluoroscopic data set, we simulated ten projections out of a 4DCT. Only ten images are used to represent a whole respiratory cycle. Considering that the duration of a respiratory cycle is about 5 seconds, the acquisition rate of our data is approximately 2 fps, more than seven time less than in the reference article.
Due to the sensitivity of mean-shift algorithm to converge to local maxima, the method is not suited to track large target movements which occur in Low-Frame-Rate (LFR) data. Moreover, a bad localization occurring at time t can have dramatic effect of all subsequent frames. This phenomenon is illustrated in Figure 6.7. For series 1 and 2, the tracking follows quite well the trajectory of the tumor. Therefore, each location in frame $t-1$ is a good initialization for the tracking in frame t. For series 2 and 3, the first predicted location is far from the trajectory. The location found at time $t-1$ is then a bad initialization at time t which can not recover from this bad detection in the previous frame. The predicted trajectory is, in such cases, is highly diverging from the true trajectory. The probability of propagating a mislocation can be reduced by increasing the frame rate.

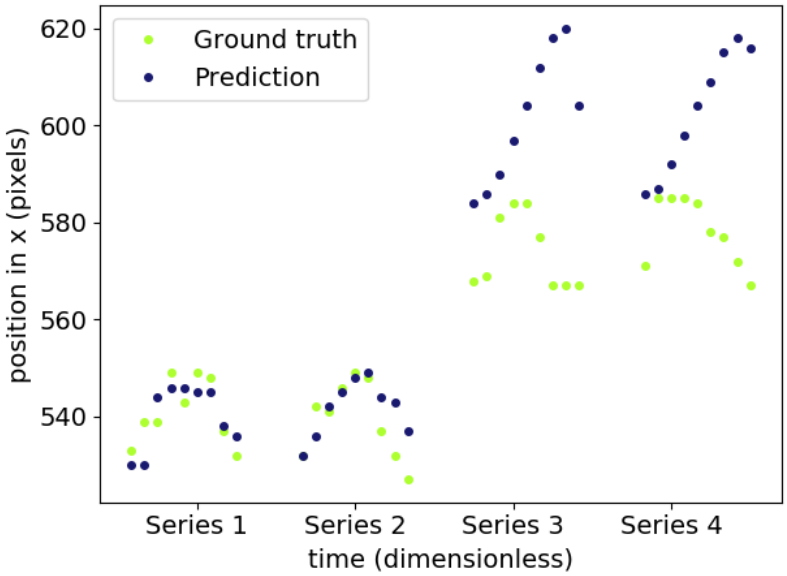


FIGURE 6.7: Kernel mean-shift tracking of the tumor along the x axis (horizontal direction) for patient 2 with $m = 10$ and $\boldsymbol{h}_4$.

A possible way of improvement to deal with LFR data is presented in [63]. By taking into account that the location of the target in the previous frame can be a bad initialization, they proposed to initialize the mean-shift algorithm at multiple locations near to the previous target estimated position. The predicted location of the tumor in the current frame is the one for which the similarity coefficient (for example, Bhattacharyya) is the highest. These initial spots are located in moving regions, i.e. regions where the difference between the current frame and its predecessor is high. In applications in which the background can be considered as relatively static, this approach can improve the results.
In addition, multiple target models are used in the article. In applications where the target appearance is variable, this extension is necessary to ensure accuracy of the results. However, for lung tumor detection, the tumor appearance can be considered as invariant regarding to scale changes. Remind that the kernel representation of the target with a histogram makes the mean-shift robust to (minor) deformations.

Impact of the image quality enhancement step. In order to asses whether the contrast enhancement method described in Section 4.2 brings some improvement on the tracking process or not, we applied the kernel mean-shift algorithm in the original DRRs. Absolute mean error depending on the parameters for patient 2 is represented in Figure 6.8. In this case, a minimum error around 9 mm is obtained with $m = 10$ and $\boldsymbol{h}_1$. In comparison to the 3.1 mm obtained with the contrast enhanced DRRs (see Figure 6.6(a)), this sustains the thinking that the image quality enhancement step involving a CLAHE approach brings a significant improvement in the accuracy of the kernel mean-shift tracking process.
Intuitively, the CLAHE method has the effect of widening the dynamic range of the target region histogram. Due to the nature of the Bhattacharyya coefficient, such a similarity metric is not very informative for narrow histograms, making the mean-shift method unable to find the most similar target candidate.

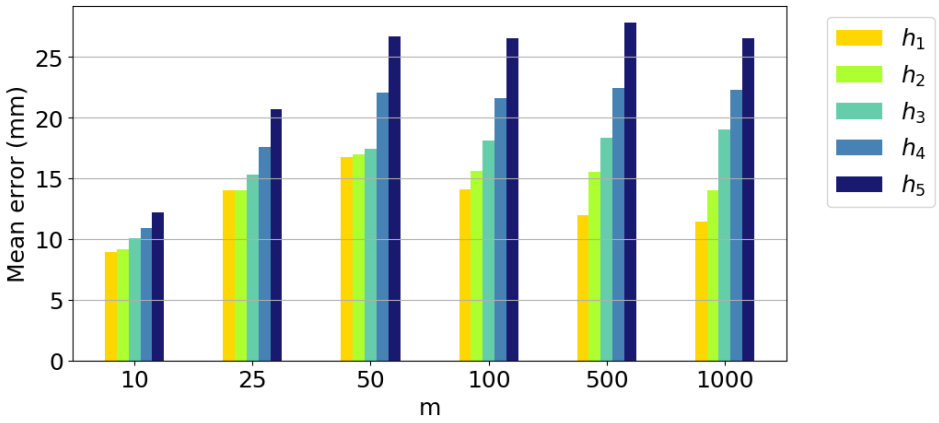


FIGURE 6.8: Absolute mean error obtained with the kernel-based tracking on the original DRRs of patient 2 (without the image quality enhancement step).

Chapter 7

Conclusion

Geometrical uncertainties induced by breathing motion degrade the conformity in the delivered dose distribution profile. In proton therapy, where a misestimation of the patient anatomy can lead to large discrepancies between planned and absorbed radiation doses, motion mitigation strategies are all the more crucial to offer a possibility of treating lung tumors with proton therapy.
This thesis addressed the problem of lung tumor detection in fluoroscopic data. The purpose of such technique is dual. On the one hand, the choice of the motion mitigation strategy relies on a characterization of tumor motions. On the other hand, both beam gating and tracking are techniques synchronized with tumor motions. Therefore, an accurate real-time assessment of the tumor shape and morphology is needed in order to ensure the success of these techniques.

First, a fluoroscopic image enhancement technique has been proposed in Chapter 4. Due to the poor contrast of fluoroscopic images, we suggested to preprocess the DRRs, which simulate a fluoroscopic data acquisition, to enhance their quality. Purpose of this step was both to increase edge visibility and enhance contrast. The proposed method combined a DoG filtering step followed by the application of the CLAHE algorithm, whose parameters were automatically determined by entropy maximization. The entropy has been computed over the entire image and within a contextual region encompassing the tumor.
Outcome of the method showed a net improvement of the contrast resulting in a better visualization of the tumor. The enhancement of the contrast within soft tissues was even more noticeable when the parameters of the proposed method maximized the entropy of the contextual region of the tumor instead of the the whole image.
As discussed in Section 4.4, the need of a DoG filtering to enhance bony structures visibility is arguable for automated tumor detection. In this context, the enhancement of the contrast between the tumor and surrounding tissues was the key objective of the preprocessing method. Indeed, higher is the contrast in an image, easier are the tumor segmentation and tracking tasks.

Second, a segmentation method has been proposed in Chapter 5 in order to characterize tumor morphology in fluoroscopic images. Segmentation in low contrast images with blurred transitions between regions is very challenging. Therefore, we proposed a method which integrated a strong prior knowledge about the tumor delineation in order to compensate for the poor information contained in the images. In this respect, the proposed solution was a graph cut approach initialized with a binary mask of the tumor instead of a manual labeling of some pixels performed by an expert. The binary mask has been drawn previously on another image series of the same patient. The mask corresponding to the same breathing phase was then selected to initialize the graph cut algorithm.

Despite the change of the graph weights computation, the method was not able to improve the delineation initially provided with any of the tested sets of parameters. A possible way of improvement, consists in using this prior binary mask in order to restrict the solution space containing the tumor delineation.
As already mentioned in this chapter, active contours could also have been a good candidate method to perform the segmentation task in fluoroscopic images. We opted for graph cut due to the intuition that the initialization of the method was less critical for graph cut than for active contours. However, in the light of the results, this assumption was not verified since graph cut results are mainly dependent on the initial binary mask location and morphology. Therefore, an active contour approach driven by some prior information about the tumor shape could have been investigated. For example, the solution could be forced to present similar shape and volume than a prior known model of the tumor.

Third, we performed tumor tracking thanks to a kernel mean-shift method in Chapter 6. Despite promising results emerging from the literature, accuracy of kernel mean-shift algorithm on our data set was disappointing. The most likely reason, that caused these large mean distance errors, is the use of ten DDRs to simulate a fluoroscopic video. This sequence of ten images representing a complete breathing cycle corresponds approximately to a 2 fps data acquisition while real fluoroscopy stands around 15 fps. In LFR, target motion is likely to be large, making mean-shift algorithm more subject to fall into a local maximum along its path to the true target location. Considering this and the almost inability of mean-shift to recover from a bad initialization of the target location, the method resulted in diverging tumor trajectories in a large number of cases. Even if multiple target initializations in high movement regions can be used to enhance tracking in data acquired at LFR, testing the kernel mean-shift on real fluoroscopic data would be the best way to produce informative results.

Despite the insufficient accuracy of the graph cut and kernel mean-shift methods, some improvement suggested in this thesis could significantly enhance the result obtained with these methods. However, the acquisition of a real fluoroscopic data set for future research is strongly recommended for a validation of the methods in a practical application purpose.

Bibliography

[1] BWKP Stewart, Christopher P Wild, et al. "World cancer report 2014". In: *Health* (2017).

[2] L Moonen and H Bartelink. "Fractionation in radiotherapy". In: *Cancer Treatment Reviews* 20.4 (1994), p. 365.

[3] Harald Paganetti. *Proton therapy physics*. CRC press, 2016.

[4] Radhe Mohan, Anita Mahajan, and Bruce D Minsky. "New strategies in radiation therapy: Exploiting the full potential of protons". In: *Clinical Cancer Research* 19.23 (2013), pp. 6338–6343.

[5] SE McGowan, NG Burnet, and AJ Lomax. "Treatment planning optimisation in proton therapy". In: *The British journal of radiology* 86.1021 (2013), pp. 20120288–20120288.

[6] Eike Rietzel and Christoph Bert. "Respiratory motion management in particle therapy". In: *Medical physics* 37.2 (2010), pp. 449–460.

[7] Dirk De Ruysscher et al. "Tumour movement in proton therapy: solutions and remaining questions: a review". In: *Cancers* 7.3 (2015), pp. 1143–1153.

[8] Christoph Bert, Sven O Grözinger, and Eike Rietzel. "Quantification of interplay effects of scanned particle beams and moving targets". In: *Physics in Medicine & Biology* 53.9 (2008), p. 2253.

[9] Nicolas Peguret et al. "Breathing adapted radiotherapy: a 4D gating software for lung cancer". In: *Radiation oncology* 6.1 (2011), p. 78.

[10] Bradley M Oborn et al. "Future of Medical Physics: Real-time MRI guided Proton Therapy". In: *Medical physics* (2017).

[11] SPM Crijns, BW Raaymakers, and JJW Lagendijk. "Real-time correction of magnetic field inhomogeneity-induced image distortions for MRI-guided conventional and proton radiotherapy". In: *Physics in Medicine & Biology* 56.1 (2010), p. 289.

[12] Shinichi Shimizu et al. "A proton beam therapy system dedicated to spot-scanning increases accuracy with moving tumors by real-time imaging and gating and reduces equipment size". In: *PLoS One* 9.4 (2014), e94971.

[13] PJ Keall et al. "Motion adaptive x-ray therapy: a feasibility study". In: *Physics in Medicine & Biology* 46.1 (2001), p. 1.

[14] Junsang Cho et al. "Development of a Real-Time Internal and External Marker Based Gating System for Proton Therapy". In: *Progress in Medical Physics* 28.3 (2017), pp. 92–99.

[15] Paolo Bifulco et al. "2D-3D registration of CT vertebra volume to fluoroscopy projection: a calibration model assessment". In: *EURASIP Journal on Advances in Signal Processing* 2010.1 (2009), p. 806094.

[16] Simon Rit et al. "The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK)". In: *Journal of Physics: Conference Series*. Vol. 489. 1. IOP Publishing. 2014, p. 012079.

[17] Thomas Martin Deserno. *Biomedical image processing*. Springer Science & Business Media, 2011.

[18] Kushal Roy and Amit Phadikar. "Automated Medical Image Segmentation: A Survey". In: *Proc. of Int. Conf. on Computing, Communication and Manufacturing*. Citeseer. 2014.

[19] Rahul Rajendran et al. "A versatile edge preserving image enhancement approach for medical images using guided filter". In: *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE. 2016, pp. 002341–002346.

[20] Kaiming He, Jian Sun, and Xiaoou Tang. "Guided image filtering". In: *IEEE transactions on pattern analysis and machine intelligence* 35.6 (2013), pp. 1397–1409.

[21] Paolo Irrera, Isabelle Bloch, and Maurice Delplanque. "A flexible patch based approach for combined denoising and contrast enhancement of digital X-ray images". In: *Medical image analysis* 28 (2016), pp. 33–45.

[22] Edward H Adelson et al. "Pyramid methods in image processing". In: *RCA engineer* 29.6 (1984), pp. 33–41.

[23] Jianfeng Qiu et al. "Automatic x-ray image contrast enhancement based on parameter auto-optimization". In: *Journal of applied clinical medical physics* 18.6 (2017), pp. 218–223.

[24] Stephen M Pizer et al. "Adaptive histogram equalization and its variations". In: *Computer vision, graphics, and image processing* 39.3 (1987), pp. 355–368.

[25] Karel Zuiderveld. "Contrast limited adaptive histogram equalization". In: *Graphics gems* (1994), pp. 474–485.

[26] Ali M Reza. "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement". In: *Journal of VLSI signal processing systems for signal, image and video technology* 38.1 (2004), pp. 35–44.

[27] Byong Seok Min et al. "A novel method of determining parameters of CLAHE based on image entropy". In: *International Journal of Software Engineering and Its Applications* 7.5 (2013), pp. 113–120.

[28] Luis G More et al. "Parameter tuning of CLAHE based on multi-objective optimization to achieve different contrast levels in medical images". In: *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE. 2015, pp. 4644–4648.

[29] Dzung L Pham, Chenyang Xu, and Jerry L Prince. "Current methods in medical image segmentation". In: *Annual review of biomedical engineering* 2.1 (2000), pp. 315–337.

[30] A Karthikeyan and M Valliammai. "Lungs segmentation using multi-level thresholding in CT images". In: *Int. J. Electron. Comput. Sci. Eng* 1.3 (2012), pp. 1509–1513.

[31] William J Kostis et al. "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images". In: *IEEE transactions on medical imaging* 22.10 (2003), pp. 1259–1274.

[32] Jamshid Dehmeshki et al. "Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach". In: *IEEE transactions on medical imaging* 27.4 (2008), pp. 467–480.

[33] Jianhua Xuan, Tülay Adali, and Yue Wang. "Segmentation of magnetic resonance brain image: integrating region growing and edge detection". In: *Image Processing, 1995. Proceedings., International Conference on*. Vol. 3. IEEE. 1995, pp. 544–547.

[34] Yu-Len Huang and Dar-Ren Chen. "Watershed segmentation for breast tumor in 2-D sonography". In: *Ultrasound in Medicine and Biology* 30.5 (2004), pp. 625–632.

[35] Yanni Zou and Bo Liu. "Survey on clustering-based image segmentation techniques". In: *Computer Supported Cooperative Work in Design (CSCWD), 2016 IEEE 20th International Conference on*. IEEE. 2016, pp. 106–110.

[36] Kalavathi Palanisamy and A Dhavapandiammal. "Segmentation of Lung Tumor in CT Scan Images using FA-FCM Algorithms". In: 18 (May 2016), pp. 74–79.

[37] Juan Eugenio Iglesias and Mert R Sabuncu. "Multi-atlas segmentation of biomedical images: a survey". In: *Medical image analysis* 24.1 (2015), pp. 205–219.

[38] Marius Erdt, Sebastian Steger, and Georgios Sakas. "Regmentation: A new view of image segmentation and registration". In: *Journal of Radiation Oncology Informatics* 4.1 (2017), pp. 1–23.

[39] Gavriil Tsechpenakis. "Deformable model-based medical image segmentation". In: *Multi modality state-of-the-art medical image segmentation and registration methodologies*. Springer, 2011, pp. 33–67.

[40] P Poudel et al. "Active contours extension and similarity indicators for improved 3D segmentation of thyroid ultrasound images". In: *Medical Imaging 2017: Imaging Informatics for Healthcare, Research, and Applications*. Vol. 10138. International Society for Optics and Photonics. 2017, p. 1013803.

[41] Yuri Y Boykov and M-P Jolly. "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images". In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Vol. 1. IEEE. 2001, pp. 105–112.

[42] Yuri Boykov and Vladimir Kolmogorov. "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision". In: *IEEE transactions on pattern analysis and machine intelligence* 26.9 (2004), pp. 1124–1137.

[43] Vladimir Kolmogorov and Ramin Zabin. "What energy functions can be minimized via graph cuts?" In: *IEEE transactions on pattern analysis and machine intelligence* 26.2 (2004), pp. 147–159.

[44] Zhengwen Shen et al. "Multi-phase simultaneous segmentation of tumor in lung 4D-CT data with context information". In: *PloS one* 12.6 (2017), e0178411.

[45] Wenlu Zhang et al. "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation". In: *NeuroImage* 108 (2015), pp. 214–224.

[46] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. "Online object tracking: A benchmark". In: *Computer vision and pattern recognition (CVPR), 2013 IEEE Conference on*. Ieee. 2013, pp. 2411–2418.

[47] De Vleeschouwer Christophe and Laurent Jacques. *LELEC2885 Image Processing and Computer vision*. Université catholique de Louvain. 2017.

[48] Satoshi Abe et al. "Fiducial marker matching versus vertebral body matching: Dosimetric impact of patient positioning in carbon ion radiotherapy for primary hepatic cancer". In: *Physica Medica: European Journal of Medical Physics* 33 (2017), pp. 114–120.

[49] Mikael Mongeon et al. "Safety and Efficacy of Endovascular Fiducial Marker Insertion for CyberKnife Stereotactic Radiation Therapy Planning in Early-Stage Lung Cancer". In: *Journal of Vascular and Interventional Radiology* 28.8 (2017), pp. 1090–1097.

[50] Tong Lin et al. "Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy". In: *Physics in Medicine & Biology* 54.4 (2009), p. 981.

[51] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[52] Ying Cui et al. "Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers". In: *Physics in Medicine & Biology* 52.20 (2007), p. 6229.

[53] Xiaoyong Zhang et al. "A kernel-based method for markerless tumor tracking in kV fluoroscopic images". In: *Physics in Medicine & Biology* 59.17 (2014), p. 4897.

[54] Mitra Basu. "Gaussian-based edge-detection methods-a survey". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 32.3 (2002), pp. 252–260.

[55] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (2nd Edition)*. Pearson, 2002.

[56] Veska Georgieva, Roumen Kountchev, and Ivo Draganov. "An adaptive enhancement of X-Ray images". In: *Advances in Intelligent Analysis of Medical Data and Decision Support Systems*. Springer, 2013, pp. 79–88.

[57] John Ashburner and Karl J Friston. "Rigid body registration". In: *Statistical parametric mapping: The analysis of functional brain images* (2007), pp. 49–62.

[58] Varduhi Yeghiazaryan and Irina Voiculescu. *An overview of current evaluation methods used in medical image segmentation*. Tech. rep. Tech. Rep. CS-RR-15-08, Department of Computer Science, University of Oxford UK, 2015.

[59] Greg Slabaugh and Gozde Unal. "Graph cuts segmentation using an elliptical shape prior". In: *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. Vol. 2. IEEE. 2005, pp. II–1222.

[60] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. "Kernel-based object tracking". In: *IEEE Transactions on pattern analysis and machine intelligence* 25.5 (2003), pp. 564–577.

[61] Dorin Comaniciu and Peter Meer. "Mean shift: A robust approach toward feature space analysis". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.5 (2002), pp. 603–619.

[62] R Collins. "Mean-shift Tracking". In: *CSE, PSU CSE598G Spring* (2006).

[63] Fatih Porikli and Oncel Tuzel. "Object tracking in low-frame-rate video". In: *Image and Video Communications and Processing 2005*. Vol. 5685. International Society for Optics and Photonics. 2005, pp. 72–80.

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve **www.uclouvain.be/epl**