UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

PhD thesis

Real-time deformable image registration

Applied to Tumor Tracking in 2D Cine-MRI

José David Tascón-Vidarte

Advisor: Sune Darkner

Submitted: January 28, 2022

This thesis has been submitted to the PhD School of The Faculty of Science, University of Copenhagen

Abstract

Registration is the process of aligning a set of images and is considered a fundamental task in computer vision and medical image analysis. For medical applications where tissue is involved, registration requires a deformable transformation. The deformation implies more parameters to solve, and the optimization is computationally expensive. More efficient and high-performing deformable registration algorithms will benefit many applications such as image-guided radiotherapy and surgery.

This thesis aims to improve deformable image registration (DIR) performance and achieve real-time operation for tumor tracking in Cine-MRIs. We explore the state-of-the-art of DIR to select a strategy that accomplishes high performance. We choose to enhance intensity-based DIR algorithms with a variational (gradient-based) approach. The reasons are generalization capabilities, accuracy, and data availability. The research path to implement and validate our strategy is as follows. First, we explore how accurate intensity-based algorithms are with low contrast organs such as the liver. Next, we design a high-performance image registration library. Then, we obtained a fast algorithm with the lowest reported time in literature to solve DIR in real-time. Finally, we applied the DIR algorithm to tumor tracking in the context of image-guided radiotherapy.

A proper evaluation of a tumor tracking application requires considerable ground truth data. Therefore we design a novel Cine-MRI simulator that creates video sequences with the underlined delineation of the organ and the tumor. Furthermore, we use treatment Cine-MRIs with manual delineations. After the data preparation, we evaluate deformable image registration and other tracking algorithms with multiple organs. We find that DIR is more accurate in tracking organs compared to tumors. As a consequence of a multiple-organ strategy, we propose and validate how well the tracking algorithms replicate novel gating control signals for image-guided radiotherapy.

Finally, we made a comprehensive study on tumor tracking with nine algorithms to find the best solution. We include liver and lung patients from simulation and treatment data with the most challenging conditions. We propose a novel tracking method that combines template matching with deformable image registration. This algorithm was among the best-performing algorithms overall for tumor tracking. In summary, we find that the best algorithms perform close to interobserver variability and we prove that tracking tumors on lung and liver patients offers similar accuracy. All the code generated during the development of this thesis is publicly available.

Resumé

Registrering er processen med at justere et sæt billeder og betragtes som en grundlæggende opgave i computersyn og medicinsk billedanalyse. Til medicinske anvendelser, hvor væv er involveret, kræver registrering en deformerbar transformation. Deformationen indebærer flere parametre at løse, og optimeringen er beregningsmæssigt dyr. Mere effektive og højtydende deformerbare registreringsalgoritmer vil gavne mange applikationer såsom billedstyret strålebehandling og kirurgi.

Denne afhandling har til formål at forbedre ydeevnen for deformerbar billedregistrering (DIR) og opnå realtidsdrift til tumorsporing i Cine-MRI'er. Vi udforsker DIR's state-of-the-art for at vælge en strategi, der opnår høj ydeevne. Vi vælger at forbedre intensitetsbaserede DIR-algoritmer med en variationsbaseret (gradientbaseret) tilgang. Årsagerne er generaliseringsmuligheder, nøjagtighed og datatilgængelighed. Forskningsvejen til at implementere og validere vores strategi er som følger. Først undersøger vi, hvor nøjagtige intensitetsbaserede algoritmer er med organer med lav kontrast, såsom leveren. Dernæst designer vi et højtydende billedregistreringsbibliotek. Derefter opnåede vi en hurtig algoritme med den laveste rapporterede tid i litteraturen til at løse DIR i realtid. Endelig anvendte vi DIR-algoritmen til tumorsporing i forbindelse med billedstyret strålebehandling.

En korrekt evaluering af en tumorsporingsapplikation kræver betydelige sandhedsdata. Derfor designer vi en ny Cine-MRI simulator, der skaber videosekvenser med den understregede afgrænsning af organet og tumoren. Ydermere anvender vi behandling Cine-MRI med manuelle afgrænsninger. Efter dataforberedelsen evaluerer vi deformerbar billedregistrering og andre sporingsalgoritmer med flere organer. Vi finder, at DIR er mere nøjagtig til at spore organer sammenlignet med tumorer. Som en konsekvens af en flerorganstrategi foreslår og validerer vi, hvor godt sporingsalgoritmerne replikerer nye gating-kontrolsignaler til billedstyret strålebehandling.

Til sidst lavede vi en omfattende undersøgelse af tumorsporing med ni algoritmer for at finde den bedste løsning. Vi inkluderer lever- og lungepatienter fra simulerings- og behandlingsdata med de mest udfordrende tilstande. Vi foreslår en ny sporingsmetode, der kombinerer skabelonmatchning med deformerbar billedregistrering. Denne algoritme var blandt de bedst ydende algoritmer generelt til tumorsporing. Sammenfattende finder vi, at de bedste algoritmer fungerer tæt på interobservatørvariabilitet, og vi beviser, at sporing af tumorer på lunge- og leverpatienter tilbyder lignende nøjagtighed. Al den kode, der genereres under udviklingen af dette speciale, er offentligt tilgængelig.

Acknowledgements

This thesis has offered me an extraordinary experience, a delightful path to knowledge, sometimes numerous concerns, and overall much personal growth. This project would not have been possible without the support of many people.

I want to thank my Supervisor, Sune Darkner, for all his support, sharing his ideas and moments of discussions. I am grateful to have such a knowledgeable, professional, and friendly mentor.

Kenny Erleben accompanied my progress closely and encouraged me to improve my work with his critical questions and reviews. Christine S. Andersen for all the administrative support during my Ph.D. studies. Julien Jomier, who hosted my research stay in Kitware Europe, provided me with insights into registration and the project in general.

I am thankful for having a very productive collaboration with some medical physicists from the Radiation Oncology Department of Rigshospitalet. Ivan R. Vogelius guided me in the radiotherapy field and offered valuable input to this project. Isak Wahlstedt, Line B. Stick, and Mirjana Josipovic for their professionalism and willingness to cooperate with me.

Thanks to my colleagues, professors, and Ph.D. fellow students from the Image Section Group of the Department of Computer Science at the University of Copenhagen.

Thanks to the RAINBOW Innovative Training Network and its coordination team, who provided me with the financial means to complete this work. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 764644.

Finally, I am grateful to my family. My parents, Diego and Raquel, for their unconditional love and support. My brother Juan Diego, who is a close friend, always available to me and willing to provide some help. My wife Catalina, for being with me on this long journey, for her big heart and warming touch that encourages me to continue. And my son Samuel, who has brought much joy and inspiration to my life.

Contents

Chapter 1

Introduction

This thesis deals with high-performance deformable image registration algorithms applied to real-time tumor tracking. The context of the application is image-guided radiotherapy. Therefore, we provide some background regarding the anatomy, the problem, and the related work.

1.1 Anatomy and Respiratory Motion

This thesis includes medical images of the liver and lungs in the radiotherapy context. Hence, we provide a short description of the anatomy involved. Figure 1.1 display the anatomy of the thorax and the abdomen. The organs in the thorax are the lungs a the heart. The ribs surround the thorax and can be modeled as a rigid object while the lungs and the heart move and deform due to respiratory motion and heart beating. The organs in the abdomen are the liver, the stomach, the spleen, the gallbladder, and the pancreas. These abdominal organs also deform during breathing. The respiratory motion represents the main concern for acquisition, planning, and treatment in radiotherapy when the target locations are in the thorax or the abdomen.

The acquisition of liver or lungs images requires to consider respiratory motion. Respiratory motion ranges between 5 to 20 mm with average cycle of 4 seconds. The breathing cycle has two settling periods at end-inspiration and end-expiration. The end-expiration tends to be more stable for longer periods [Keall et al., 2006]. Furhtermore, respiratory motion cause a pattern with hysteresis for tumors in the lungs [Seppenwoolde et al., 2002]. The AAPM generated guidelines for respiratory motion management in radiotherapy [Keall et al., 2006]. The recommended constraint of real-time tumor tracking within organs affected by respiratory motion is 0.5 s (2 fps) [Keall et al., 2006]. The algorithms presented here aim for lower computational times than the typical acquisition time of 250 ms (4 fps) of commercial MR-linacs [Olsen et al., 2015, Raaymakers et al., 2017].

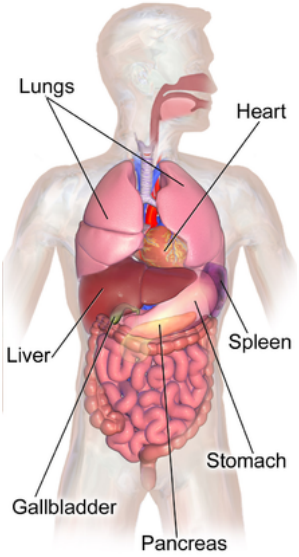


Figure 1.1: Abdominal organs anatomy (Image source: https://en.wikipedia.org/wiki/File:Abdominal_Organs_Anatomy.png).

1.2 Problem

Improving deformable image registration performance is beneficial for multiple areas in the medical field, such as diagnosis and prognosis of tumors, orthopedics, radiotherapy, surgeries, and microbiology. Some applications require short times or even real-time execution to align the images. This thesis aims to reduce the computational time of deformable image registration and demonstrate the benefits of a real-time registration solution for tumor tracking in Cine-MRIs. Therefore, we provide a formal description of the registration and tumor tracking problems.

Registration: Having a fixed image $I_0(x)$ and a moving image $I_1(x)$ find the spatial transformation $\varphi(x)$ between them.

Tumor tracking: Having a reference image $I_0(x)$ with a defined tumor structure $S(x)$, find the new tumor location in an input image $I_i(x)$ within a real-time constraint (4 fps).

The two problems are closely related in terms of formulation. However, the complexity in tumor tracking is higher due to the time constraint. In order to solve tumor tracking with a registration algorithm, it is required to find the transformation between the reference and each input image and then apply this transformation to the reference tumor structure (mask) to obtain the predicted tumor location (mask). The time constraint implies a trade-off between accuracy and performance for registration algorithms solved with a variational approach (gradient-based optimization). Therefore, we also seek

to find the right balance to achieve real-time execution.

1.3 Related Work

This thesis is related to deformable image registration with high-performance variants and its application to real-time tumor tracking. The list of DIR methods in the literature is overwhelmingly long. Thus, we offer a survey of medical image registration with some well-known and distinguished methods in Chapter 2. In this section, we provide the reader an insight into high-performance and parallel DIR algorithms in addition to some tumor tracking algorithms that are directly comparable to our work.

For more than a decade, parallel strategies for DIR algorithms have been on the scene. Multiple surveys of medical image registration algorithms in parallel hardware confirm the performance gains and the parallelization capabilities [Shams et al., 2010, Fluck et al., 2011, Eklund et al., 2013]. These authors introduce the parallel hardware, the processing and the memory pipelines, the considerations for each stage, and examples of efficient algorithms. We use some of these concepts in our work.

The GPU has arisen as the popular device of choice for implementing parallel DIR algorithms. Notorious implementation examples are [Sharp et al., 2007, Muyan-Ozcelik et al., 2008, Samant et al., 2008, Gu et al., 2009, Oh et al., 2011]. Regarding the platform, CUDA and OpenCL are both widely extended. An early example and comparison of several demons algorithms on the GPU is presented in [Gu et al., 2009]. More recent examples are [Ekström et al., 2021] and [Brunn et al., 2021] who present a novel library that support multi-node and memory distributed registration. We observe that none of them have reported a real-time deformable registration application.

Tumor tracking can be solved automatically using image analysis. Some proposed strategies for tumor tracking are based on template matching [Cervino et al., 2011, Tryggestad et al., 2013, Shi et al., 2014, Menten et al., 2018, Fast et al., 2017], feature detection [Paganelli et al., 2015, Mazur et al., 2016, Fast et al., 2017], optical-flow methods [Zachiu et al., 2015, Seregni et al., 2018], deformable image registration [Fast et al., 2017, Friedrich et al., 2021], modeling based [Garau et al., 2019], segmentation with variational methods [Gou et al., 2014], or segmentation with neural networks [Cervino et al., 2011, Yun et al., 2015, Friedrich et al., 2021]. Current tracking systems used in clinical practice may fail to track unexpected movements and have difficulty in tracking motion in the out-of-plane direction [Paganelli et al., 2018]. This leaves significant room for improvement in tumor tracking.

Regarding the algorithms, only a few studies cover multiple algorithm evaluations of tumor tracking [Cervino et al., 2011, Fast et al., 2017, Friedrich et al., 2021]. The most comprehensive was [Fast et al., 2017] who compared four algorithms. The authors found that the two best-performing algorithms

are deformable image registration (DIR) with B-splines and multiple template matching (MTM). Recently, [Friedrich et al., 2021] compared tumor tracking between DIR B-splines and UNET segmentation, the latter performed better.

Both [Fast et al., 2017] and [Friedrich et al., 2021] B-splines algorithms did not achieve real-time operation at four frames per second. In contrast, we exhibit in this project a B-spline algorithm working at one frame per second, and the first reported DIR algorithm to achieve real-time performance using a demons strategy operating at eight frames per second. All the compared algorithms by Fast et al. and Friedrich et al. use ten templates images to work. Under normal conditions, the oncologist only delineates organs over a single image on treatment day, and therefore we use algorithms in Chapter 7 that fulfill this condition.

Other studies have implemented DIR algorithms for tumor tracking applications in other image modalities. [Reaungamornrat et al., 2016] implemented demons algorithms for aligning MR-CT images of the spine. The performance is 1 min a far from real-time. Another work is [Cifor et al., 2013] that performed ultrasound tumor tracking with deformable image registration. The authors achieve a performance of 5 min per frame. Tumor tracking is extensively evaluated with ultrasound images [De Luca et al., 2018]. The authors offered four methods, CNN, KCF, Lucas-Kanade tracker, and one DIR algorithm, the Log-Demons. The Log-Demons is a close proposal to our work. However, their algorithm is featured-based and works only in a localized image region (4.8 fps). In contrast, our algorithm works at a full imaging scale and offers better performance (8.9 fps). Our template matching combined with demons (TMDEM) algorithm presented in Chapter 7, that represent our localized version of DIR works at 53.3 fps.

1.4 Outline of thesis

This thesis consists of chapters written as papers. The papers are to be submitted, in review, or published. We do not modify the contents of the papers under review or published, and the only applied changes relate to the format to fit the thesis layout.

The overall goal of this thesis is to achieve real-time performance for deformable image registration. In order to prove its usefulness, we focus on the tumor tracking problem in image-guided radiotherapy.

In Chapter 2, we present a state-of-the-art in deformable image registration. We formulate mathematically the commonly used framework in the variational approach that includes transformation, interpolation, similarity metrics, regularization, and optimization routine. Furthermore, we include a description of the most successful algorithms, open-source implementations, and publicly available datasets to evaluate DIR accuracy. Finally, we present the considerations to parallelize registration algorithms.

When the project started in 2018, the most successful DIR algorithms were based on the variational approach. In contrast, the deep learning approaches emerged with good performance times but still with accuracy drawbacks. The factors in choosing one approach versus the other were: generalization capabilities, data availability, and accuracy. We determined that there was still potential for improvement in the variational approach, and we decided to implement a real-time intensity-based DIR.

In Chapter 3, we evaluate the accuracy of intensity-based deformable image registration with tumors in the liver. We use a dataset of 4DCT scans from patients with fiducial markers implanted near the tumor. We exploit this data to test registration under respiratory motion and to propose the model used in Chapter 5. We demonstrate that intensity-based DIR is accurate in radiotherapy (range $2-3mm$ [Brock et al., 2017]) to estimate tumor motion.

In Chapter 4, we design a high performance computing registration library. The library supports image analysis operations, filtering, and deformable image registration on CPU and GPU. We focused on the transformation-interpolation operation, which is highly repeated during each optimization step in registration. Furthermore, we implemented demons and LDDMM strategies. The demons algorithm is the only suitable for real-time operation. As a result, the high-performance demons algorithm is proven and tested along with Chapter 6 and 7.

In Chapter 4, we design a high performance computing registration library. The library supports image analysis operations, filtering, and deformable image registration on CPU and GPU. We focused on the transformation-interpolation operation, which is highly repeated during each optimization step in registration. Among the registration algorithms, we implemented the demons and the large deformations diffeomorphic metric mapping (LDDMM) strategies. The demons algorithm is the only one suitable for real-time operation. As a result, the high-performance demons algorithm is proven and tested along with Chapter 6 and 7.

In Chapter 5, we propose a patient-specific simulator of 2D Cine-MRI in free-breathing motion. The system can generate a Cine-MRI with simulated ground truth contours of the desired organs using pre-treatment images. The simulator includes the out-of-plane motion because we model the 3D respiratory motion. We validate that the simulation represents tumor motion in free breathing. We demonstrate its usefulness for evaluating tumor tracking algorithms by generating multiple Cine-MRI per patient with varying motion amplitude and noise.

In Chapter 6, we explore tracking multiple organs on 2D Cine-MRI treatment data. The data includes two liver and two lung patients with a field of view between the thorax and the abdomen. We aim to track the liver, the lung, and the tumor simultaneously. We compare four real-time algorithms: B-splines registration, demons registration, UNET segmentation and object tracking with a siamse region proposal network. We introduce a UNET

algorithm for tumor tracking that is trained with a single image using augmentations. Finally, we validate how the multiple-organ tracking algorithms replicate convenient gating control signals for image-guided radiotherapy.

In Chapter 7, we compare the accuracy and performance of nine (9) tracking algorithms using 2D Cine-MRIs from simulation and treatment data. We include in the comparison state-of-the-art medical imaging and computer vision algorithms. We propose a new algorithm named TMDEM that is based on template matching for global tracking combined with demons deformable image registration for local refinement. The results indicate that our algorithm is among the best performing algorithms with the advantage that generalize better compared to a UNET segmentation approach. In summary, we prove that tracking tumors on lung and liver patients offers no statistical significant differences regarding accuracy and that the best-performing algorithms achieved results close to inter-observer variability for treatment data.

1.5 Contributions

The contributions of this thesis are listed below:

- IMART, a high performance library for medical image registration (Chapter 4).
- First implementation of deformable image registration applied for tumor tracking on 2D Cine-MRI at full imaging scale that fulfill the real-time constraint of 4 fps for commercial MR-linacs (Chapter 4, 6 and 7).
- Novel pipeline for simulation of 2D Cine-MRI with the underlined delineations of organ and tumors. (Chapter 5 and 7).
- First evaluation of multiple-organ tracking (Chapter 6)
- Proposed gating control signals from multiple-organ tracking useful for image-guided radiotherapy (Chapter 6).
- Comprehensive comparative study with nine algorithms applied to tumor and organ tracking under challenging conditions for liver and lung patients (Chapter 7).
- Novel method for tumor tracking based on combined template matching with demons deformable image registration (Chapter 7).

We publish all our code as open source:

- IMART: `https://github.com/josetascon/imart`.
- Cine-MRI simulator: `https://github.com/josetascon/cinemri-simulation`.
- Tracking Algorithms: `https://github.com/josetascon/tracking-tissue`.

Chapter 2

A Survey in Deformable Image Registration

The work presented in this chapter is an original paper written for this thesis, that is to be submitted as: Tascón-Vidarte, J. D. (2022). A Survey in Medical Image Registration.

2.1 Abstract

Registration is the process of aligning two images into a single coordinate space. From the medical image analysis point of view, registration is a primary task and a mature subject. Nevertheless, too many medical applications require more accurate, faster, and automated registration algorithms. This situation generates multiple algorithm proposals per year. Thus, we present a modern survey of medical image analysis with a focus on deformable image registration. Furthermore, we propose a unified taxonomy to classify the registration methods. We cover the registration components such as transformation, interpolation, similarity metric, regularization, and optimization. We express the registration components in a unified mathematical framework and include a complexity analysis. Due to a large amount of research, we only include some distinguished registration methods. In addition, we develop and summarize validation, open-source implementations, and real-time considerations. Finally, we discuss the current research trends and prospects.

Keywords: *Deformable image registration, Medical Images, Similarity metric, Optimization*

2.2 Introduction

Registration is the process of aligning a set of images into a single coordinate space. In medical image analysis, registration is a primary task and is the subject of several research projects every year. Figure 2.1 depicts the research activity of medical image registration in the last twenty years. The trend is still growing with more than three thousand outputs in the last couple of years. From 2000 to 2010, the research focus was voxel intensity-based deformable image registration solved with a variational approach [Oliveira and Tavares, 2014]. Later, the deep learning approaches revitalized the research in the last five years [Haskins et al., 2020]. We will cover these topics, but a proper definition of the registration problem is required.

A registration problem is defined formally as: Having a fixed image $I_0(x)$ and a moving image $I_1(x)$ find the spatial transformation $\varphi(x)$ between them. If we consider registration as a system, the inputs are the fixed and the moving images, and the output is the transformation. Nevertheless, since the goal of registration is to align the images into a single coordinate space, another desired output is the warped moving image.

The components of a registration algorithm are the transformation, the interpolation, the similarity metric and the optimization. Figure 2.2 shows the registration pipeline and its components.

A registration algorithm solves a cost function E with a similarity metric E_M and regularization E_R components. An optimization routine will find the transformation $\varphi(x)$ parameters. The mathematical expression is defined as:

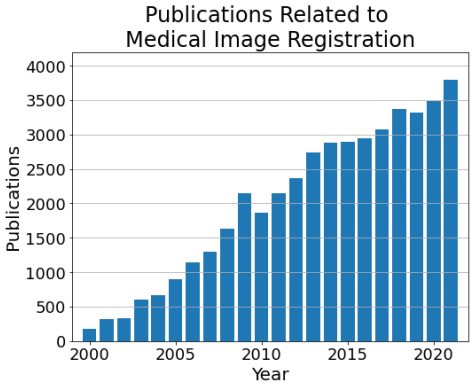


Figure 2.1: Research activity on medical image registration during the last twenty years. The publication numbers include the terms "medical image registration", "deformable image registration", "non-rigid registration", "deformable registration" and "non-rigid image registration"

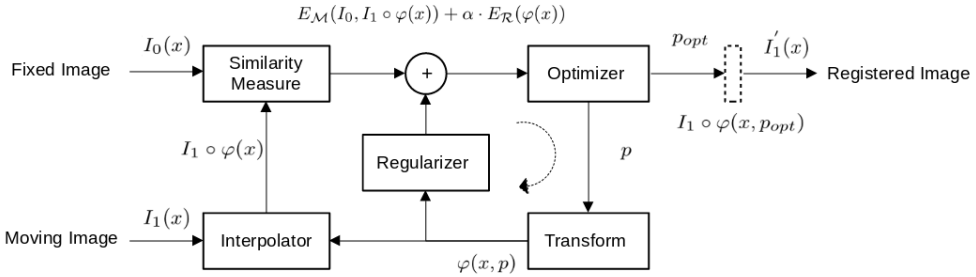


Figure 2.2: Medical image registration components.

$$\arg\min_{\varphi(x)} E = E_{\mathcal{M}}(I_0, I_1 \circ \varphi(x)) + \alpha \cdot E_{\mathcal{R}}(\varphi(x)) \tag{2.1}$$

Where:

$$x_j \, \epsilon \, \Omega_j, \; \Omega_j \subset \mathbb{R}^d, \; j = 0, 1$$

$$I_j : \Omega_j \to \mathbb{R}, \; \varphi(x) : \Omega_0 \to \Omega_1$$

Although the problem statement of registration is short and straightforward, it is a complex problem to solve. According to [Hadamard, 1923] and its definition of well-posed problems, medical image registration is an ill-posed problem, i.e., there is no unique solution. This is because different transformations can be found as possible solutions. Furthermore, the cost function nonlinearity and nonconvexity may produce solutions with different local minimums depending on the optimization method used. This is more evident for deformable transformations that have a high number of parameters.

Deformable image registration (DIR) refers to the medical image application where the objects in the images are non-rigid. The principle is that

deformable objects contract, expand, or move over time. The organs are good examples of these objects.

The desired conditions of a transformation found by a DIR algorithm are:

- Inverse Consistency: The algorithm produces a consistent solution independent of the choice of the fixed and the moving image [Christensen and Johnson, 2001]. This means that if the fixed image is exchanged with the moving image, the registration algorithm will produce the inverse transformation.

- Topology preservation: One-to-one transformation where no folding of space occurs [Noblet et al., 2005]. This preserves structure which is essential for the study of anatomy.

- Diffeomorphism: A differential mapping that has a differentiable inverse [Younes, 2010]. When a transformation is a diffeomorphism, the smoothness of surfaces are preserved, and the coordinates are transformed consistently. Therefore, a diffeomorphism implies inverse consistency and inverse consistency.

These conditions are usually imposed in the cost function of the registration algorithm, primarily in the regularization. However, some transformations implicitly define these constraints [Rueckert et al., 2006]. However, there are particular cases in medical image registration where these properties are not desired to model singularities, such as lung sliding motion [Hua et al., 2017] or brain tumors [Nielsen et al., 2019]. Thus, a contextual understanding of the medical imaging problem is required to propose an adequate registration solution.

The straightforward registration problem involves two similar images. However, for medical images, the complexity is increased due to multiple options of image acquisition, involved subjects (patients), desired targets, and the components in mathematical solutions. Therefore, a proper taxonomy is required to define the universe of possibilities.

2.3 Classification of Registration Methods

We propose a taxonomy of medical image registration in Figure 2.3. Our taxonomy is based on Maintz and Viergever [Maintz and Viergever, 1998] and includes the deformation models proposed by Sotiras et al. [Sotiras et al., 2013]. We divide the context of the registration problem in two: the image acquisition and the mathematical solution. We formulate this division because the image acquisition context is established and defined by the problem domain conditions, while the mathematical solution side is user-specific and selected based on a desired registration accuracy and performance.

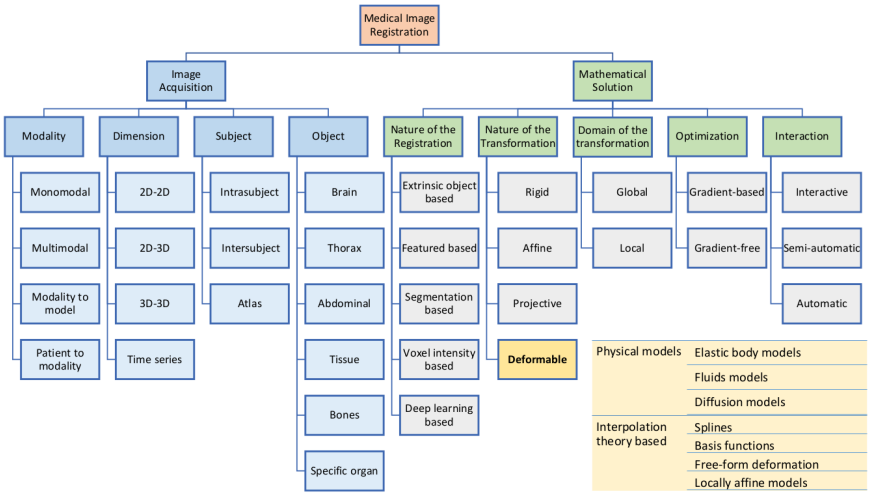


Figure 2.3: Medical image registration taxonomy.

The taxonomy presented by Maintz and Viergever in 1998 had an update in 2016 [Viergever et al., 2016]. The authors conclude that the classification originally proposed is still valid. Consequently, our taxonomy preserves a similar structure to [Maintz and Viergever, 1998]. Regarding deformable image registration and the transformations models, the survey made by [Sotiras et al., 2013] covers the work of [Holden, 2008] regarding deformable transformations. The latter presented a classification for deformable transformations consisting of physical models and function representations. In their review [Sotiras et al., 2013] follow this taxonomy, and thus, we consider that classification similarly. We will discuss and provide some specifics for each taxonomy domain below.

Image Acquisition Domain

The modality refers to the imaging device where the fixed and the moving images were acquired. In medical imaging a wide variety of imaging devices are used, such as: Ultrasound, X-ray, CT, PET, MR, CCD cameras, among others. The monomodal case is when the images to register were acquired with the same device. The multimodal case is when the images come from different devices, for instance registration of PET-CT [Goerres et al., 2002, Mattes et al., 2003], CBCT-CT [Zachiu et al., 2017], CT-MR [Roy et al., 2014] or Ultrasound-MR [Mercier et al., 2012]. Modality to model refers to the case when an image is registered to an existing 3D model (digital model). Patient to modality refers to the registration of pre-treatment images to intra-treatment images, although this distinction is covered by the former types.

The dimension refers to the image coordinates. The typical registration application consist of two images of the same dimension, as 2D-2D or 3D-3D. Other option is to register 2D to 3D images, a common example is Ultrasound to MR. The more complex dimension involves time series where registration is done pairwise or group-wise.

The subject refers to the patient captured in the images. Intrasubject is the case where the images are from the same patient. Intersubject refers to images from difference patients. Where an atlas is a registration perform to a image that represent multiple patients or a generalization of a patient.

The object refers to the field of view or the target that the images enclose. Examples of different objects are: brain, head-and-neck, thorax, abdomen, tissue, bones, extremities or specific organs (lung, liver, heart, stomach, pancreas, prostate, among others).

Mathematical Solution Domain

The nature of the registration refers to the algorithmic solution and the overall strategy to find the transformation.

Feature-based algorithms require landmarks annotated in both input images. The landmarks can correspond to anatomical features or intensity features. The cost function usually includes a euclidean distance [Modersitzki, 2004]. Extrinsic object based algorithms are feature-based registration methods with the difference that they have an external apparatus for guidance. The registration typically required an intervention to locate the extrinsic object, and therefore, these kinds of methods are in decline [Viergever et al., 2016].

Voxel intensity-based is the most popular choice to solve deformable image registration. The cost functions incorporate similarity metrics that only consider the intensity values of the input images [Oliveira and Tavares, 2014]. These methods are solved with a gradient-based optimizer and thus, represent a variational approach.

Segmentation-based algorithms are required to identify specific objects or organs in the images as a first step. Then, the segmented structures are typically used to improve algorithm convergence and fit boundaries.

Finite element models registration algorithms are a hybrid between segmentation-based and intensity-based. The method defines a pipeline that requires organ segmentation, surface mesh conversion, boundary conditions, manual parametrization of material properties, and then the finite element analysis [Brock et al., 2005].

Deep learning approaches are relatively new on the stage. These algorithms are based on Convolutional Neuronal Network and require large amounts of data for training a model [Litjens et al., 2017]. One issue that may arise is the lack of generalization of the algorithm, and several proposals aim to augment the training dataset to improve the results [Haskins et al., 2020].

The nature of the transformation refers to the model or function that represents the spatial displacement. This is also clasiffied as rigid and non-rigid. The rigid is a general term that covers rigid, affine and projective [Hartley and Zisserman, 2003]. The non-rigid is analog to deformable transformations, for further details see 2.4. The most common transformation are displacement vector fields, which is a dense representation of the deformation. Another common deformable transformations are Bsplines that reduce the transformation

parameters taking advantage of interpolations to approximate the displacements.

The domain of the transformation refers to the transformation coverage within the images. A global transformation applies for every discrete point inside the images. A local transformation split the image in subdomains or regions and have local parameters to represent the transformation in this area.

The optimization refers to the strategy used to solve the cost function in registration. The gradient-based methods models the cost functions as diferrentiable equation and try to find a global minimum. We describe gradient-based methods in section 2.8. The gradient free methods refers to non analytical optimization algorithms such as: bayesian optimization, particle swarm optimization, evolution strategies and genetic algorithms. The later methods are less popular choices for solving registration.

The interaction refers to the automation degree of the registration algorithm. Interactive refers to manually configured and setup algorithms. Featured based or segmentation based solutions may require some user input and therefore are examples of interactive. The semi-automatic algorithms require less input from the users. The automatic algorithms tend to work without any configuration or supervision and the only input are the fixed and moving images. The intensity-based algorithms are a typical example.

In the following sections, we cover the components of the registration algorithm depicted in Figure 2.2 because they represent the fundamental theory of any proposed method.

2.4 Transformations

The term linear transformation refers to a rigid, affine or projective model. The linear transformations are also called the group of projective transformations [Hartley and Zisserman, 2003]. Some authors use indistintivbly the term rigid and non-rigid transformations [Crum et al., 2004, Holden, 2008], where rigid refers to the group of linear transformation.

A linear transformation represents a global transformation and is defined with a matrix product and a vector addition. The matrix represent a modification (scale, shear and rotation) of the coordinates and the vector a displacement. Mathematically we defined the linear transformation as:

$$\varphi(x) := \mathbf{A}\mathbf{x} + \mathbf{t} \tag{2.2}$$

Where $\mathbf{A}$ is called the affine matrix and $\mathbf{t}$ the translation vector. This equation represent the affine transformation. The rigid matrix $\mathbf{R}$ arise for the particular case where $\mathbf{A}$ is an orthogonal matrix, i.e., $A^{-1}A = I$. Some authors define a similarity transformation [Hartley and Zisserman, 2003] which encodes a rigid transformation with a scaling.

A projective transformation comprises the matrix product and the vector addition in a single matrix product. This originates from the use of projective coordinates defined as:

$$\begin{aligned} \varphi(x) :=& \mathbf{Hx} \\ =& \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \end{aligned} \tag{2.3}$$

The projective transformation is a convenient expression to unify the transformation in a single operation but also becomes convenient to simplify transformation compositions or inverse transformation operations.

The linear transformations in 2d and 3d differ in the number of parameters. We summarize in Table 2.1 and 2.2 the expresssions of linear transformations in 2d and 3d respectively. Furthermore, we also describe in Table 2.1 and 2.2 the invariant properties for each transformation.

Group	**Matrix**	**Simplified form**	**Invariant properties**
Projective (8 dof)	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$	$\begin{bmatrix} A & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix}$	Concurrency, collinearity, intersection (1 pt contact), tangency (2 pt contact), inflections (3 pt contact), tangent discontinuities and cross ratio.
Affine (6 dof)	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$	Parallelism, ratio of areas, ratio of lengths on parallel lines, linear combination of vectors.
Similarity (4 dof)	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} sR & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$	Ratio of lengths, angles.
Euclidean (3 dof)	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} R & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$	Length, area.

Table 2.1: 2D linear transformations

For deformable image registration, the typical requirements of a non-rigid transformation are that it is smooth and invertible [Rueckert and Aljabar, 2010], i.e., that it does not lead to effects such as tearing or collapsing regions to a point. Such requirements reflect the variations in anatomy where changes in size and shape are common, but changes in topology are rare. A flexible transformation that allows the imposition of different constraints is a deformation vector field.

Group	Matrix	Invariant properties
Projective (15 dof)	$\begin{bmatrix} A_{[3\times3]} & \mathbf{t}_{[3\times1]} \\ \mathbf{v}^T_{[1x3]} & v \end{bmatrix}_{[4\times4]}$	Intersections and tangency of surfaces in contact. Sign of Gaussian curvature.
Affine (12 dof)	$\begin{bmatrix} A_{[3\times3]} & \mathbf{t}_{[3\times1]} \\ \mathbf{0} & 1 \end{bmatrix}_{[4\times4]}$	Parallelism of planes, volume ratios, centroids.
Similarity (7 dof)	$\begin{bmatrix} sR_{[3\times3]} & \mathbf{t}_{[3\times1]} \\ \mathbf{0} & 1 \end{bmatrix}_{[4\times4]}$	Conic shapes.
Euclidean (6 dof)	$\begin{bmatrix} R_{[3\times3]} & \mathbf{t}_{[3\times1]} \\ \mathbf{0} & 1 \end{bmatrix}_{[4\times4]}$	Volume.

Table 2.2: 3D linear transformation.

A deformation vector field is defined as:

$$\varphi_x : \varphi(x) = \mathbf{x} + u(\mathbf{x}) \tag{2.4}$$

Where $\mathbf{x}$ are the spatial coordinates (in $\mathbb{R}^d$) and $u(\mathbf{x})$ are the displacement vectors.

The B-splines transformation was introduced in the computer graphics field [Lee et al., 1996, Lee et al., 1997] and first applied in registration of medical images (breast MR) [Rueckert et al., 1999]. The B-splines transformation approximate the dense displacement $u(\mathbf{x})$ of equation 7.1 as:

$$u(\mathbf{x}) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3}\beta_x\beta_y\beta_z c_{i,j,k} \tag{2.5}$$

Where c are control points that parametrize the transformation. B-splines represent a popular choice of transformation due to the implicit regularization given because the local approximations [Holden, 2008].

Another transformation is the thin-plate-splines that represent and alternative non-rigid model [Goshtasby, 2005]. The thin-plate-splines transformation has useful smoothing properties and is convenient for undersample regions. Furthermore, it is usually used with sets of anatomical landmarks or paired features which are typically manually located in the images. The transformation is defined as:

$$u(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{B} + \mathbf{I}\sum_{i=0}^{N} F_i r_i^2 log(r_i) \tag{2.6}$$

Where $\mathbf{A}$ and $\mathbf{B}$ are affine matrices, $\mathbf{I}$ is the identity matrix, F_i are the coeficients and $r_i^2 log(r_i)$ is the logarithmic basis function.

2.5 Interpolation

Interpolation is the process of estimating new data points from a set of discrete values. The discrete nature of images usually requires an interpolation operation to estimate the intensity value after a transformation. Interpolation are also used for estimating the local displacement field for transformations with different space coordinates or transformation defined with functional representations (e.g. B-splines).

The interpolation operation is represented by:

$$I(\varphi(x)) \equiv I \circ \varphi(x) \tag{2.7}$$

The interpolation function offers intensity values outside the image grid. For a comprehensive survey covering the theory and performance of interpolation for medical image application, see [Thévenaz et al., 2000]. Multiple methods of interpolation exist. However, the commonly used methods are nearest-neighbor, linear, cubic, b-splines. For an overview in complexity of these interpolation methods see [Guven et al., 2016]. Linear interpolation is the most widely used method in practice with the right balance between complexity and accuracy.

2.6 Similarity Metrics

The similarity metric is a mathematical and functional representation that defines the closeness between two objects. In registration, how close the moving image resemble the fixed image. The mathematical function refers to the $E_{\mathcal{M}}$ term in Equation 7.2. Since registration solves a spatial transformation the ideal similarity metric is a distance function of paired sample points, i.e. a feature-based similarity. When a direct measurement of the desired optimization is not available a surrogate measurement can be used. In registration intensity-based similarity is a surrogate of target registration errors in metric units.

Similarity measures are discussed and mathematically defined by several authors [Hajnal and Hill, 2001], [Modersitzki, 2004], [Crum et al., 2004], [Shams et al., 2010], [Fluck et al., 2011]. The authors define feature-based and intensity-based similarity metrics. As remarked by [Crum et al., 2004], combining geometric features and intensity-based approaches in registration should result in more robust methods. Some authors proposed hybrid similarity metrics that use organ edges and an intensity-based metric [Motegi et al., 2019]. These methods are suitable for radiotherapy, where manual delineations of the organ are available during treatment planning.

The feature-based similarity metric, also called Feature Registration Error (FRE), is defined as an euclidean distance of paired points:

$$E_{\mathcal{M}} := FRE = \frac{1}{N}\sum_{i=0}^{n}(x_{0,i} - \varphi \circ x_{1,i})^2 \tag{2.8}$$

Where N is the total number of paired features i in the fixed x_0 and moving $x1$ spatial domain, respectively. The drawback of the feature-based method is that the available features are typically not dense and cannot describe the deformation accurately. Instead, the intensity-based approach is fully dense because it describes the entire imaging space.

We discuss below intensity-based similarity metrics. Our descriptions are based on [Hajnal and Hill, 2001] who explained the advantages and disadvantages, on [Shams et al., 2010] who precise the modality scope, and on [Fluck et al., 2011] who presented the strengths and weaknesses of intensity-based metrics.

The Sum of Squared Differences (SSD) is one of the most common metrics used in practice. The advantages of SSD are simplicity, fast to compute, and suitability for sequential registrations. The disadvantages of SSD are: only applicable to images with the same intensity profiles (e.g., mono-modality) and limited to Gaussian noise. The SSD metric is defined as:

$$E_{\mathcal{M}} := SSD = \frac{1}{N}\sum_{x\epsilon\Omega_0}^{n}(I_0(x) - I_1 \circ \varphi(x))^2 \tag{2.9}$$

The Sum of Absolute Differences (SAD) has similar advantages to SSD but offers a reduced impact of intensity variations compared to it. The SAD is defined as:

$$E_{\mathcal{M}} := SAD = \frac{1}{N}\sum_{x\epsilon\Omega_0}^{n}|I_0(x) - I_1 \circ \varphi(x)| \tag{2.10}$$

The Normalized Cross Correlation (NCC) is a more robust metric than the former (SSD, SAD) because performs and handle better different noise patterns in the image. The metric is technically restrictive to monomodality but it has been applied when there is a linear relation of intensity values (e.g. CBCT-CT registration [Zachiu et al., 2017], US-CT registration [Pandey et al., 2018]). For implementation purposes some authors use the negative of normalized cross correlation in order to maintain the optimizer as a minimization (e.g. [ITK, 2021]). The NCC metric is defined as:

$$E_{\mathcal{M}} := NCC = \frac{\left(\sum_{x\epsilon\Omega_0}^{n}(I_0(x) - \bar{I}_0)\cdot(I_1 \circ \varphi(x) - \bar{I}_1)\right)^2}{\sum_{x\epsilon\Omega_0}^{n}(I_0(x) - \bar{I}_0)^2 \cdot \sum_{x\epsilon\Omega_0}^{n}(I_1 \circ \varphi(x) - \bar{I}_1)^2} \tag{2.11}$$

The presented intensity-based similarity metrics represent arithmetic operations, the SSD is based on differences and the NCC is based on multiplication. An alternative version is the Ratio of Uniformity (RIU) that based on division. The metric was introduced by [Woods et al., 1992] and is defined as:

$$\begin{aligned} E_{\mathcal{M}} := RIU &= \frac{\sqrt{\frac{1}{N}\sum_{x\epsilon\Omega_0}^{n}(R(x)-\bar{R})^2}}{\bar{R}} \\ R(x) &= \frac{I_0(x)}{I_1 \circ \varphi(x)} \end{aligned} \tag{2.12}$$

The intensity-based similarity problem can be seen a maximization of shared information between the two images. This point of view comes from the information theory perspective. In this field the most commonly used measurement of information is the Shannon-Wiener entropy. The entropy is defined as:

$$H(I) = -\sum_i p_i log(p_i) \tag{2.13}$$

A very successful intensity-based metric that incorporate the entropy metric is Mutual Information (MI). [Mattes et al., 2001] presented the first formulation of non-rigid MI for medical imaging by performing registration to PET-CT images, and the authors latter extended this work in [Mattes et al., 2003]. For multi-modality registration, mutual information has arise as the standard. For further details of MI see [Pluim et al., 2003], an early survey in mutual information algorithms.

The mutual information is defined in terms of entropy as:

$$E_{\mathcal{M}} := MI = H(I_0) + H(I_1 \circ \varphi(x)) - H(I_0, I_1 \circ \varphi(x)) \tag{2.14}$$

The development of Equation 2.14 leads to:

$$E_{\mathcal{M}} := MI = \sum_{i_0}\sum_{i_1} p_{I_0 I_1^{\varphi}}(i_0, i_1) log \frac{p_{I_0 I_1^{\varphi}}(i_0, i_1)}{p_{I_0}(i_0) p_{I_1^{\varphi}}(i_1)} \tag{2.15}$$

MI may lead to numerical problems due to different intensity ranges. For instance, CT scans intensity values in Hounsfield units range from negative values to represent air to positive values to represent bone, while MR scans range only for positive values. An improved alternative for those cases is the normalized version of mutual information (NMI). The metric is defined as:

$$E_{\mathcal{M}} := NMI = \frac{2 \cdot MI}{H(I_0) + H(I_1 \circ \varphi(x))} \tag{2.16}$$

A factor to consider in medical images is to align the organ edges. The edge registration is achieved with intensity-based metrics that incorporate the image gradient. A notorious similarity metric is the Normalized Gradient

Fields (NGF) [Haber and Modersitzki, 2006] that incorporate the gradients that a numerically stable computation. In order to define the metric, the stable normalized gradient is defined first as:

$$G_f(I) := \frac{\nabla I(x)}{\|\nabla I(x)\|_\varepsilon}, \quad \|\nabla I(x)\|_\varepsilon := \sqrt{\nabla I(x)^T \nabla I(x) + \varepsilon^2} \tag{2.17}$$

The NGF is then defined as:

$$E_{\mathcal{M}} := NGF = \sum_{x \epsilon \Omega_0}^{n} \left(G_f(I_0(x)) \cdot G_f(I_1 \circ \varphi(x))\right)^2 \tag{2.18}$$

Other alternatives to incorporate gradients is to compare the gradients with the former intensity-based such as SDD or CC. As an example, [Shams et al., 2007] proposed an intensity-based registration using the cross-correlation of the gradient. A more complex approach is [de Senneville et al., 2016], that proposed a data fidelity term, a similarity metric based on a patch-based integral criterion to compare the image gradients magnitudes and orientations. The overall goal of these alternatives is to improve the edge registration.

We summarize all the intensity-based similarity metrics and their complexity in Table 2.3. We even detail the number of operations involved for consideration in high-performance applications. Similarity metrics such as SSD, SAD, NCC, and RIU have linear complexity while the others such as MI, NMI, and NGF involve high order complexity.

Abbrev.	Equation	Operations				Complexity
		Reduction	Arithmetic	Trfm-Interpolation	Total	
SSD	$\frac{1}{N}\sum_{x \epsilon \Omega_0}^{n}(I_0(x) - I_1 \circ \varphi(x))^2$	n	2n	2n	5n	$\mathcal{O}(n)$
SAD	$\frac{1}{N}\sum_{x \epsilon \Omega_0}^{n}\lvert I_0(x) - I_1 \circ \varphi(x)\rvert$	n	2n	2n	5n	$\mathcal{O}(n)$
NCC	$\frac{\left(\sum_{x \epsilon \Omega_0}^{n}(I_0(x)-\bar{I}_0)\cdot(I_1\circ\varphi(x)-\bar{I}_1)\right)^2}{\sum_{x \epsilon \Omega_0}^{n}(I_0(x)-\bar{I}_0)^2 \cdot \sum_{x \epsilon \Omega_0}^{n}(I_1\circ\varphi(x)-\bar{I}_1)^2}$	5n	5n	2n	12n	$\mathcal{O}(n)$
RIU	$\frac{\sqrt{\frac{1}{N}\sum_{x \epsilon \Omega_0}^{n}(R(x)-\bar{R})^2}}{\bar{R}}$, $R(x) = \frac{I_0(x)}{I_1 \circ \varphi(x)}$	2n	3n	2n	7n	$\mathcal{O}(n)$
MI	$\sum_{i_0}\sum_{i_1} p_{I_0 I_1^\varphi}(i_0, i_1) log \frac{p_{I_0 I_1^\varphi}(i_0,i_1)}{p_{I_0}(i_0) p_{I_1^\varphi}(i_1)}$	n^2	6n	2n	$n^2 + 8n$	$\mathcal{O}(n^2)$
NMI	$\frac{2 \cdot MI}{H(I_0)+H(I_1 \circ \varphi(x))}$	$n^2 + 2n$	6n	2n	$n^2 + 10n$	$\mathcal{O}(n^2)$
NGF	$\sum_{x \epsilon \Omega_0}^{n}\left(G_f(I_0(x)) \cdot G_f(I_1 \circ \varphi(x))\right)^2$	n	$c \cdot n + 5n$	2n	$c \cdot n + 8n$	$\mathcal{O}(c \cdot n)$

Table 2.3: Complexity of intensity-based similarity metrics, where n refers to the number of voxel

2.7 Reguralization

Regularization is considered in deformable image registration to enforce in the displacement vector field the desired properties such as topology preservation,

symmetry, smoothness and invertible. The diffeomorphic definition collects the latter three. The regularization refers to the mathematical function $E_{\mathcal{R}}$ in Equation 7.2.

An indicative parameter of topology preservation is the Jacobian J. The condition $detJ > 0$ is equivalent to determine realizable volumetric changes and to avoid incompressibility.

Modersitzki [Modersitzki, 2004] summarized the regularization methods inspired on mechanical and geometrical models. The elastic, fluid and curvature regularization exposed here are based on Modersitzki. Some regularization schemes are summarized in [Werner et al., 2014].

The classic regularization (references Amit 1994, Cristensen 1994).

$$E_{\mathcal{R}} = \|Lu(x)\|^2_{L^2} \tag{2.19}$$

In elastic registration the regularization is defined as the linearized elastic potential of the displacement $P(u)$, such as:

$$E_{\mathcal{R}} := P(u) = \int_{\Omega} \frac{\mu}{4} \sum_{j,k=1}^{d} (\partial_{x_j} u_k + \partial_{x_k} u_j)^2 + \frac{\lambda}{2}(divu)^2 dx \tag{2.20}$$

Where λ and μ denote the Lamé constants.

In fluid registration the regularization is the elastic potential of the time derivative of the displacement, such as:

$$E_{\mathcal{R}} := P(\partial_t u(x)) \tag{2.21}$$

Approximations of fluid and diffussion -¿ smoothing sigmas.

LDDMM regularization

$$E_{\mathcal{R}} := \int_0^1 \|v_t(x)\|^2_V dt \tag{2.22}$$

SyN regularization

$$E_{\mathcal{R}} := \int_0^{0.5} \left(\|v_1(x)\|^2_L + \|v_2(x)\|^2_L\right) dt \tag{2.23}$$

2.8 Optimization

The optimizer is the iterative component of the registration algorithm that is in charge of finding a local minimum for the cost function E (Equation 7.2). The optimizer search for the best set of parameters in the transformation $\varphi(x)$. The optimization is usually solved with gradient-based methods. For

an overview of the optimization theory, we refer to [Nocedal and Wright, 2006]. Here, we describe the most relevant gradient-based optimizer options.

The straightforward method of optimization is gradient descent. The gradient complexity of the algorithms is $\mathcal{O}(k{\cdot}n)$ with k as the number of iterations. The optimizer has an iterative update defined as:

$$\varphi(x)_{k+1} = \varphi(x)_k - \lambda \cdot \nabla E \tag{2.24}$$

The partial differential equation (PDE) inspired method seeks to impose a constraint in the optimization based on the regularization term. The PDE optimizer is defined as:

$$\varphi(x)_{k+1} = \varphi(x)_k - \lambda(I + \lambda\alpha\nabla E_{\mathcal{R}}) \cdot \nabla E \tag{2.25}$$

The Gauss-Newton is a second-order derivative method. The Gauss-Newton optimizer update is defined as:

$$\varphi(x)_{k+1} = \varphi(x)_k - \lambda \cdot (J_e^T J_e)^{-1} \cdot \nabla E \tag{2.26}$$

Levenberg–Marquardt algorithm is a damped least-squares method that interpolates between the Gauss–Newton algorithm and the gradient descent method [Moré, 1978]. This algorithm represents a more popular optimization choice than the Gauss-Newton method because it can converge better even with poor initialization parameters.

The Newton method include the Hessian computation in the iterative update. The Newton method has a higher complexity than the earlier methods, i.e. $\mathcal{O}(k{\cdot}n^3)$ and each iteration is computationally demanding. The algorithm is defined as:

$$\varphi(x)_{k+1} = \varphi(x)_k - \lambda \cdot H \cdot \nabla E \tag{2.27}$$

The Quasi-Newton approach computes a matrix P as an approximation of the Hessian. Quasi-Newton methods tend to have a complexity $\mathcal{O}(k \cdot n^2)$ lower than the Newton methods. The Quasi-Newton update is defined as:

$$\varphi(x)_{k+1} = \varphi(x)_k - \lambda \cdot P^{-1} \cdot \nabla E \tag{2.28}$$

Mutiple Quasi-Netwon methods exist and the most popular algorithms are: the Boyden's algorithm, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm and the limited memory version L-BFGS, Davidon–Fletcher–Powell formula (DFP) and the Symmetric Rank 1 method (SR1). For a comprehensive review see [Dennis and Moré, 1977].

2.9 Deformable Image Registration Methods

Several surveys on medical image registration and deformable image registration exist. Those surveys are a useful starting point to overview the existing methods. [Goshtasby, 2005] present linear and non-linear registration algorithms specific for feature-based (landmarks) images. [Modersitzki, 2004] focus on numerical methods for medical image registration, specify derivatives of different similarity metrics, and cover landmark-based and intensity-based registration. [Younes, 2010] defines the mathematical background for all the methods behind deformable image registration with diffeomorphism. [Ashburner and Miller, 2015] provide a brief introduction to some diffeomorphic registration algorithms with methods such as: viscous fluid registration, large deformation diffeomorphic metric mapping and geodesic shooting. Several authors report among that the most successful method are intensity-based and, more recently, the deep learning methods. We cover those two strategies in this section.

Voxel Intensity-Based Methods

Voxel intensity-based methods are registration algorithms that use an intensity similarity metric and incorporate a gradient-based optimizer. This solution is sometimes called a variational approach.

The demons algorithm was an early proposal of intensity-based DIR algorithms and was introduced by Thirion [Thirion, 1998]. The algorithm treats registration as a diffusion process similar to an optical flow approach. The algorithm is still a subject of research activity due to its high-performance computation. The demons algorithm can be considered as an approximation of second-order gradient descent on the sum of squared differences (intensity similarity metric) criterion [Pennec et al., 1999] with Gaussian regularization. The diffusion equation is defined as:

$$F(\nabla E(\varphi_x)) \equiv \frac{(I_0 - I_1(\varphi_x))\nabla I_1(\varphi_x)}{\|\nabla I_1(\varphi_x)\|^2 + (I_0 - I_1(\varphi_x))^2} \tag{2.29}$$

The iterative process to update the transformation $\varphi(x)$ is described as:

$$\begin{aligned} compute: &\quad h = F(\nabla E(\varphi(x))) \\ update: &\quad \varphi(x) = \varphi(x) + \lambda \cdot h \\ regularize: &\quad \varphi(x) = G^{\sigma} * \varphi(x) \end{aligned} \tag{2.30}$$

The demons algorithm is one of the most developed approaches of intensity-based DIR. There are other relevant variants of the demons algorithms. Some examples are the fluid demons [Pennec et al., 1999], the spherical demons [Yeo et al., 2009], the Log-demons [Vercauteren et al., 2009] , the LCC demons [Lorenzi et al., 2013]. We develop here the fluid and the log demons.

[Pennec et al., 1999] expose the variant of fluid demons. This proposal applies the Gaussian regularization to the gradient update. Several implementation algorithms in [ITK, 2021] incorporate this type of regularization and conjunction with the diffusion regularization. The fluid demons is defined as:

$$\begin{aligned} compute: &\quad h = G^{\sigma} * F(\nabla E(\varphi(x))) \\ update: &\quad \varphi(x) = \varphi(x) + \lambda \cdot h \\ regularize: &\quad \varphi(x) = \varphi(x) \end{aligned} \tag{2.31}$$

The diffeomorphic demons algorithm is an extension of the original algorithm with a constraint to the transformation to be diffeomorphic [Vercauteren et al., 2009].

$$\begin{aligned} compute: &\quad u(x) = F(\nabla E(\varphi_x)) \\ fluid\,update: &\quad u(x) \leftarrow G^{\sigma_f} * u(x) \\ diffeomorphic: &\quad \varphi(x)' \leftarrow exp(u(x)) \\ diffusion\,update: &\quad \varphi(x) \leftarrow G^{\sigma_d} * \varphi(x)' \end{aligned} \tag{2.32}$$

The expression $exp()$ refers to the intrinsic update on the Lie group of diffeomorphisms, and G refers to a Gaussian kernel. The kernels can be applied to the deformation update $u(x)$ or the overall transformation $\varphi(x)$ and behave as an approximation of fluid or diffusion regularization, respectively.

The diffeomorphic B-splines algorithms are popular choices of DIR because they maintain regularity due to the transformation. [Rueckert et al., 2006] offered a method that use NMI as the similarity metric and incorporate a penalty function in the optimization. [Ashburner, 2007] presented DARTEL, an algorithm with a diffeomorphic B-splines transformation and the optimization is perform with Levenberg-Marquart algorithm. [Tustison et al., 2013] expose the implementation details of how to explicitly define the regularization in a diffeomorphic approach. The methods is available in [ITK, 2021].

Beg et al. introduced the LDDMM algorithm [Beg et al., 2005]. The algorithm has proven to be an effective approach to enforce the diffeomorphic constraint while allowing for large deformations. The result is a symmetric and accurate algorithm but computationally expensive.

The LDDMM cost function is defined as:

$$\underset{v_t:\dot{\varphi}_t = v_t(\varphi_t)}{\arg\min}\ E(I_0, I_1, v_t, \varphi) = \frac{1}{\sigma^2}||I_0 \circ \varphi_1^{-1} - I_1||_{L^2}^2 + \int_0^1 ||v_t||_L^2 dt \tag{2.33}$$

Where $L = \alpha\nabla + \gamma$ is the Cauchy-Navier operator that induces regularity. The cost function gradient is defined as:

$$\nabla E(I_0, I_1, v_t, \varphi) = 2v_t + \frac{2}{\sigma^2} K\left(|Det\varphi_{t,1}|\nabla J_t^0 (J_t^0 - J_t^1)\right) \tag{2.34}$$

Where $J_t^0 = I_0 \circ \varphi_{t,0}$, $J_t^1 = I_1 \circ \varphi_{t,1}$, $K(g) = (L^\dagger L)^{-1} g$ is the regularization function. The algorithm is solved with a gradient descent approach to solve the velocity field and the deformation field is found by integration. The algorithm steps are summarized as:

$$\begin{aligned} update: &\quad v_t(x) = v_t(x) + \lambda h \\ integrate\ backward: &\quad v_t \rightarrow \varphi_{t,1} \\ integrate\ forward: &\quad v_t \rightarrow \varphi_{t,0} \\ compute: &\quad h = \nabla E(I_0, I_1, v_t, \varphi) \end{aligned} \tag{2.35}$$

The algorithm rapidly evolved to incorporate different similarity metrics such as CC [Avants et al., 2008] and MI [Lorenzen et al., 2006]. The LDDMM complexity is high and its computationally demanding in terms of memory and processing capabilities. The stationary LDDMM algorithm proposed by [Arsigny et al., 2006] is an alternative to reduce computational complexity while a memory efficient version is proposed by [Polzin et al., 2016]. These algorithms represent attempts to make LDDMM available and suitable for multiple application where the time is a constraint.

Avants et al. [Avants et al., 2008] introduce Symmetric Image Normalization (SyN) a DIR algorithm. The algorithm is available with ANTs toolbox [ANTs, 2021] and supported by ITK toolbox [ITK, 2021]. SyN algorithm uses cross correlation as the similarity metric $E_{\mathcal{M}}$ defined as:

$$CC(I_0, I_1, \varphi(x)) \equiv \frac{\left(\sum_{x \epsilon \Omega_0}^{n} (I_0(x) - \bar{I}_0) \cdot (I_1 \circ \varphi(x) - \bar{I}_1) \right)^2}{\sum_{x \epsilon \Omega_0}^{n} (I_0(x) - \bar{I}_0)^2 \cdot \sum_{x \epsilon \Omega_0}^{n} (I_1 \circ \varphi(x) - \bar{I}_1)^2} \tag{2.36}$$

The regularization $E_{\mathcal{R}}$ is defined as the Sobolev norm of the velocity field v_t, similar to the term proposed by Beg et al. [Beg et al., 2005]. The velocity field is related to the deformation field $\varphi(x)$ with the ordinary differential equation $\frac{d\phi_t(x)}{dt} = v_t(\phi_t(x), t)$, with solution $\varphi(x) = \phi_0 + \int_0^1 v_t(\phi_t(x))dt$. In order to address consistency and reduce computation time, SyN algorithm divide the deformation field in two, such as: $\varphi(x) = \phi_1 \circ \phi_2^{-1}(x, t)$. Consequently, the regularization term involves two velocity fields. The mathematical setup is expressed as follows:

$$E_{\mathcal{R}}(\varphi(x)) = \int_0^1 \|v_t\|_L^2 \, dt = \int_{t=0}^{0.5} \left(\|v_1(x,t)\|_L^2 + \|v_2(x,t)\|_L^2 \right) dt \tag{2.37}$$

Where:
$L = a\nabla + bId$, is a linear differential operator that induces regularity.

The optimization in SyN is described by the minimization of the energy equation:

$$\varphi^*(x) \equiv \arg\min_{\varphi(x)} \int_{t=0}^{0.5} \left(\|v_1(x,t)\|_L^2 + \|v_2(x,t)\|_L^2 \right) dt + \int_\Omega CC(I_0, I_1, \varphi(x)) d\Omega \tag{2.38}$$

SyN algorithm is a successful algorithm that is ranked among the best intensity-based approaches for multiple studies [Klein et al., 2009a, Murphy et al., 2011]. Some authors still consider SyN as an state-of-the-art method.

Deep Learning Methods

The core architecture of the deep learning methods usually has a CNN encoder-decoder setup. The pipeline is similar to the one presented in Figure 2.2 with the difference that the CNN model is placed before the optimizer.

Litjens et al. presented an early survey on deep learning approaches for medical image analysis [Litjens et al., 2017] and registration. They reported the first deep learning strategies that incorporate a deep metric on registration algorithms [Wu et al., 2013a, Simonovsky et al., 2016]. However, the first strategies to estimate the transformation parameters were [Miao et al., 2016, Yang et al., 2016], i.e., proper registration algorithms.

Haskins et al. offered a modern survey with medical image registration [Haskins et al., 2020]. The authors collect different architectures of deep learning approaches. The more relevant are supervised [Yang et al., 2016, Yang et al., 2017b], unsupervised [Zhang, 2018, Balakrishnan et al., 2019] and deep feature-based strategies [Wu et al., 2013a, Simonovsky et al., 2016, Cheng et al., 2018].

Some well-known frameworks of deep learning DIR algorithms are QuickSilver [Yang et al., 2017b], and VoxelMorph [Balakrishnan et al., 2019]. QuickSilver is a supervised method developed with a CNN encoder-decoder architecture. The algorithm is evaluated with registrations of Brain MRIs. VoxelMorph is an unsupervised method and has a CNN architecture similar to U-Net. The authors impose a diffeomorphic transformation in their method. The application was also the registration of Brain MRIs. These methods are the foundations of multiple deep learning DIR algorithms and are still considered state-of-the-art.

2.10 Validation and Accuracy

Since deformable image registration is an ill-posed problem, there is no unique solution to find the spatial transformation. Thus, the goal is to obtain a metrical unit to validate how aligned the fixed and moving images are. The desired direct metric is the Target Registration Error (TRE). Defined as the euclidean distance between paired points (features or landmarks) in both input images (fixed and moving). The TRE, also called residual euclidean distance, is defined mathematically as:

$$TRE(\mathbf{x}_f, \mathbf{x}_m) = \|\varphi(x) \circ \mathbf{x}_f - \mathbf{x}_m\| \tag{2.39}$$

The TRE requires ground truth paired points in both images, which is difficult to have in practice or manually intensive [Brock et al., 2017]. Instead of using the TRE, indirect metrics could be used to validate registration. However as exposed by [Rohlfing, 2011], a properly evaluated registration algorithm has to be evaluated with features. A new way to compute TRE is arising, consisting into automatically detect and match feature points, e.g. [Vickress et al., 2016], [Yang et al., 2017a]. These feature mathching methods represent an automatic solution for validation purposes even if the registration algorithm itself was optimized for any other similarity metric.

[Crum et al., 2006] presented an overview of overlap measures for evaluation and validation in medical image analysis. The authors include a set of indirect metrics that are described below.

The Dice Similarity Coeffient (DSC) serves to quantify the overlapping of the segmented structures. The DSC is defined as:

$$DSC(X,Y) \equiv \frac{2|X \cap Y|}{|X| + |Y|} \tag{2.40}$$

where X is a control segmentation structure and Y is the estimated structure.

A similar validation metric to DSC is the Jaccard Index (JAC). The JAC metric is defined as:

$$JAC(X,Y) = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \tag{2.41}$$

THe DSC is widely used for specific segmentation measurement but are also used for evaluation of registration algorithms when features are not available.

The Hausdorff distance is defined as:

$$d_H(X,Y) \equiv \max \left\{ \max_{x \in A_X} \left(\min_{y \in B_Y} \|(x-y)\| \right), \max_{y \in B_Y} \left(\min_{x \in A_X} \|(y-x)\| \right) \right\} \tag{2.42}$$

where x is a point in the contour A_X related to structure X, and y is a point in the estimated contour B_Y related to structure Y.

The mean contour distance is defined as:

$$d_H(X,Y) \equiv \max \left\{ \underset{x \in A_X}{mean} \left(\min_{y \in B_Y} \|(x-y)\| \right), \underset{y \in B_Y}{mean} \left(\min_{x \in A_X} \|(y-x)\| \right) \right\} \tag{2.43}$$

The centroid distance is useful in tracking or localization applications using registration. The centroid distance is defined as:

$$d_{com}(X,Y) \equiv \|X_{com} - Y_{com}\| \tag{2.44}$$

where X_{com} is the control structure center of mass and Y_{com} the estimated structure center of mass.

Some researchers include the voxel similarity metric (see section 2.6) to measure the registration accuracy. The use of similarity measure as metric is discouraged and only represent an indicative measurment [Rohlfing, 2011].

The validation procedure of registration is not an straithforward task. It requires one or multiple metrics and preferebly to include the TRE. We shortly describe some comprehensive examples of registration validation available in the literature. [Klein et al., 2007] evaluate different similarity metrics, multiresolution and sampling in registration. [Klein et al., 2009a] present and evaluation of 14 deformable image registration with brain MRI images. [Mogadas et al., 2018] offer a comparison of 6 open source DIR libraries. ANTs, Elastix, VarReg, DIRART, NiftyReg, Plastimatch. The authors found that the most accurate open-source options are the ANTs and Elastix. The mentioned methods are methodological examples of registration validation.

The publicly available datasets represent a powerfull asset to compare the performance of different registration algorithms between multiple authors. We describe below some well-known datasets.

RIRE: The Retrospective Image Registration Evaluation Project [West et al., 1997] described in the retrospective evaluation of inter-subject brain registration [Hellier et al., 2003] is an early multimodal dataset that contains CT and MR images avialable at [RIRE, 2021].

DIR-LAB: [Castillo et al., 2009b] introduce a framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. This framework is extended in [Castillo et al., 2009a] and [Castillo et al., 2013]. The dataset is available at [dir-lab, 2021]. The dataset contains 4DCT images of 10 patients with manually annotated landmarks between each phase. The larger amount of paired landmarks are located between the inspiration and the expiration phase.

POPI: This data is fully described in [Vandemeulebroucke et al., 2011]. The POPI-model is a point-validated pixel-based breathing thorax model. Available at [POPI-model, 2021]. The dataset contains 4DCT images of 6 patients with paired landmarks similar to DIR-LAB.

BITE: Brain Images of Tumors for Evaluation database, this dataset is described in [Mercier et al., 2012]. An MRI and Ultrasound dataset with manually annotated landmarks. The dataset is available at [BITE, 2021].

2.11 Considerations for Real-Time Implementations

The most promising applications for real-time medical imaging registration are for surgery and image guided radiotherapy. In surgery field, minimal invasive

procedures avoids the trauma and faster recovery time, this is achieved with image guidance during surgery [Shams et al., 2010]

In radioteraphy the aid comes from usign medical images to improve the accuracy in the radiation applied to the target tissue. Uncertainties due to motion during radio therapy can cause underdosing the target and/or overdosing organs-at-risk [Caillet et al., 2017]

As remarked by [Murphy, 2004] with regards to radiotherapy, the best way to accommodate intra-fraction tumor motion would be to dynamically shift the dose in space so as to follow the tumor's changing position. This is referred as real-time tracking. A radiotherapy system able to use real-time tracking can eliminate the need for a tumor motion margin in the dose distribution while maintaining a 100% beam duty cycle for efficient dose delivery.

A radiotherapy tracking system requires four principles: (1) determine the tumor position, (2) anticipate the tumor motion to allow for time delays in realign-ment of the beam, (3) reposition the beam, and (4) adapt the dosimetry to allow for changing tumor and critical structure configuration. Among those four principles, medical image registration can contribute to the first two.

Regarding algorithm parallelization, two major surveys have been present of how to implement medical image registration. The first one by [Shams et al., 2010] and the other one by [Fluck et al., 2011]. As remarked by [Shams et al., 2010], the computational bottleneck of registration is not the optimization steps but the computation of the transformation-interpolation and the similarity metric.

The interpolation-transformation operation is inherently parallel and requires matrix operations. Furthermore, they are straightforward to implement in processing kernels. The similarity metric allows vectorization and is suitable to implement in parallel hardware. The operation involves and piece-wise arithmetic and a reduction The complexity is typically $\mathcal{O}(n)$.

The optimizer is a gradient-based with a inherently serial dependency. The optimization has little parallelization opportunities. One option is to compute in parallel the partial derivatives, however the expected gain may be limited. A more relevant feature in optimization is a multiresolution or multipyramid framework [Thevenaz et al., 1998]. This strategy improves the convergence time in the algorithms.

Some examples of high performance algorithms are implemented on the GPU. The computational efficiency of the demons algorithm have promoted its implementation in parallel architectures like the GPU. Some examples are [Sharp et al., 2007, Muyan-Ozcelik et al., 2008, Samant et al., 2008, Gu et al., 2009, Oh et al., 2011]. [Gu et al., 2009] expose the most comprehensive evaluation with six (6) demons algorithms on GPU. The methods are called passive force (PF), evolved passive force (EPF), active force (AF), double force (DF), adjusted doble force (ADF) and inverse consistency (IC). The passive force is the original Thirion's proposal where the gradient of the moving image

is computed only at start. The evolved passive force is also proposed by [Thirion, 1998] where the fixed and moving gradients are used to compute the update. The active force is the same passive equation but in this case the gradient of the moving image is computed iteratively. The double force is the summation of the passive and active force. The adjusted double force is the inclusion of a normalization factor in the double force equation as proposed by [Pennec et al., 1999]. Finally the inverse consistency is the algorithm proposed by [Yang et al., 2008]. The algorithms are tested with dir-lab dataset [dir-lab, 2021]. In terms of performance the methods are faster as they were described (i.e. PF the fastest towards IC being the slower). Computational time range from 7 to 11 seconds of full 3D registration. In terms of accuracy the best performing algorithms are ADF, AF, PF and EPF. The authors also suggest that the latter four algorithms are all suitable for radiotherapy applications.

2.12 Open Source Implementations

Since medical image registration is a mature field, several open-source implementations exist. The most well-known framework is the Insight Segmentation and Registration Toolkit (ITK). ITK is an open-source library developed in C++ with an object-oriented design. The source code is available at [ITK, 2021]. ITK is a cross-platform system that provides developers with an extensive suite of software tools for image analysis. Regarding registration, ITK supports multiple configurations of transformations, interpolations, similarity metrics, and optimizers. Thus, ITK is a versatile tool that works as the cornerstone for other registration frameworks.

ANTS: Advanced Normalization Tools is an ITK based toolkit exposed in [Avants et al., 2008]. The source code is available at [ANTs, 2021]. ANTs is based on ITK. SyN algorithm is included in ANTs and was the best overall algorithm on a brain MRI comparison in [Klein et al., 2009a] and on lung comparison [Murphy et al., 2011].

Elastix: A general registration toolkit presented in [Klein et al., 2010]. The source code is available at [Elastix, 2021]. The open source software is based on ITK, it consists of a collection of algorithms that are commonly used to solve (medical) image registration problems. The parametrization is posible with configuration files that are shared publicly.

NiftyReg: an open-source software for efficient medical image registration. The source code is available at [NiftyReg, 2021].

Plastimatch: an open source software for image computation. The main focus is high-performance volumetric registration of medical images. The source code is avialable at [Plastimatch, 2021].

CLAIRE: a novel library that support multi-node and memory distributed registration described in [Mang et al., 2019, Brunn et al., 2021]. The source code is avialable at [CLAIRE, 2021]

2.13 Discussion

We presented a modern survey on medical image registration focusing on deformable image registration. Our approach defines a common mathematical framework of registration with a novel approach that considers the algorithms' complexity. Furthermore, we discussed and detailed some well-known intensity-based and deep learning methods. Finally, we exposed other aspects of registration such as the validation metric, publicly available datasets, consideration for high-performance implementations, and open-source implementations. The result is a comprehensive registration study for medical imaging that works as a theoretical background.

Chapter 3

Registration Accuracy with Liver

The work presented in this chapter is based on a paper that is to be submitted as: Tascón-Vidarte, J. D., Stick, L. B., Josipovic, M., Risum, S., Jomier, J., Erleben, K., Vogelius, I. R., & Darkner, S. (2021). Accuracy and consistency of deformable image registration in 4DCT for tumor motion estimation in liver radiotherapy planning.

3.1 Abstract

We investigate the accuracy of deformable image registration for tumor localization in liver stereotactic body radiotherapy (SBRT). We included eight patients, recieving SBRT for liver metastases within a retrospective clinical study. Each patient had three fiducial markers implanted. 4DCT images were used to capture breathing motion. The liver and the tumor were delineated in the midventilation phase, their positions in the other phases were estimated with deformable image registration. Referenced and sequential registrations were tested. The accuracy of registration was evaluated with the fiducial markers as gold standard. The registration errors related to measured versus estimated fiducial markers showed a mean value less than $1.6mm$. Positions of some fiducial markers appeared not stable on the 4DCT through out the respiratory phases. Markers' center of mass tends to be a more reliable measurement. Distance errors of tumor location based on registration versus markers center of mass were less than $2mm$. There were no statistically significant differences between the reference and the sequential registration, i.e. consistency and errors were of comparable magnitude to resolution error. We demonstrated that deformable image registration is accurate up to resolution level for locating the tumor in the liver during breathing motion.

3.2 Introduction

Stereotactic body radiotherapy (SBRT) is one of the treatment modalities used for managing tumors in the liver [Aitken and Hawkins, 2015]. The SBRT delivery is, however, affected by the respiratory motion of the target [Worm et al., 2013], resulting in potential underdosing of the target and/or overdosing to organs-at-risk [Romero et al., 2009, Swaminath et al., 2015, Caillet et al., 2017]. Different methods such as breath-hold, shallow breathing, and abdominal compression have been proposed to control or reduce the effect of the breathing motion [Høyer et al., 2012, Dhont et al., 2020]. Nonetheless, breathing motion has to be accounted for during planning and treatment.

The feasibility of using deformable image registration (DIR) to estimate breathing motion and tumor positions in the liver has been demonstrated [Brock et al., 2006], aiding the evaluation of the positional uncertainties in both target and risk organs, induced by the respiration [Mogadas et al., 2018]. Although DIR has been increasingly used in radiotherapy [Rosenman et al., 1998, Brock et al., 2017], its accuracy is difficult to quantify and is considered as a non-trivial task [Murphy et al., 2011]. This difficulty is caused by the lack of ground truth data. Several DIR algorithms have been developed,

but they have been mostly evaluated on images from the thorax [Murphy et al., 2011] which have manually annotated landmarks [Castillo et al., 2009b, Vandemeulebroucke et al., 2011]. In contrast, the accuracy in low contrast organs (abdominal area) has only been studied sparsely [Mogadas et al., 2018].

Liver tumors are often not visible without admission of intravenous contrast on either CT or cone-beam CT (CBCT) [Beddar et al., 2008, Bertholet et al., 2016]. Therefore gold fiducial markers can be implanted in the tumor vicinity [Wunderink et al., 2010] and used as a surrogate of the tumor position during SBRT [Seppenwoolde et al., 2011]. Image guidance with fiducial markers is reported as a more accurate method compared to image guidance with the liver contour or diaphragm position [Shimohigashi et al., 2017]. Therefore, we use fiducial markers to evaluate tumor registration accuracy in our study.

Regarding the DIR algorithms on the liver, the finite element model-based DIR proposed by Brock et al. [Brock et al., 2005] is the only algorithm to be extensively evaluated in CT with low contrast organs [Brock et al., 2006, Velec et al., 2012, Velec et al., 2017]. The MIDRAS study [Brock et al., 2010] compared several algorithms for the abdomen and the liver with only one 4DCT scan. Finite element model-based DIR methods seem to perform slightly better than intensity-based algorithms for the liver. The drawback is that one patient is not representative of the general performance of the algorithms. Furthermore, the intensity-based methods here only included demons [Vercauteren et al., 2009] and B-splines [Rueckert et al., 2006, Rietzel and Chen, 2006] algorithms and not the modern large diffeomorphic metric mapping algorithms [Beg et al., 2005, Avants et al., 2008]. An advantage of intensity-based DIR is that it computes registration without prior information or preprocessing. In contrast to the finite element model-based DIR that requires organ segmentation, surface mesh conversion, boundary conditions, and manual parametrization of material properties [Brock et al., 2005]. We consider that a modern intensity-based algorithm requires a proper evaluation of registration accuracy for liver radiotherapy due to its implicit automation capabilities.

In this paper, we aim to demonstrate that the use of intensity-based DIR is accurate in estimating the position of liver tumors in low contrast images. We applied DIR to respiratory correlated 4DCT and used implanted fiducial markers as a surrogate of tumor locations.

3.3 Materials and Methods

Patients

This study was based on data from eight patients treated with SBRT for metastases in the liver at Rigshospitalet (Copenhagen, Denmark) between March 2018 and September 2019. The patients were selected from an ethical committee approved study (approval nr. H-17033786) that investigates liver

SBRT in breath-hold. See details on patient selection in Appendix A. Patient Selection.

Image acquisition

Respiratory correlated 4DCT with intravenous contrast injection was performed for all patients on a SOMATOM Definition AS scanner (Siemens Healthineers, Germany). 4DCT image data were phase-sorted into ten phase bins throughout a respiratory cycle based on an external respiratory signal monitored with Real-Time Position Management (RPM, Varian Medical Systems, Palo Alto, CA). The slice separation in each phase of the 4DCT was $2mm$. The image resolution in each slice was 512×512 pixels and a pixel size of 0.98×0.98 mm.

In addition, MR scans in deep inspiration breath-hold without visual guidance were performed for all patients to achieve better visualization of the tumor. The MR images (T1 VIBE with iv contrast, T1 FL2D, T2 HASTE, TrueFISP) were adquired in a 1.5T SIEMENS MAGNETOM Avanto scanner (Siemens Healthineers, Germany).

Delineations

The mid-ventilation phase of the 4DCT was determined based on fiducial marker (tumor) motion for each patient and used for treatment planning [Wolthaus et al., 2006]. Further details of treatment are out of the scope of the paper. The mid-ventilation phase was rigidly registered (in 6D) with the MR scan focusing on the tumor area to guide delineation of the tumor. The Gross Tumor Volume (GTV) was delineated on mid-ventilation phase by a senior radiologist and approved by a senior oncologist. Risk organs, including the liver, were also delineated. Fig 3.1 shows the mid-ventilation phase for patient 2.

Fiducial markers

Three fiducial markers were implanted percutaneously near the tumor in each patient using ultrasound guidance. The fiducial markers were placed within one week prior to the imaging for treatment planning. Goldlock fiducial markers (cylinder with star-shaped cross-section, 1 x 3 mm, Beampoint AB, Sweden) were implanted in the first five patients and Gold Anchor fiducial markers (cylinder with multiple cut-outs, 0.4 x 10 mm, Naslund Medical AB, Sweden) were implanted in the latter three patients.

The centroid position of each fiducial marker was calculated automatically in each phase and chosen as its ground truth locations. The automatic procedure consisted of three steps. First, a threshold of 500 Hounsfield units was applied to the images. Second, the fiducial markers were delineated in the

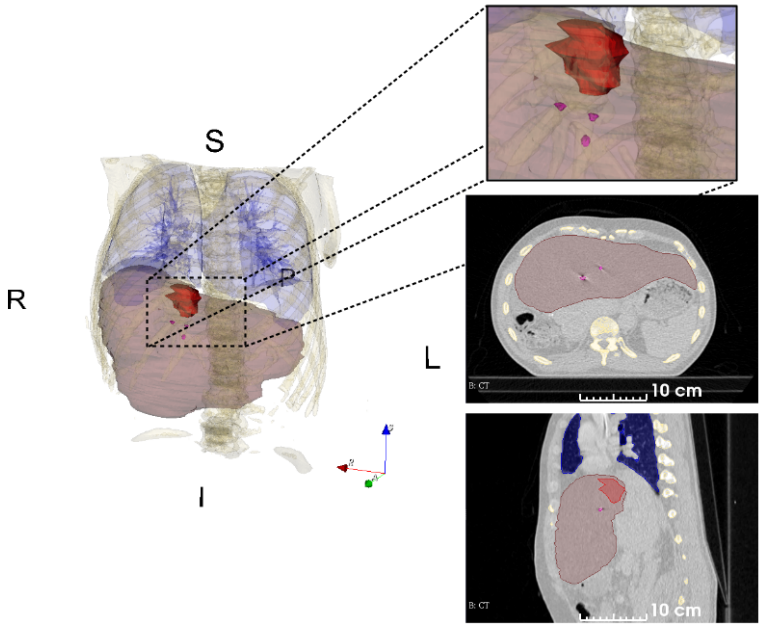


Figure 3.1: **A** single CT image of the 4DCT with organ delineations. The image corresponds to mid-ventilation phase of patient 2. Delineations of liver (translucent brown), gross tumor volume (red) and fiducial markers (magenta). The fiducial markers were implanted near the tumor. The delineated volumes were used to compute the centroids of the liver, the gross tumor volume (GTV) and the fiducial markers

axial plane in all 4DCT phases. Third, the delineations were used to create a volume of the fiducial marker and then the centroid is computed. The accuracy of this method was compared to the manual method for locating the center of the fiducial markers [Beddar et al., 2007]. The comparison was made with all 4DCT phases of a single patient and the values were approximately the same (differences $< 0.1mm$). The uncertainty of the fiducial markers centroids were expected to be greater in the SI plane due to the 2 mm slice separation.

The fiducial markers center of mass (COM) are calculated for each patient as the average position of the three fiducial markers. From here, we will refer to this measurement simply as markers COM, while 'fiducial markers' refer to the (individual) positions of the markers centroids.

Registration Algorithm

The publicly available Symmetric Image Normalization algorithm (SyN) [Avants et al., 2008] was used as the image registration algorithm, being one of the best publicly available algorithms for DIR [Mogadas et al., 2018]. Briefly, the

SyN algorithm uses Cross Correlation (CC) as similarity metric and the L2 norm of the velocity field as regularization. The transformation or displacement field is then computed by integration of the differential equation that relates velocity with displacement. The mathematical details are exposed in Appendix B. Registration Algorithm.

In this work, the DIR was computed in two ways, depicted in Fig 3.2. The first method used a single image as reference and the remaining images were subsequently registered to this reference image. The reference image was chosen as the 50% phase of the 4DCT which approximately corresponded to the end of expiration. The second method registered all the images sequentially. For the sake of simplicity, we will refer to the methods as reference and sequential registration through out this manuscript. We compared the registration errors of the two methods in order to investigate if sequential registration introduces a drift error due to composition of the estimated deformations.

The liver and GTV delineations were propagated to the other phases of the 4DCT with the DIR algorithm. The appropriated transformation was chosen as the lowest number of composition of deformations from the mid-ventilation phase to the other phases. After DIR, a volume was created in each phase for each patient to compute the center of mass (COM) of the tumor and the liver. From here, we will refer to this positions as liver COM and GTV COM. This methodology is referred as segmentation based on DIR or contour propagration [Hardcastle et al., 2013]. The accuracy of this estimation is therefore related to the DIR error.

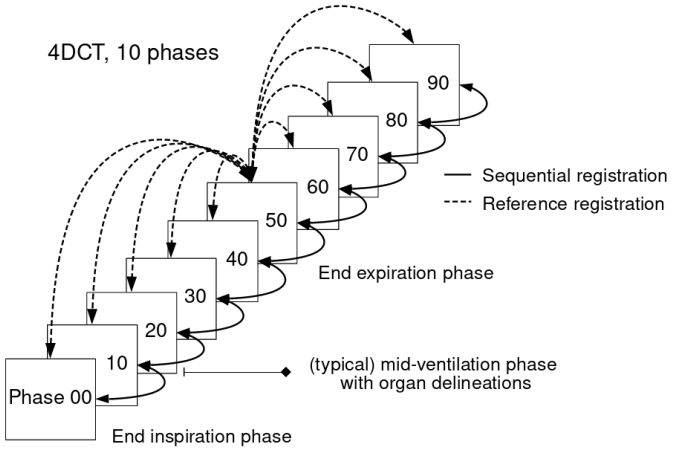


Figure 3.2: 4DCT scan with a representation of reference and sequential registration. Delineations of the liver and gross tumor volume (GTV) are propagated from mid-ventilation phase to the other phases. The 50% phase corresponds (approximately) to end expiration and chosen as the reference image.

The DIR algorithm is voxel intensity-based. In this scenario, the fiducial

markers became a problematic surrogate for registration, since they create high intensity values around their location [Brook et al., 2012]. In order to overcome this, image inpainting was applied to remove the high intensity voxels.

Image Inpainting

The image inpainting was performed with the algorithm proposed by Telea et. al. [Telea, 2004]. For extended details of the inpainting see Appendix C. Inpainting. After all the 4DCT images are inpainted, the DIR algorithm is applied to all the patients. Only the reference registration approach is used here. The measured fiducial markers coordinates are still used for computing the registration errors. The registration of the inpainted images was compared to the registration without the inpainting. In case of similar results we could conclude that the registration was not driven by the high intensities of the fiducial markers.

Distance Metrics

The fiducial markers were used as the direct measurement to evaluate the accuracy of image registration. Two distance metrics were used: the first one for DIR evaluation and the second one to verify the relative distance variations of markers, tumor and liver. Fig 3.3 presents all the points used to compute the distance metrics.

The first metric was the registration error calculated as the euclidean distance between the ground truth positions of the fiducial markers and the estimated positions of the fiducial markers. The ground truth coordinate was a point $x_{f,j}$, where the subscripts f and j refer to the particular fiducial marker and the image phase respectively. The transformation found by registration φ_j was applied to the fiducial markers in the reference phase $x_{f,\text{ref}}$, producing the estimated fiducial markers $\hat{x}_{f,j} = \varphi_j \circ x_{f,\text{ref}}$. The metric is defined as:

$$r_{\text{error}} \equiv \left\| x_{f,j} - \underbrace{\varphi_j \circ x_{f,\text{ref}}}_{\hat{x}_{f,j}} \right\| \tag{3.1}$$

The registration errors are calculated for the reference and sequential registration algorithms, as well as for images registered with and without inpainting. The statistical tool used to compared the data is the Kolmogorov-Smirnov test [Massey Jr, 1951].

The second metric was the relative distance variation. In this case, the distance between two points was compared to check if the distance was the same across the phases. The metric was calculated between the fiducial markers, between the GTV COM and the fiducial markers, between the GTV COM and the markers COM, and between the GTV COM and the liver COM. As

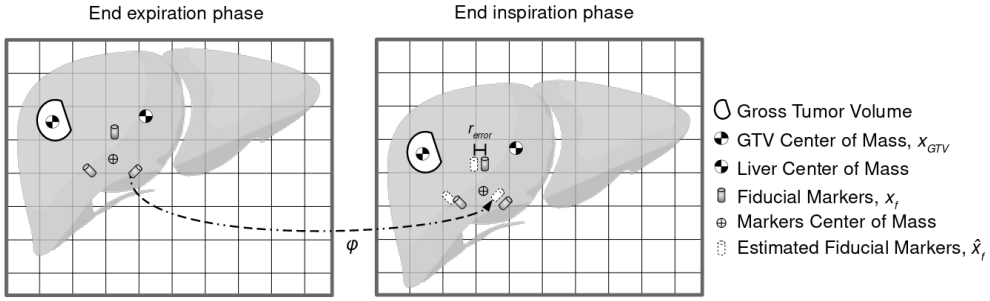


Figure 3.3: **L**andmarks used to compute registration and distance metrics. Center of mass (COM) are calculated for fiducial markers, gross tumor volume (GTV) and liver. GTV and liver volumes are estimated in all phases via contour propagation based on DIR from the mid-ventilation phase. Estimated fiducial markers' positions are computed with the transformation φ found by registration. The registration error (r_{error}) from Eq (3.1) is illustrated here as the euclidean distance of estimated fiducial markers' position versus measured fiducial markers' position.

an example, the relative distance of the fiducial marker 1 and fiducial marker 2 is presented here, as:

$$d \equiv \|(x_{f2,j} - x_{f1,j})\| - \|(x_{f2,\text{ref}} - x_{f1,\text{ref}})\| \tag{3.2}$$

3.4 Results

Breathing motion

Table 5.1 summarizes breathing motion characteristics of the patients. The average displacement of the fiducial markers between expiration and inspiration for all the patients was $\mu = 7.9$ mm with a standard deviation of $\sigma = 2.9$ mm. The fiducial markers distances to the GTV COM ranged from 17 to 55 mm, with mean value $\mu = 32$ mm and standard deviation of $\sigma = 10$ mm.

The sequential registration was used to estimate volume variations in the liver and the GTV calculated on every phase. The baseline volume measured in mid-ventilation phase is shown in Table 5.1. The patients in this study had a liver volume coefficient of variation (CV) of 6.5% in average and a maximum of 12.5% (patient 5). The GTV coefficient of variation (CV) was on average 1.0% and a maximum of 2.4% (patient 4). The registration indicated that compressibility in general was low for the liver, and was higher for the liver tissue compared to the tumor tissue.

Table 3.1: **S**ummary of patients information.

	Patient	1	2	3	4	5	6	7	8
Tumor location (geometric)		I-A-L	S-A-R	S-P-R	S-P-L	I-P-R	S-A-L	S-P-L	I-P-L
Breathing cycle (mean) [s]		2.8	3.5	3.0	4.1	5.8	4.0	5.3	3.3
Tumor displacement [mm]	LR	1.2	0.9	3.6	2.5	2.5	1.3	0.4	0.8
	AP	1.3	0.6	3.4	0.5	3.3	1.8	2.3	2.9
	SI	5.1	7.8	10.2	6.3	12.6	3.7	8.2	7.9
	3D	5.4	7.9	11.2	6.7	13.3	4.0	8.5	8.3
Distance GTV COM to [mm]	Marker 1	20.6	19.0	30.8	43.5	33.8	49.0	25.0	30.8
	Marker 2	29.5	24.0	41.1	41.5	28.3	33.1	21.1	25.7
	Marker 3	17.9	41.6	18.0	44.6	54.9	31.5	36.9	40.8
	Marker COM	22.3	40.3	26.6	32.2	26.1	18.1	38.7	29.9
Distance between markers [mm]	Marker 1 - 2	12.0	15.9	14.1	30.4	25.3	19.6	23.3	21.3
	Marker 2 - 3	16.6	29.5	21.6	35.7	20.1	15.2	17.3	31.7
	Marker 3 - 1	29.2	58.9	31.8	14.2	41.9	40.8	19	43.7
GTV [cm3]	Mid-ventilation	7.6	29.0	29.3	11.6	2.4	4.0	4.5	64.8
	Mean	7.4	29.7	28.8	11.4	2.5	4.2	4.4	65.4
	CV [%]	0.5	1.2	1.7	2.4	0.1	0.4	0.9	1.0
Liver volume [cm3]	Mid-ventilation	1904	2324	1343	1458	912	2156	1744	1596
	Mean	1908	2321	1343	1458	903	2172	1722	1594
	CV [%]	7.6	1.4	3.7	5.1	12.5	10.1	10.7	3.5

Table notes: Tumor location is the geometric octant of where the tumor is with regards to the liver COM. The abbreviations used correspond to Superior-Inferior, Anterior-Posterior and Left-Right. Breathing cycle time was determined from the external respiratory signal, acquired during the 4DCT scan. Tumor displacements are computed using the markers COM as surrogate of the tumor. Distance of GTV COM to fiducial markers and between markers are measured in the mid-ventilation phase. GTV and liver volumes are estimated for every phase based on DIR. COM = center of mass, CV = coefficient of variation.

Accuracy and Consistency of Registration

We tested the registration algorithm with end expiration and end inspiration phases (the 50% phase and 0% phase were used for all patients) where the maximum displacement occurs. The algorithm was parametrized to achieve convergence in this scenario. For further details see Appendix D. Registration Parametrization. The average time of the DIR algorithm was 170.5 minutes (per registration), running in a workstation with two processors Intel Xeon Silver 4110 (8 cores, 16 threads each cpu).

We tested the registration errors (Eq (3.1)) in all patients and all markers per phase. The registration errors of the estimated fiducial markers versus measured fiducial markers had a maximum mean value of $1.4mm$ for the referenced registration and $1.6mm$ for the sequential registration, see Fig 3.4a and 3.4b. The corresponding median values were up to $1.2mm$ and $1.4mm$ respectively. An increased mean and standard deviation was noted for the images further away of the reference phase for both reference and sequential registration. There were no statistically significant differences per phase related to the registration method. See more details in Appendix E. Statistical

Results.

The maximum registration errors were 5.3mm and 3.7mm for the referenced registration and the sequential registration respectively. More details are presented in Appendix F. Registration Errors per Patient. The registration errors always have at least one fiducial marker with an error lower than 2mm for each patient.

The inpainted images were tested only for the referenced registration case with the registration error metric and we found no statistically significant differences compared to the original images, see Fig 3.4c and Appendix E. Statistical Results.

As we observed with the registration errors that the average maximum obtained was equivalent or lower than the resolution in the axial slices. Since the reference and the sequential registration did not produce statistically significant differences this indicates a consistency in the registration method to accurately align the images.

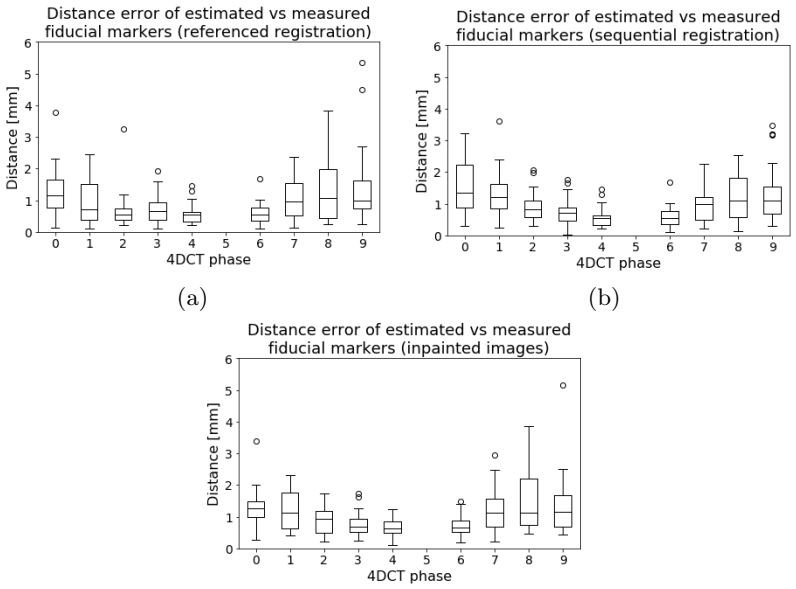
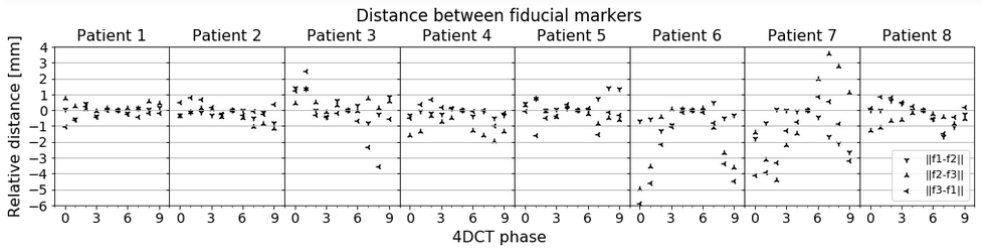


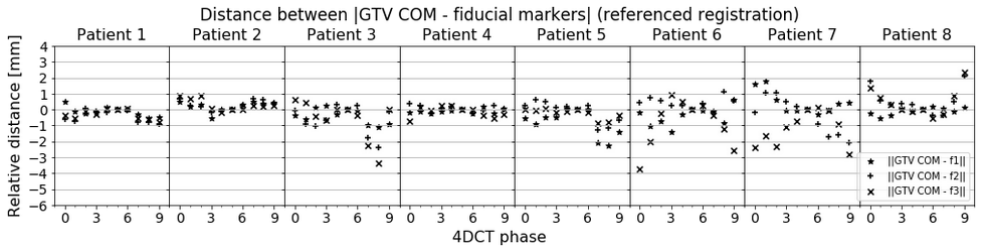
Figure 3.4: **D**istance metric (Eq (3.1)) of fiducial markers for all the patients computed with the reference (left) and the sequential registration (right). End expiration was chosen as reference of measurements, i.e. 50% phase. Top figures: 3.4a and 3.4b show the distance between the estimated and the measured fiducial markers in all phases for all patients. The estimated fiducial markers are the transformed markers from the reference phase to the corresponding 4DCT phase using the transformation found with registration. Bottom figure: 3.4c depicts the registration errors for the inpainted images. Only the reference registration approach is tested for the inpainted images.

Distance variations of Markers, GTV and Liver

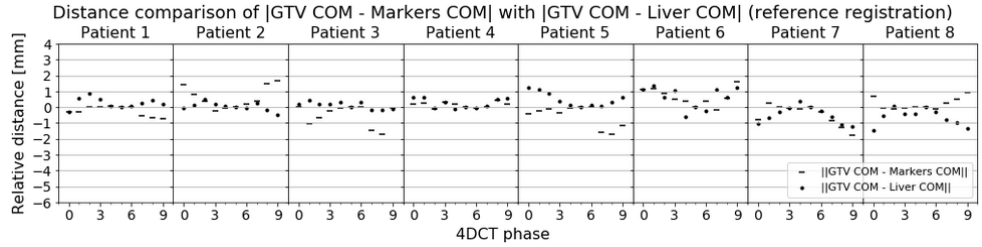
Fig 3.5 shows the distance variations of markers, GTV and liver for each patient. The corresponding distances were computed similar to Eq (3.2). For reference values see Table 5.1. Fig 3.5a depicts the euclidean distance between the markers for each phase. The distance between markers varied up to 6 mm.



(a)



(b)



(c)

Figure 3.5: **C**omparisons of markers distance stability. The distances within each phase are measured relatively to the base distance measured at reference phase (50% phase). The top panel 3.5a shows the distance between the markers. For some patients the marker to marker distance differ in the breathing cycle up to $6mm$. This clearly demonstrates that individual markers are problematic surrogates of the tumor in the liver. Middle panel 3.5b shows the distance between the GTV COM and each fiducial marker. Some of the markers change their distance with the tumor up to $4mm$. Bottom panel 3.5c depicts the relative distance of the GTV COM with the liver COM. GTV COM and liver COM are estimated with deformable image registration. Errors are up to $1.2mm$.

Fig 3.5b presents the relative distance variations of GTV COM and fiducial

markers per patient. Since the GTV COM is estimated with registration, this particular metric is related to the registration error. It can be seen that when a high distance variation between fiducial markers occur, it corresponds to a high variation of GTV COM and fiducial markers, and therefore, a high registration error. This is also observed in Fig 3.9 of Appendix F. Registration Errors per Patient. This pattern is noticed for patient 3, 6 and 7. As an example, if we analyze fiducial marker 3 of patient 6 at 0% phase, the GTV COM to fiducial marker distance is close to $4mm$. This error is justified in Fig 3.5a where there are distance variations of fiducial marker 3 with the other markers. We could confirm an uncertainty in the measurement of fiducial marker 3. Individual analysis could be made for the others fiducial markers. In order to avoid individual fiducial marker errors the markers COM is recommended.

For patient 8 a high registration error of marker 1 in phase 90% is related to the position of the marker close to the edge of the liver. For this case, image qualitative verification is a valid tool of assessment, as suggested in the AAPM TG-132 report on image registration [Brock et al., 2017].

Fig 3.5c shows the relative distance variations between GTV COM and markers COM, and between GTV COM and liver COM. The GTV COM - liver COM relative distance variation was computed with registration. GTV COM to markers COM distance variations were up to $2mm$. Assuming that the real position of the GTV was followed by the marker COM the accuracy of registration was as good as the resolution of the image. GTV COM to liver COM distance errors were lower than $1.2mm$. Both COMs were computed from registration and they followed the same motion pattern.

3.5 Discussion

We have demonstrated that the application of a publicly available SyN deformable image registration algorithm on a specific dataset, acquired for liver SBRT was suitable for liver and GTV localization. We investigated the accuracy of estimated versus measured fiducial markers' positions and shown that the mean error was less than $1.6mm$. This suggested that intensity-based DIR is accurate for liver radiotherapy ($< 2 - 3mm$ [Brock et al., 2017]).

We found no statistically significant differences between registering the images sequentially or to a reference. These results are in accordance to the results found by Boldea et. al. for lungs [Boldea et al., 2008]. In contrast to rigid motion problems where a drift is presented to sequentially registered images [Schonberger and Frahm, 2016], we found that DIR errors in liver 4DCT scans were of smaller magnitude than resolution errors.

It could be argued that the high intensity artifacts created by the fiducial markers influenced the convergence and accuracy of the registration algorithm. In order to investigate this hypothesis, we performed a test with the same algorithm on the same images where inpainting was applied to remove the

fiducial markers. Only the reference registration is tested. No statistically significant difference was found from the results obtained using the original images (without inpainting). This also suggest that the DIR is driven by the whole abdomen (liver) intensities values.

We expose the other capabilities of DIR applied to the liver. These are the estimation of liver and tumor volumes in the breathing cycle. The liver tissue is not highly compressible, but some deformation is expected [Marchesseau et al., 2017]. Another capability is contour propagation. Contour propagation with DIR have been evaluated before for lung and head and neck tumors [Hardcastle et al., 2013]. To the best of our knowledge, this is the first time that is evaluated for the liver. Although, this is not validated with overlapping metrics, the visual quality and the registration errors suggested succesful results.

We always evaluated the registration accuracy with features or points, i.e. fiducial markers. Other metrics such as image similarity or volume overlapping are less recommended in DIR [Rohlfing, 2011]. The registration errors always had at least one fiducial marker with an error lower than $2mm$ for each patient. Some of the maximum values in registration errors could be seen as outliers, and may be due to uncertainties in the measurement of individual fiducial markers.

Analysis of the individual results of each patient in Fig 3.5 revealed that the errors came from the fiducial markers positions that seemed unstable between the respiratory phases. These variations could be related to two factors, measuring errors or liver compression/deformation. These relatively high values demonstrated that a single fiducial marker cannot be a reliable surrogate for tumor positioning. One explanation of a measuring error could be related to an artefact, induced during 4DCT reconstruction due to an unstable respiratory pattern [Keall et al., 2006]. The other reason is that compression/deformation inherently change the distance between markers. This behavior may induce errors in the routine location of the tumor inside the liver which is dependent on the markers. We propose then the use of a more consistent reference for locating the GTV COM with the liver COM computed with registration. This reference shows a maximum relative error of 1.2 mm inter-phases. Therefore, the approach taken here could reduce potential errors.

Compared to some of the previous studies related to fiducial markers in liver radiotherapy [Beddar et al., 2007, Wunderink et al., 2008, Wunderink et al., 2010, Seppenwoolde et al., 2011, Park et al., 2012, Worm et al., 2016], to the best of the author's knowledge, this is the first study reporting variations in marker to marker distances between phases. This factor points out to additional induced errors with fiducial markers as a surrogate due to measurement errors or to liver deformations. Similarly to the previous studies, we also confirmed that the fiducial markers COM is a better surrogate of the liver compared to single fiducial markers. It is also expected that some uncertainty exists in the markers COM calculation. As suggested by Wunderink et al [Wunderink et al., 2010], the fiducial markers should be implanted near and

surrounding the tumor. In our study the fiducial markers were closely located between $17mm$ and $55mm$.

One of the limitations of our study is the small number of patients. Including more patients would strengthen the results, however for a proof of principle, the included number seems sufficient, since a wide range of tumour positions within the liver and respiratory motion amplitudes was represented in our patient material. Another drawback prior to a potential clinical implementation is current DIR computational time. This may be overcome in a near future with SyN algorithm implementations on GPU [Nielsen et al., 2019] or emerging deep learning DIR methods [Balakrishnan et al., 2019] where results are obtained in a couple of minutes or even seconds respectively.

3.6 Conclusions

We have demonstrated that intensity-based DIR is accurate for liver radiotherapy. Furthermore, we found no statistically significant differences between the images sequentially or referenced DIR methods. We propose a GTV localization strategy on 4DCT using DIR, that is consistent with GTV localization based on fiducial markers' COM with less variability than what is seen between individual markers. The DIR algorithm using the liver COM as a reference to locate the GTV, resulted in our small patient cohort in a maximum error of $1.2mm$. Such accuracy also seems adequate for radiotherapy.

3.7 Appendix

Appendix A. Patient Selection

The patients were selected within a breath-hold study of SBRT for metastases. The patients were treated at Rigshospitalet, Copenhagen, Denmark. The initial 21 patients were reduced to 8 patients based on fiducial markers implanted, image quality and regular breathing patterns. See details on patient selection in Figure 3.6.

As a part of the study protocol the patients were also imaged using CT in three deep inspiration breath-holds and three expiration breath-hold using the RPM system with visual feedback to obtain consistent breath-hold levels.

Appendix B. Registration Algorithm

Image registration is the process of estimating the transformation between two images in order to align them in a single coordinate system. The algorithm is called deformable image registration (DIR) when the images need to be transformed in a deformable manner, typically with a deformation field. This is the case of images with internal organs of the human body acquired during breathing motion.

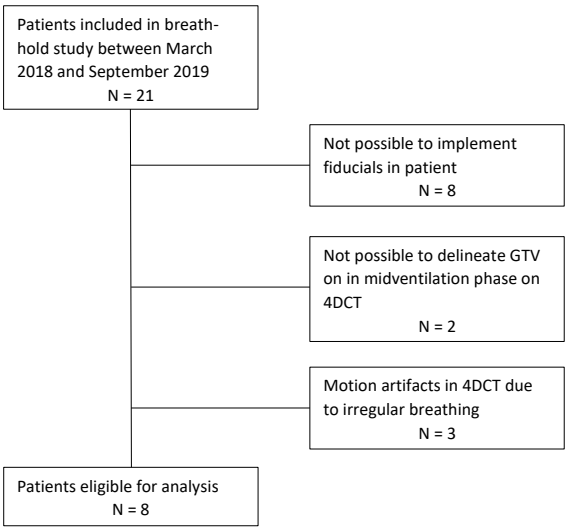


Figure 3.6: Flow chart of patient selection. A total of 8 patients with fiducial markers were selected to test the accuracy of registration to locate the tumor.

DIR is expressed as computing the diffeomorphic mapping $\varphi(x)$ of the moving image I_1 into the reference image I_0. The registration problem can be posed as a cost- or energy optimization problem involving a similarity metric $E_{\mathcal{M}}$ and a regularization term $E_{\mathcal{R}}$, such as:

$$\varphi^*(x) \equiv \arg\min_{\varphi(x)} \quad \underbrace{E_{\mathcal{M}}(I_0, I_1 \circ \varphi(x)) + E_{\mathcal{R}}(\varphi(x))}_{\equiv E(\varphi(x))} \tag{3.3}$$

Where:
$\varphi(x) : \Omega_0 \to \Omega_1\,,\ x \in \Omega,\ \Omega \subset \mathbb{R}^d,$
$I_j : \Omega_j \to \mathbb{R},\ \forall j \in \{0, 1\}.$

Avants et al. [Avants et al., 2008] introduce Symmetric Image Normalization (SyN) a DIR algorithm. The algorithm is available with ANTs toolbox [ANTs, 2021] and supported by ITK toolbox [ITK, 2021].

SyN algorithm uses cross correlation as the similarity metric $E_{\mathcal{M}}$ defined as:

$$E_{\mathcal{M}}(I_0, I_1 \circ \varphi(x)) = CC(I_0, I_1, \varphi(x)) \equiv \frac{\left(\sum_{x \epsilon \Omega_0}^{n} (I_0(x) - \bar{I}_0) \cdot (I_1 \circ \varphi(x) - \bar{I}_1) \right)^2}{\sum_{x \epsilon \Omega_0}^{n} (I_0(x) - \bar{I}_0)^2 \cdot \sum_{x \epsilon \Omega_0}^{n} (I_1 \circ \varphi(x) - \bar{I}_1)^2} \tag{3.4}$$

The regularization $E_{\mathcal{R}}$ is defined as the Sobolev norm of the velocity field v_t, similar to the term proposed by Beg et al. [Beg et al., 2005]. The velocity field is related to the deformation field $\varphi(x)$ with the ordinary differential equation $\frac{d\phi_t(x)}{dt} = v_t(\phi_t(x), t)$, with solution $\varphi(x) = \phi_0 + \int_0^1 v_t(\phi_t(x))dt$. In order to address consistency and reduce computation time, SyN algorithm divide the deformation field in two, such as: $\varphi(x) = \phi_1 \circ \phi_2^{-1}(x, t)$. Consequently, the regularization term involves two velocity fields. The mathematical setup is expressed as follows:

$$E_{\mathcal{R}}(\varphi(x)) = \int_0^1 \|v_t\|_L^2 \, dt = \int_{t=0}^{0.5} \left(\|v_1(x,t)\|_L^2 + \|v_2(x,t)\|_L^2 \right) dt \tag{3.5}$$

Where:
$L = a\nabla + bId$, is a linear differential operator that induces regularity.

The optimization in SyN is described by the minimization of the energy equation:

$$\varphi^*(x) \equiv \arg\min_{\varphi(x)} \underbrace{\int_{t=0}^{0.5} \left(\|v_1(x,t)\|_L^2 + \|v_2(x,t)\|_L^2 \right) dt + \int_\Omega CC(I_0, I_1, \varphi(x)) d\Omega}_{\equiv E(\varphi(x))} \tag{3.6}$$

Appendix C. Inpainting

The image inpainting algorithm proposed by Telea et. al. [Telea, 2004] is publicly available in OpenCV [OpenCV, 2021]. The fiducial marker centroid is used as the center point of the region where the inpainting is applied. A square area of 30 by 30 pixels is removed of the axial slices. The region is removed for a total of three slices around the marker. Figure 3.7 depicts the inpainting of a patient.

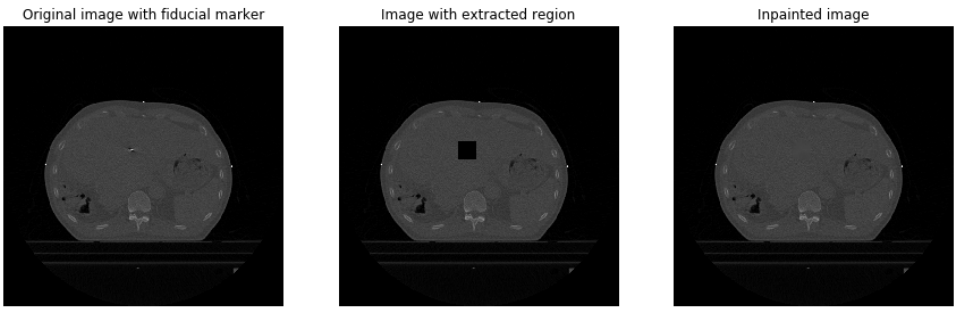


Figure 3.7: Example of image inpainting in one axial slice. The inpainted area is a square of 30 pixels around the fiducial marker centroid.

Appendix D. Registration Parametrization

The algorithm is setup in a multi-resolution framework. We computed the distance error per iteration to compare with the cross correlation (CC) metric. Figure 3.8 depicts the registration for one patient with a maximum of 300 iterations per level. In the multi-resolution scheme it was observed that the intensity metric could induce some overfitting behavior to the metric position of the markers. Furthermore, the initial scale down 1/8 worked as the initialization and most of the convergence occurred at the scale 1/4. The finals steps were a refinement process.

We varied the parameters of registration such as: iteration, multi-resolution levels, gradient step and radius neighbor for CC. We analyzed the convergence and chose gradient step as 0.1, radius 4 in cross correlation, 4 levels of resolution with scales [8,4,2,1] and iterations per level [120,100,90,60]. We observed that the algorithm converged properly for all the patients after parametrization.

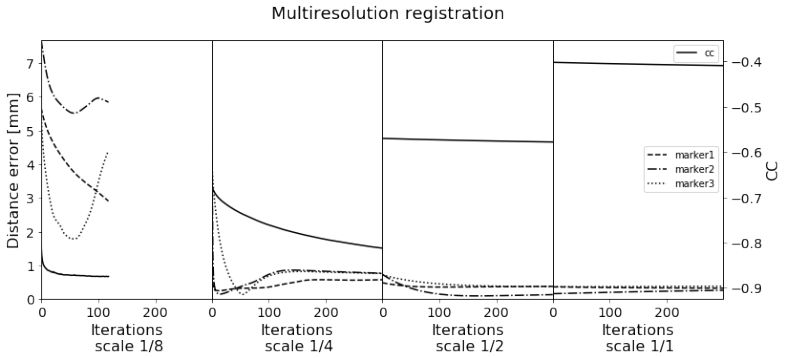


Figure 3.8: Comparing distance error of fiducial markers with cross-correlation similarity metric in multi-resolution registration. The voxel intensity metric achieves convergence for the distance metric.

Appendix E. Statistical Results

Table 3.2 shows the computed values of the Kolmogorov-Smirnov statistic on 2 samples. This is a two-sided test for the null hypothesis that 2 independent samples are drawn from the same continuous distribution. The test is applied to the distribution of registration errors for all the patients, comparing the results per phase obtained with the different methods. The compared methods are the reference versus sequential registration, and the inpainted versus the original images with reference registration.

	Phase	00	10	20	30	40	50	60	70	80	90
Reference vs sequential	Statistic	0.292	0.333	0.333	0.167	0.0	0.0	0.0	0.250	0.208	0.167
	P-value	0.263	0.140	0.140	0.902	1.0	1.0	1.0	0.449	0.686	0.902
Original vs inpainted	Statistic	0.167	0.333	0.375	0.167	0.250	0.0	0.208	0.125	0.250	0.167
	P-value	0.902	0.140	0.068	0.902	0.449	1.0	0.686	0.994	0.449	0.902

Table 3.2: Summary of Kolmogorov-Smirnov statistical tests. Comparisons are made for reference versus sequential registration methods, and the inpainted versus original images (reference registration). The p-values are higher than 0.05 which cannot reject the null hypothesis.

Appendix F. Registration Errors per Patient

The registration errors were comprised in a single plot for all patients. For expanding this information the following figures offer patient specifics of registration errors (Figure 3.9) and registration errors of inpainting (Figure 3.10).

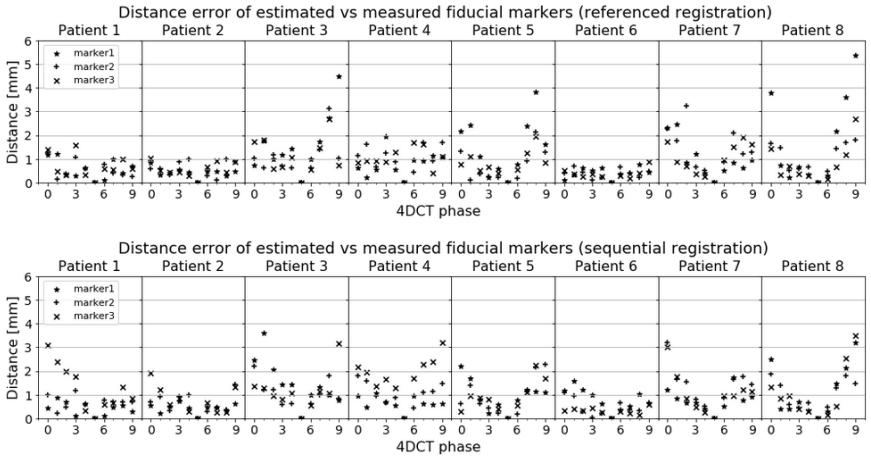


Figure 3.9: Registration errors per patient. Top panel relates to refence registration and botton panel relates to sequential registration.

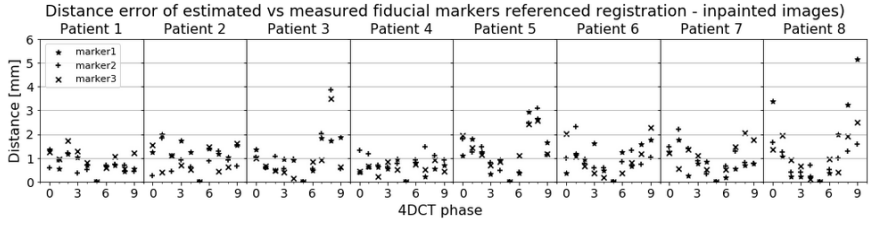


Figure 3.10: Registration errors with inpainting per patient. Only referenced registration is applied for the inpainted images

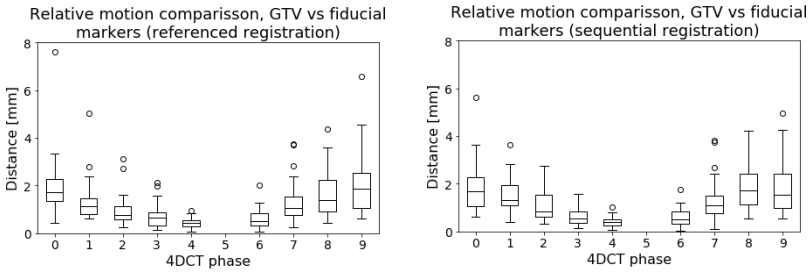


Figure 3.11: Relative motion discrepancy of GTV COM compared to the relative motion of the fiducial markers. The distance metric is comprised for all the patients computed with the reference (left plot) and the sequential registration (right plot). Relative motion refers to having as zero coordinates the positions of 50% phase.

Appendix G. Complementary Metric

In our study, we evaluated another metric to test registration errors. We exclude this from the paper for simplicity reasons. The complementary metric was defined as the relative motion discrepancy. The metric compare the GTV COM and the fiducial markers. The fiducial markers are the ground truth. The GTV COM position was calculated for every image based on the transformed GTV of the reference image. For both GTV COM and fiducial markers, the respective reference positions are substracted, i.e. the same zero coordinate point at the reference. After this, the relative motion discrepancy was then the distance between the GTV COM and the fiducial markers in the other respiratory phases of the 4DCT. This was defined as:

$$m \equiv \|(x_{f,j} - x_{f,\text{ref}}) - (\varphi_j \circ x_{\text{GTV},j} - x_{\text{GTV,ref}})\| \tag{3.7}$$

The relative motion discrepancy comprised results are shown in Figure 3.11. Individual patients results are shown in Figure 3.12. The same patern of high distance variations in patient 3, 6 and 7 where the fiducial markers have variations is observed. The relative motion discrepancy between GTV COM and fiducial markers had a maximum mean of $2.0mm$ and $1.9mm$ for the referenced registration and the sequential registration respectively. The maximum median values are $1.8mm$ for both reference and sequential registration. Similar to the registration error, no statistically significant differences between the sequential and referenced registration methods was found.

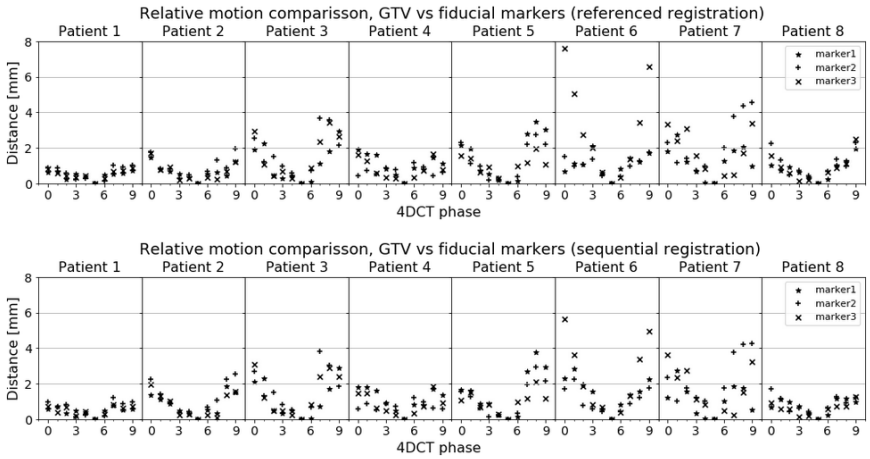


Figure 3.12: Relative motion discrepancy per patient of GTV COM to the relative motion of the fiducial markers. Top panel relates to refence registration and botton panel relates to sequential registration.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 764644. This paper only contains the author's views and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

Conflict of interest declaration

The authors LBS, MJ, SN and IRV received grants and/or educational fees from Varian Medical Systems.

Chapter 4

High Performance Registration

The work presented in this chapter is an original paper written for this thesis, that is to be submitted as: Tascón-Vidarte, J. D. (2022). IMART: IMAge Registration Toolkit. A high performance computing library for deformable image registration.

4.1 Abstract

Deformable image registration is computationally expensive and is generally considered an offline procedure. This paper presents IMART, a high-performance library of image processing and registration. Our desing supports various parallel hardware with OpenMP, OpenCL, and CUDA. We validated the framework execution on CPU and GPU. We focused on the transformation-interpolation operation to efficiently parallelize registration. The result is an improved registration algorithm capable of achieving real-time performance in 2d and reducing 3d computational times. We compare our library performance in terms of vectorial operations with ViennaCL and interpolation/registration with ITK. IMART achieves a lower or similar computational time to those optimized libraries.

Keywords: *Deformable Image Registration, High Performance Computing, Parallel Hardware, GPU*

4.2 Introduction

The most successful frameworks of deformable image registration are based on the Insight Segmentation and Registration Toolkit (ITK) [ITK, 2021]. Some examples are [ANTs, 2021], [Elastix, 2021] and [Plastimatch, 2021]. ITK is highly optimized for the CPU and supports multi-threading. However, there is still no full support for image processing on the GPU.

For more than a decade, GPU acceleration for DIR algorithms has been on the scene. Multiple surveys of medical image registration algorithms in parallel hardware confirm the performance gains and the parallelization capabilities [Shams et al., 2010, Fluck et al., 2011]. An early example is presented in [Gu et al., 2009], who offered a comparison of several demons algorithms on the GPU. More recent examples are [Ekström et al., 2021] and [Brunn et al., 2021] who present a novel library that support multi-node and memory distributed registration. We observe that none of them have reported a real-time deformable registration application.

We design a framework of high-performance computing for image registration. Our novel implementation allows defining the algorithms on the higher level while the device computation runs on the low level. This architecture enables device independence. We support multiple devices with OpenCL (CPU, GPU), multithreading on the CPU with OpenMP, and Nvidia GPUs with CUDA. To the best of our knowledge, our library implementation obtains the lowest computational time for a DIR algorithm in 2d ($< 4fps$). Thus, we offer a suitable algorithm for a real-time application.

4.3 Methods

Our goal is to design and implement a high-performance deformable image registration library. In order to achieve this, we formulate the registration problem, identify the computational bottleneck and propose a parallel algorithm. Furthermore, we intend to create a high-level and scalable framework to develop multiple intensity-based registration algorithms.

Mathematical Formulation of Registration

Having a fixed image $I_0(x)$ and a moving image $I_1(x)$ find the spatial transformation $\varphi(x)$ between them. Figure 4.1 shows the registration setup.

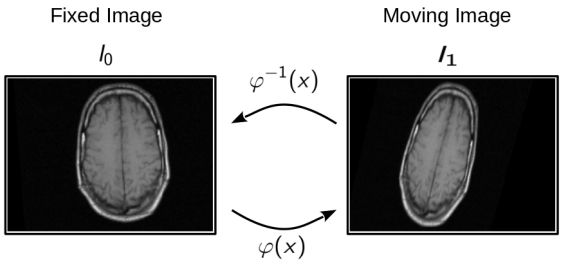


Figure 4.1: Deformable image registration

The registration problem can be formulated as an energy function as follows:

$$\arg\min_{\varphi(x)} E(I_0, I_1, \varphi) = E_{\mathcal{M}}(I_0, I_1 \circ \varphi) + E_{\mathcal{R}}(\varphi) \tag{4.1}$$

Where:

$$x_j \, \epsilon \, \Omega_j, \; \Omega_j \subset \mathbb{R}^d, \; j = 0, 1$$
$$I_j : \Omega_j \to \mathbb{R}, \; \varphi(x) : \Omega_0 \to \Omega_1$$

The cost function of equation 7.2 represents an optimization problem and can be solved with an iterative gradient-based method in the form:

$$\varphi(x)_{k+1} = \varphi(x)_k - \lambda \cdot A \cdot \nabla E \tag{4.2}$$

Where A represents a high order approximation, if A is the identity matrix, the equation is equivalent to a gradient-descent algorithm. The algorithm requires to compute the cost function and the update derivative on each iteration. The commonly repeated operation for both equations is the transformation-interpolation. In fact, the computational bottleneck of registration is not the optimization steps but the computation of the transformation-interpolation and the similarity metric, as remarked by [Shams et al., 2010]. We primary focus on this stage to improve performance.

Transform-Interpolation

The most straightforward deformable transformation is a displacement vector field. The field offers a vector translation for each voxel coordinate in the image domain. The advantage is the parallelization capabilities of this transformation. The disadvantage is the number of parameters required, i.e., $dimension \times voxels$. The deformation field transformation is defined as:

$$\varphi(x) = x + u(x) \tag{4.3}$$

In the case of having the fixed and moving images with identical world coordinates, the registration becomes a vectorial sum operation. Otherwise, interpolation is required to estimate the deformation vector at locations outside the grid. Therefore, we explicitly develop two transformation functions, one highly vectorized for images with the same coordinates and another parallelized supporting interpolation.

Interpolation is the process of estimating new data points from a set of discrete values. The discrete nature of images usually requires an interpolation operation to estimate the intensity value after a transformation. Its functional representation is given by:

$$I(\varphi(x)) \equiv I \circ \varphi(x) \tag{4.4}$$

The interpolation function offers intensity values outside the image grid. For a comprehensive survey covering the theory and performance of interpolation for medical image application, see [Thévenaz et al., 2000]. Multiple methods of interpolation exist. However, the commonly used methods are nearest-neighbor, linear, cubic, b-splines. Linear interpolation is the most widely used method in practice, and thus we focus on this algorithm.

A one dimensional linear interpolation is estimated as:

$$I(x) = I(x_a) + (I(x_b) - I(x_a)) \cdot \frac{x - x_a}{x_b - x_a} \tag{4.5}$$

Where x is any coordinate point located inside $[x_a, x_b]$. For discrete or structured data (e.g. images), the unit distance $x_b - x_a = 1$ and therefore we can simplify Equation 4.5 as:

$$I(x) = I(x_a) \cdot (1 - \dot{x}) + I(x_a + 1) \cdot \dot{x} \tag{4.6}$$

Where $\dot{x} = x - x_a$ and x_a is the nearest discrete value. For images we use the floor operator $x_a = floor(x)$. Similarly, we can express the linear interpolation of a 2 dimensional image as:

$$\begin{aligned} I(x, y) = I(x_a, y_a) \cdot (1 - \dot{x} - \dot{y} + \dot{x}\dot{y}) + I(x_a + 1, y_a) \cdot (\dot{x} - \dot{x}\dot{y}) \\ + I(x_a, y_a + 1) \cdot (\dot{y} - \dot{x}\dot{y}) + I(x_a + 1, y_a + 1) \cdot \dot{x}\dot{y} \end{aligned} \tag{4.7}$$

The Equation 4.7 is usually called bi-linear interpolation. The mathematical extension to 3-dimensional images is straightforward. For simplicity, we only refer to the 2-dimensional case.

We define a single kernel for interpolation in order to parallelize interpolation. During 2d interpolation, there is a probability to have one coordinate or both close to the discrete values. For these cases, the interpolation computation using Equation 4.6 is more efficient than using the full expression of Equation 4.7. Our interpolation kernel takes advantage of this situation to reduce some computational burden. Algorithm 1 summarize the kernel. The kernel function is executed during each interpolation request.

Algorithm 1 Linear Interpolation Kernel

Input: Reference Image $I_r(x_a, y_a)$
Parameters: Image coordinates (x, y), $\varepsilon \to 0$
Output: Interpolated Image $I_o(x, y)$

Find $x_a = floor(x)$, $y_a = floor(y)$
if (x_a, y_a) inside Ω_r **then**
  Find $\dot{x} = x - x_a$, $\dot{y} = y - y_a$
  if $(\dot{x} < \varepsilon)$ & $(\dot{y} < \varepsilon)$ **then**
    Compute $I_o(x, y) = I_r(x_a, y_a)$
  else if $\dot{x} < \varepsilon$ **then**
    Compute $I_o(x, 0)$, using Eq. 4.6
  else if $\dot{y} < \varepsilon$ **then**
    Compute $I_o(0, y)$, using Eq. 4.6
  else
    Compute $I_o(x, y)$, using Eq. 4.7
  end if
end if

We implement the kernel function in C++ with OpenMP, OpenCL, and CUDA. We consider the device architecture and the memory model for GPUs with OpenCL and CUDA. Figure 4.2 shows the OpenCL memory model as reference. The image is partitioned and transferred to the local memory to avoid a memory-bound algorithm. Afterward, the kernel is applied. We replicate this procedure locally to achieve better performance. Finally, we integrate all these principles into our IMART library.

IMART Design

We designed and implemented a general image processing and registration library that supports parallel hardware. The library is an object-oriented framework written in C++ (std=c++17) as header-only. This simplifies the building process, template creation, and portability. The library uses ITK as a dependency to support multiple formats of image reading and writing (formats: jpg, png, dcm, nii, nrrd, among others). We implement real-time visualizations during registration with VTK. Other dependencies are boost::options,

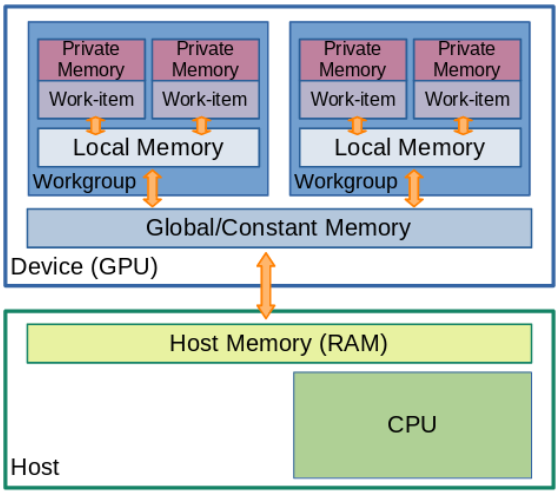


Figure 4.2: Memory model

fftw3, and clfft. The library has automatic testing with google/test (gtest) and benchmark with google/benchmark (benchmark). The software development and version control are supported by CMake and Git, respectively. Furthermore, there is integration with interface functions for other computer vision libraries such as ITK and OpenCV. Finally, the interoperability is available with Docker, and the documentation is built with Doxygen.

The overall design of IMART is presented in Figure 4.3. The imaging classes are templates with intuitive names, short definitions, and brief descriptions. IMART has three main classes: *data_object*, *space_object* and *process_object*. All other classes below the main classes take advantage of inheritance and polymorphism. Regarding image processing, IMART has a core class named "*image*". This class incorporates functions related to initialization (zeros, ones, random), overloaded arithmetic $(+, -, *, /, \& scalar)$, reductions (min, max, sum, prod, dot), and image processing utilities (pad, unpad, cast, normalize, fft, ifft, gradient). Regarding registration, IMART includes transformations (Rigid, Affine, Deformation Vector Field), interpolations (Nearest, Linear, Cubic), similarity metrics (SSD, CC), and optimizer (Gradient Descent).

In terms of parallel hardware, the library supports multiple devices, such as: CPU (OpenMP) and GPU (OpenCL, CUDA). Our design is inspired on ViennaCL [Rupp et al., 2016, ViennaCL, 2021]. The image class requires a container class to store the data and to operate with an specific device. The available containers are: *vector_cpu* $<$ *type* $>$, *vector_opencl* $<$ *type* $>$, *vector_cuda* $<$ *type* $>$. The functions and operators of the image class are overloaded to support device independence on the high level. An example is

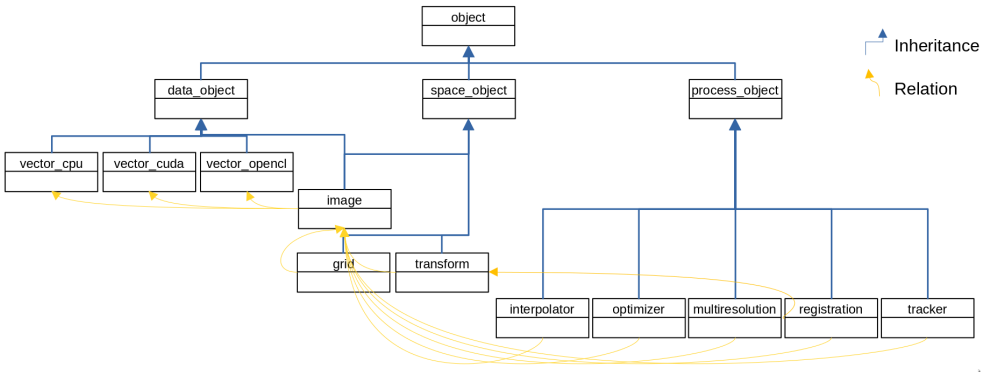


Figure 4.3: IMART design. Inheritance and class relations diagram

shown in Algorithm 2. The code example works on CPU because the image is defined in line 14 with the cpu container "$vector\_cpu < type >$". To change the whole operation to GPU we only modify the same line 14 with the desired container, for instance "$auto\ img = image < type, vector\_opencl < type >>::new\_pointer()$". The result is a highly versatile library with a straightforward way to change the processing hardware.

Implemented Algorithms

We incorporated in IMART two well-established deformable image registration algorithms, the diffeomorphic demons and large deformations diffeomorphic metric mapping (LDDMM). The algorithms are described below.

Demons Algorithm

The demons algorithm was introduced by Thirion [Thirion, 1998]. The algorithm treats registration as a diffusion process that is similar to an optical flow approach. The diffeomorphic demons algorithm is an extension of the original algorithm with a constraint to the transformation to be diffeomorphic [Vercauteren et al., 2009]. We choosed to implement the demons algorithm because it has proven to be versatile and very efficient to perform deformable image registration [Gu et al., 2009].

The demons algorithm can be considered as an approximation of a second order gradient descent on the sum of square of intensity differences (similarity metric) criterion [Pennec et al., 1999] with Gaussian regularization. The diffusion equation is defined as:

$$F(\nabla E(\varphi_x)) \equiv \frac{(I_0 - I_1(\varphi_x))\nabla I_1(\varphi_x)}{\|\nabla I_1(\varphi_x)\|^2 + (I_0 - I_1(\varphi_x))^2} \tag{4.8}$$

The iterative process to update the transformation $\varphi(x)$ is described as:

Algorithm 2 IMART C++ code example on CPU

```
// std libs
#include <iostream>

// imart libs
#include "imart/image.h"
#include "imart/image_utils.h"

using namespace imart;

int main()
{
    using type = unsigned char;

    auto img = image<type, vector_cpu<type>>::new_pointer();
    img->read("input.png");
    img->print("input");

    *img = *img + 10;
    img = pad(img, std::vector<int>{8,4}, std::vector<int>{8,4});

    img->write("output.png");
    img->print("output");

    return 0;
};
\label{}
```

$$\begin{aligned} compute: &\quad u(x) = F(\nabla E(\varphi_x)) \\ fluid\,update: &\quad u(x) \leftarrow G^{\sigma_f} * u(x) \\ diffeomorphic: &\quad \varphi(x)' \leftarrow exp(u(x)) \\ diffusion\,update: &\quad \varphi(x) \leftarrow G^{\sigma_d} * \varphi(x)' \end{aligned} \tag{4.9}$$

The expression $exp()$ refers to the intrinsic update on the Lie group of diffeomorphisms, and G refers to a Gaussian kernel. The kernels can be applied to the deformation update $u(x)$ or the overall transformation $\varphi(x)$ and behave as an approximation of fluid or diffusion regularization, respectively.

LDDMM Algorithm

Beg et al. introduced the LDDMM algorithm [Beg et al., 2005]. The algorithm has proven to be an effective approach to enforce the diffeomorphic constraint while allowing for large deformations. The result is a symmetric and accurate algorithm but computationally expensive.

The LDDMM cost function is defined as:

$$\underset{v_t:\dot{\varphi}_t=v_t(\varphi_t)}{\arg\min} E(I_0, I_1, v_t, \varphi) = \frac{1}{\sigma^2}||I_0 \circ \varphi_1^{-1} - I_1||_{L^2}^2 + \int_0^1 ||v_t||_L^2 dt \tag{4.10}$$

Where $L = \alpha\nabla + \gamma$ is the Cauchy-Navier operator that induces regularity. The cost function gradient is defined as:

$$\nabla E(I_0, I_1, v_t, \varphi) = 2v_t + \frac{2}{\sigma^2} K\left(|Det\varphi_{t,1}|\nabla J_t^0 (J_t^0 - J_t^1)\right) \tag{4.11}$$

Where $J_t^0 = I_0 \circ \varphi_{t,0}$, $J_t^1 = I_1 \circ \varphi_{t,1}$, $K(g) = (L^\dagger L)^{-1} g$ is the regularization function. The algorithm is solved with a gradient descent approach to solve the velocity field and the deformation field is found by integration. The algorithm steps are summarized as:

$$\begin{aligned} update: &\quad v_t(x) = v_t(x) + \lambda h \\ integrate\ backward: &\quad v_t \to \varphi_{t,1} \\ integrate\ forward: &\quad v_t \to \varphi_{t,0} \\ compute: &\quad h = \nabla E(I_0, I_1, v_t, \varphi) \end{aligned} \tag{4.12}$$

Performance Evaluation

We developed performance tests of our IMART library concerning memory transfer (CPU to GPU and vice versa), vectorial operations, linear interpolation, and deformable image registration. We implemented all the tests using google benchmark [Google/Benchmark, 2021]. We ran each test 11 times and reported the median values. Furthermore, we included a device and platform comparison for each test, i.e., CPU (OpenMP) and GPU (OpenCL, CUDA). Additionally, we contrasted our library to other optimized libraries. We compared the performance of vectorial operations (BLAS level 1) with ViennaCL [Rupp et al., 2016, ViennaCL, 2021]. We compared the computational time of IMART versus ITK ([ITK, 2021]) in regards to interpolation and deformable image registration. The vectors and the images were initialized with random values for all the tests, except for deformable image registration, where we tested the algorithm intrapatient with ten pairs of 2-dimensional images of size 256×256. The images belong to the DIR-LAB dataset [Castillo et al., 2009b] where we extracted two random slices per patient. We only performed the registration tests in 2d because we aim for real-time applications.

Performance Metric

Due to different hardware and software implementations an unified way to compare registration algorithms is using the performance metric proposed by Shams et al [Shams et al., 2010]. The metric is defined as:

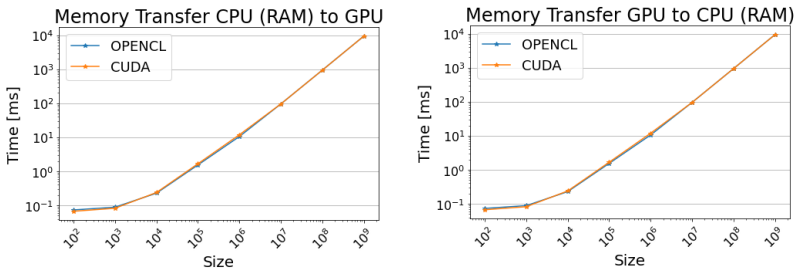
$$P = \frac{milliseconds}{MVoxel \cdot iteration} \tag{4.13}$$

We use the metric accompanied by the total time to compare the deformable image registration algorithms. The algorithms used a multiresolution framework with 3 levels. The performance metric is only computed for the final resolution (images at full scale)

4.4 Results

All the tests are run on a workstation with two CPUs, one GPU, and 128 GB of RAM. Each CPU is an Intel(R) Xeon(R) Silver 4110 @ 2.10GHz, 8 cores, 16 threads. The GPU is a Nvidia(R) GeForce(R) RTX 3090.

The algorithms executed on GPU require a memory transfer operation from RAM. This operation is made when the algorithm starts and finishs. We summarize on Figure 4.4 the transfer time depending on the memory size (each element is 4 Bytes, float) for OpenCL and CUDA. There are no significant differences between OpenCL and CUDA, since the maximum difference is around 0.02ms. Similarly, the time taken from CPU to GPU and from GPU to CPU is virtually the same. For instance, a 2 dimensional image of size 256×256 requires 1.04 ms to transfer, and a 3 dimensional image of size $256 \times 256 \times 256$ requires around 150 ms to transfer. The advange of our library desing is that for an expensive registration algorithm the memory is only transfered once, when the image is read.

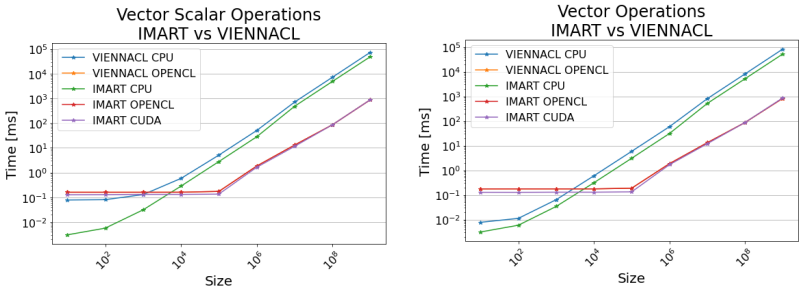


(a) (b)

Figure 4.4: Memory transfer time versus element size, from CPU to GPU (4.4a) and from GPU to CPU (4.4b).

Figure 4.5 shows the computational time of vectorial operations of our library compared to ViennCL. Figure 4.5a involves a vector scalar multiplication, i.e. $v_{out} = s \cdot v_{in}$. The operation in Figure 4.5b is a vector addition, $v_{out} = v_1 + v_2$. We observe a converging performance improvement of 1.5 times in our CPU implementation compared to ViennaCL. The performance for OpenCL between IMART and ViennaCL is virtually the same. CUDA

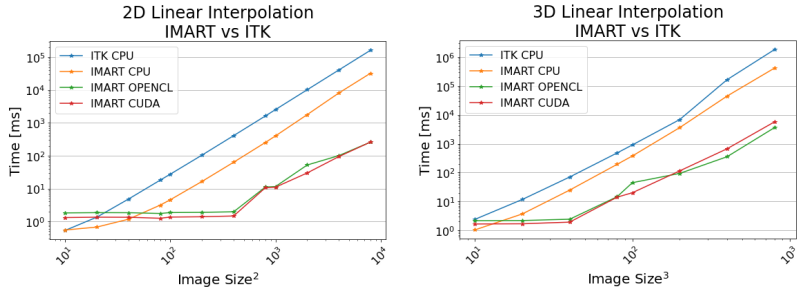
seems to perform slightly better against OpenCL for vector operation with size lower than 10^5. There is a kernel launching cost for the GPU that generate a threshold value of performance where the GPU surpass the CPU. This occurs in the operations with vector size near 10^4.



(a) (b)

Figure 4.5: Performance comparison of vector operations between IMART and VIENNACL. 4.5a corresponds to a vector scalar operation $v_{out} = s \cdot v_{in}$ and 4.5a corresponds to a vector addition $v_{out} = v_1 + v_2$.

Figure 4.6 depict the interpolation comparison between ITK and IMART in 2d (Figure 4.6a) and in 3d (Figure 4.6b). The CPU interpolation in 2d is improved by a factor 5, while the 3d interpolation by a factor of 4.3. The computational times of OpenCL and CUDA are close with a slighly better performance for CUDA. The kernel launching cost of the GPU in 2d is improved when the interpolation contain an image of 100×100, while in 3d for an image $20 \times 20 \times 20$. The GPU reduces the computational time of interpolation converging to 72.8 times compared to the CPU in our implementation.

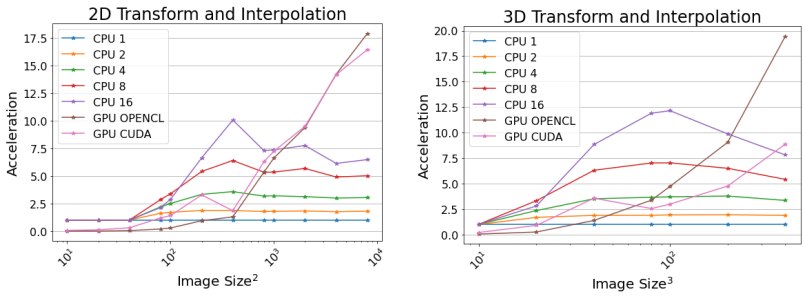


(a) (b)

Figure 4.6: Computational time of linear interpolation using IMART versus ITK.

We compare the performance of multicore CPU versus GPU for a transformation-interpolation using IMART. The transformation is a deformation vector field and the interpolation is linear. Figure 4.7 summarize the acceleration factor

given by the number of CPU threads or the GPU. We observe an increase in acceleration as the multithread on the CPU increase. We recognize again the kernel laaunching cost of the GPU, that surpass the CPU 16 (threads) performance in 2d for images with size 1000×1000 and in 3d for images with size $200 \times 200 \times 200$. This suggest that the used GPU is more efficient for large image processing applications.



(a) (b)

Figure 4.7: Acceleration results obtained with IMART for CPU and GPU executing a transformation and a linear interpolation. The number next to the CPU legend represent the threads used.

We summarize the 2d registration results in Table 4.1. We compare our demons algorithm implemented in IMART with a demons algorithm implemented in ITK. Furthermore, we collect the performance results of our LDDMM implementation. We only compare LDDMM within IMART and only with respect to the device.

We observe in the worst case 6.24 times of improvement in computational time for the CPU of our demons algorithm compared to ITK. Again, the GPU obtains good performance but the kernel launch cost makes the CPU more suitable for small image sizes. Nevertheless, the GPU improves the performance for a complex algorithm such as LDDMM.

Algorithm	Metric	CPU					GPU	
		CPU 1	CPU 2	CPU 4	CPU 8	CPU 16	OpenCL	CUDA
Demons ITK	Time [ms]	5281	2966	1807	1341	**1012**		
	Performance	1506	821	467	346	**258**		
Demons	Time [ms]	355	240	204	**159**	162	290	265
	Performance	130	65	39	23	**19**	35	32
LDDMM	Time [s]	32.41	20.55	16.18	15.4	15.1	8.9	**8.2**
	Performance	21100	11290	8350	7720	7010	5050	**4720**

Table 4.1: Registration performance tests in 2d. Our demons algorithm is compared to ITK. The implemented LDDMM algorithm is compared only with respect to the device used. The performance refers to the metric shown in Equation 4.13. In bold the best device performance for each algorithm.

4.5 Discussion and Conclusion

Our library design represents a methodology to enhance performance in deformable image registration. Our strategy of improving the transformation-interpolation offered high-performance and satisfactory results compared to ITK, the core framework of various open-source registration libraries. Therefore, we deliver an attractive implementation for further applications.

The fact that IMART supports multiple devices allows us to compare the CPU and the GPU performance. The kernel launching cost for the GPU seen in the vector operations means that the CPU performs better with small data sizes. Since deformable image registration has a better convergence in a multiresolution framework [Thevenaz et al., 1998], the computational advantage of the GPU may be limited for 2-dimensional applications of registration. We prove this statement in the registration results. Nevertheless, the GPU is an outstanding device for larger data sizes and outperforms the CPU for 3-dimensional applications.

We designed a high-performance library for medical image registration. The library supports image analysis operations, filtering, and deformable image registration on CPU and GPU. We focused on the transformation-interpolation operation, which is highly repeated during each optimization step in registration. Furthermore, we implemented demons and LDDMM strategies to prove the library's versatility. The high-performing demons registration algorithm is suitable for a 2-dimensional real-time application. The code is open source and publicly available at `https://github.com/josetascon/imart`.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 764644. This paper only contains the author's views and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

Chapter 5

Cine-MRI Simulation

The work presented in this chapter is based on the paper: Tascón-Vidarte, J. D., Wahlstedt, I., Jomier, J., Erleben, K., Vogelius, I. R., & Darkner, S. (2021). Cine-MRI Simulation to Evaluate Tumor Tracking. In International Workshop on Simulation and Synthesis in Medical Imaging (pp. 131-141). Springer, Cham.

5.1 Abstract

Conventional evaluations of tumor tracking algorithms require inter-observer segmentations from radiation oncologists on the Cine-MRI (2D sagittal MR video). Instead of performing intensive manual annotations on images, we present a 2D video simulator that uses the pre-treatment images, including a breathing model, that generates Cine-MR images in parallel with the underlined segmentation of the tumor. We include the data of seven patients within a retrospective clinical study that received stereotactic body radiation therapy for liver metastases. Each patient has a pre-treatment 4DCT scan, a pre-treatment 3D MR with tumor and liver delineations, and the treatment Cine-MRI. We augment the data with the simulator by changing breathing motion parameters and adding noise. The simulator generates a total of 84 Cine-MRI sequences, thus having 12 videos per patient. We validate the simulated versus the real Cine-MRI in terms of tumor motion. Finally, we used the simulator to evaluate the performance of real-time tumor tracking algorithms with this dataset.
Keywords: *Cine-MRI, Simulation, Tumor Tracking, Real-Time, Image-guided Radiotherapy*

5.2 Introduction

An MR-linac is a device that combines magnetic resonance imaging with a linear accelerator. Stereotactic body radiotherapy (SBRT) of liver metastases on the MR-Linac system is advantageous due to the improved soft-tissue contrast compared to cone-beam computed tomography [Kontaxis et al., 2017]. In addition, the MR-Linac has beam gating, i.e. the treatment accelerator beam is triggered in response to patient movement [Crijns et al., 2011]. The clinical advantages of using an MR-linac with online tracking have been demonstrated [Al-Ward et al., 2018], indicating the potential to reduce the liver target volume and lower the radiation dose to adjacent organs at risk.

Liver tumors deform and move during treatment mainly caused by breathing motion [Murphy, 2004]. During treatment, the scanner acquires 2D cine-MR sagittal images at four (4) frames per second [Paganelli et al., 2018]. Tumor tracking is one of the main components of beam gating in the MR-Linac system. Tumor tracking is solved automatically using image analysis. Some proposed strategies for tumor tracking are based on template matching [Cervino et al., 2011, Shi et al., 2014], feature detection [Paganelli et al., 2015], optical-flow methods [Zachiu et al., 2015, Seregni et al., 2018], deformable image registration [Fast et al., 2017], segmentation [Gou et al., 2014], neural networks [Yun et al., 2015] or modeling based [Garau et al., 2019]. The difficulty in evaluating tumor tracking is the need for ground truth data as no publicly available datasets that comprise tumor tracking on Cine-MRI exist.

This requires manually delineation of the tumor in the entire video by a radiation oncologist [Cervino et al., 2011,Shi et al., 2014,Yun et al., 2015,Fast et al., 2017]. Evaluations are therefore only comparable at the institutional level. A generalized methodology to easily evaluate tumor tracking is required.

In addition, current tracking systems used in clinical practice may fail to track unexpected movements and have difficulty in tracking motion in the out-of-plane direction [Paganelli et al., 2018]. The breath-hold treatment is the most used respiratory motion management in practice [Feldman et al., 2019]. After breath-hold, the patients can exhibit a fast motion, and thus tracking becomes very difficult. All the previously reported studies evaluate tumor tracking under free-breathing conditions [Cervino et al., 2011,Shi et al., 2014, Paganelli et al., 2015, Yun et al., 2015, Zachiu et al., 2015, Fast et al., 2017,Seregni et al., 2018,Garau et al., 2019]. Thus, there is a need for improved tracking algorithms under varying breathing motion.

Our contribution is a straightforward evaluation methodology to quantify tumor tracking performance without the need of manual segmentations. The method is patient-specific and simple to implement. We demonstrate the capabilities of our method to create multiple simulated Cine-MRI, and to evaluate tumor tracking algorithms under varying conditions.

5.3 Related Work

Respiratory motion modeling is an extensively studied field [McClelland et al., 2013]. Deformable image registration generates the most suitable models [Stemkens et al., 2016]. Likewise, our breathing model works with deformable image registration. Fu et al. [Fu et al., 2019] use known deformation fields to create ground truth images and landmarks to validate feature detection on 4DCT. Our simulator works in a similar way, but in contrast the known transformation model is applied to the images and the tumor contours.

The use of pre-treatment 4D imaging data has been exploited for treatment. Harris et al. [Harris et al., 2018] use 4DMRI to create synthetic 4DCT during treatment on conventional linear accelerators. Garau et al. [Garau et al., 2019] use pre-treatment 4DCT with treatment Cine-MRI to estimate a 3DCT and compare planning versus treatment. The mentioned methods suggest multi-modal integration of images and that pre-treatment time-sequence images are valid to model breathing motion during treatment. We aim for a similar goal with a distinction, to build a breathing model based on registration, use the model to simulate treatment sequences, in order to improve tumor tracking algorithms. To our knowledge this is the first Cine-MRI simulation based on a respiratory motion model.

An alternative option to manually delineate the tumor is to use matched landmarks [Paganelli et al., 2015, Seregni et al., 2018]. Although this process can be automated with feature detection algorithms some outliers occur.

Furthermore, the landmark distance alone does not represent how well the tracking algorithm performs with respect to the tumor structure and the contours. Most tumor tracking algorithms have been tested on lung patients with manual delineations of the tumor [Cervino et al., 2011, Shi et al., 2014, Yun et al., 2015, Fast et al., 2017] and only a small set has been evaluated on liver patients with landmarks [Paganelli et al., 2015, Seregni et al., 2018]. We test the proposed methodology and the tracking algorithms with liver patients. Identifying a tumor contour on a sagittal slice is difficult in the liver due to the lack of contrast. Hence, the liver is a remarkable example of why manual tumor delineations for evaluating tumor tracking are not always feasible.

5.4 Methods

Data

This study uses image data from seven patients already treated with SBRT for metastases in the liver at Rigshospitalet (Copenhagen, Denmark) between April and December 2019. The patients provided informed consent and approval for the usage of their anonymized data for research purposes.

Respiratory correlated 4DCT with intravenous contrast injection was performed for all patients on a SOMATOM Definition AS scanner (Siemens Healthineers, Germany). 4DCT image data were phase-sorted into ten phase bins throughout a respiratory cycle based on an external respiratory signal monitored with Real-Time Position Management (RPM, Varian Medical Systems, USA). The slice separation in each phase of the 4DCT was 2 mm. The image resolution in each slice was 512×512 pixels and a pixel size of 0.98×0.98 mm.

A 0.3T MRIDIAN MR-Linac (ViewRay, USA) is used to acquired a pre-treatment 3D MRI and the Cine-MRI sequences. The 3D MR scans were performed for all patients in inspiration breath-hold position without visual guidance. The acquisition technique is balanced steady-state free precession (bSSFP). The pre-treatment image resolution is [$512 \times 512 \times 128$ pixels] and [$1.5 \times 1.5 \times 3.0$ mm] spacing. The Cine-MRI sequences (bSSFP-Sagittal) have a resolution of [256×256 pixels] and [1.5×1.5 mm] spacing. The clinical gross tumor volume was delineated on the 3D MR by a senior radiologist and approved by a senior oncologist. Organs at risk, including the liver, were also delineated. These delineated contours are used to segment the region of interest (ROI).

Cine-MRI Simulation

We developed a patient-specific Cine-MRI simulator capable to generate a simulated ground truth contour of the desired organs using pre-treatment images. The input images are a 4DCT scan and an MR with organ contours.

The video simulator has the following input parameters: video time, frames per second, breathing cycle time, breathing amplitude, and additive noise. Figure 5.1 illustrates the video simulation process. Algorithm 3 details the simulator pseudo code. The simulator has two stages: breathing modeling and video synthesis. Each stage is described in the following.

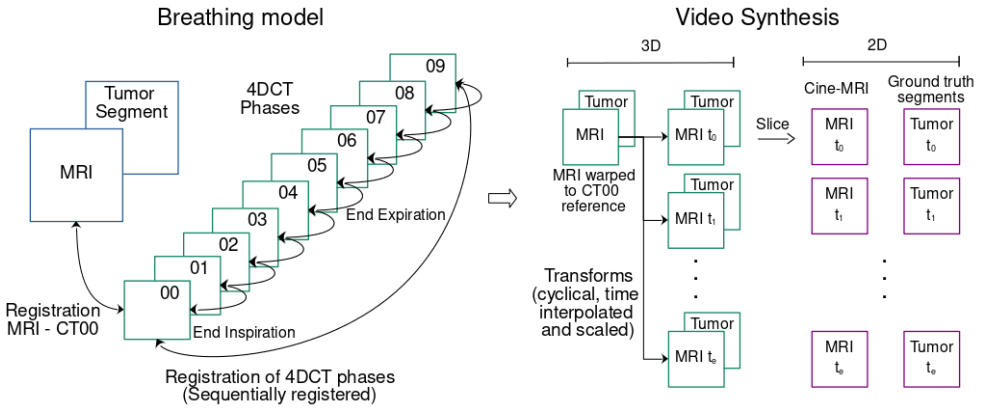


Figure 5.1: 2D Cine-MRI Simulator. The process comprises two stages: breathing modeling and video synthesis. The breathing model is a pre-processing stage and uses full 3D information to consider out-of-plane motion in the 2D Cine-MRI. The video synthesis stage can be run several times changing the simulation parameters to create different variants and motion conditions.

The breathing modeling is a pre-processing stage. It is computed once and stored in order to create several videos. This model is based on the 4DCT scan that represents the full respiratory cycle of each patient. Initially, all phases in the 4DCT are registered sequentially with the symmetric normalization algorithm [Avants et al., 2008]. Subsequently, the MR image is registered to phase 00 of the 4DCT scan, since both images are at inspiration position. This transformed MR is the starting video frame.

The video synthesis stage is an iterative stage. The simulator produces a new video frame as a composition of sequential transformations related to the 4DCT. The corresponding transformation is interpolated over time to match the proportion of the respiratory cycle with the video sample time. The video is created based on 3D images and transformations. From this a 2D slice is extracted in the sagittal view where the tumor has better visibility. Thus the simulated video has the same complexity as real 3D motion in the 2D images and simulates the MR-Linac imaging setup, where 2D real-time images are acquired and tracked. Since the initial contour of the tumor and organs are known in the reference MR, we create independent files with those regions of interest (ROI) and the ROI are transformed in parallel with the raw video image generation. For simplicity we only generate the tumor contours.

Algorithm 3 Cine-MRI Simulation

Input: 4DCT, 3D MRI (bSSFP) and clinical tumor contour
Parameters: Video time, frames per second, breathing cycle time, breathing amplitude, and additive noise
Output: Cine-MRI (Video containing 2D bSSFP-Sagittal images)

Register sequentially 4DCT images, and store transformations $\varphi_{00\to01}, \varphi_{01\to02}, ..., \varphi_{09\to00}$ where $j = 00, 01, ..., 09$ correspond to the phase;
Register MRI image to 4DCT phase 00 obtain $\varphi_{MR\to00}$;
Transform MRI image to CT phase 00 using $\varphi_{MR\to00}$, obtain M_{00};
Transform tumor delineation X_{MR} using $\varphi_{MR\to00}$ to get contour X_{00};
Compute simulation *time* using parameters: video time, frames per second;
for t in *time* **do**
    Compute time point t location in breathing cycle;
    Calculate time proportion for time interpolation t_δ;
    Find the shortest path of sequential transformations;
    Compose a transformation $\varphi_c(x)$ with the shortest path;
    Multiply $\varphi_c(x)$ with breathing amplitude parameter α;
    Transform image M_{00} using $\varphi_c(x)$, obtain M_t;
    Extract a slice of M_t to create the image $M_{2d,t}$;
    Add noise to image $M_{2d,t}$;
    Transform contour X_{00} using $\varphi_c(x)$, obtain X_t;
    Extract a slice of X_t to create the image $X_{2d,t}$;
end for

Therefore, we have the ground truth ROI for each video sample.

The simulator supports noise with two different probability distributions: Gaussian and Rician distributions. Noise in MR images is often modeled as Rician, and for signal to noise ratio greater than two, the noise behaves like Gaussian [Gudbjartsson and Patz, 1995]. Furthermore, other sources of noise from the MR device are still modelled as Gaussian.

Tracking Algorithms

The MR-Linac typical rate of acquisition is four (4) frames per second [Paganelli et al., 2018]. Tracking algorithms must meet this time requirement. The first image and its corresponding tumor contour in the Cine-MR sequence is used as the reference and the tracking algorithms uses the subsequent Cine-MRI input images to estimate a new tumor contour. Fast et al. [Fast et al., 2017] presented a comparative study where they analyzed four tumor tracking techniques. The authors concluded that all the algorithms had a relatively similar performance but among them deformable image registration and template matching provided slightly better results. We implemented these two algorithms to evaluate tumor tracking simulator. Both algorithms are implemented on C++ and parallelized on CPU with OpenMP.

We chose the diffeomorphic demons [Vercauteren et al., 2009] as a fast solution of deformable image registration. Our approach uses a multiresolu-

tion framework with three pyramidal levels. The computational bottleneck of registration is the computation of the transformation and the similarity metric [Shams et al., 2010]. We focused on this stage to improve performance. Regarding the optimization, we fix the iteration values to be able to achieve the time restriction, this means that full convergence is not always guaranteed.

We implemented the generalized template matching algorithms [Cervino et al., 2011, Shi et al., 2014]. The algorithm defines a template in a reference image, usually the tumor bounding box, and search for it in a local region of the input image by maximizing an objective function to determine a good match. The preferred function is cross-correlation for Cine-MR tracking. The wider the search region is, the more computational time, while with a more limited search region, there is a risk of not capturing the tumor motion.

Metrics

We used three metrics to validate the real and simulated Cine-MRI and to evaluate the tumor tracking algorithms. First, the Dice Similarity Coefficient, which serves to quantify the whole segmented structure. Secondly, the centroid distance, which quantify the algorithm's ability to follow the center of mass (COM) of the tumor. And third, the Hausdorff distance, which provides a measurement of the effectiveness to detect and track the tumor contours. Further information regarding the metrics are detailed in Fast et al. [Fast et al., 2017].

5.5 Experiments and Results

Patient Summary

Table 5.1 summarizes the patients tumor and breathing motion. The patient set is small in numbers but represents a wide variety of anatomical tumor location, tumor size, breathing motion, and breathing cycle times. The most challenging conditions for the tracking algorithms are a short tumor displacement or a small tumor area.

Cine-MRI Image Quality

Figure 5.2 shows a comparison between the real Cine-MRI and the simulation for one patient. All the resulting videos are visually close to the real Cine-MRI. To validate the real versus the simulated Cine-MRI, we verify how consistent the tumor motion is for all the patients. We select 12 images (approximately a breathing cycle) of the real Cine-MRI with the first image in the inspiration position. The tumors are segmented manually on the real Cine-MRI. Since the real and the simulated data have a slightly different field of view, we perform rigid registration around the tumor on the first image and align the remaining

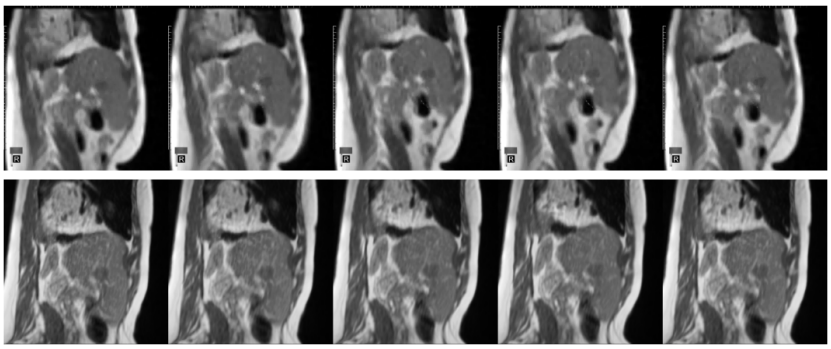


Figure 5.2: Video comparison of real Cine-MRI and simulated of one patient. Top images depict the real Cine-MRI and the bottom images the simulation.

simulation images using the same transformation. We calculate the metrics for all the patients with the real versus the simulated Cine-MRI. The Dice score is 0.89 ± 0.05 ($mean \pm std.dev.$), the centroid distance $0.78 \pm 0.32mm$ and Hausdorff distance $2.11 \pm 0.91mm$.

Tumor Tracking Performance

We evaluate tumor tracking with a full factorial experiment between breathing amplitude and noise. The breathing amplitude is varied with values 1.0, 1.5, 2.0, or random. The noise is varied between none, Gaussian and Rician. Gaussian noise and Rician noise applied in all the tests are equivalent to 20% of added noise. These experiments generate 12 videos per patient for a total of 84 Cine-MRI sequences. The breathing cycle parameter is patient-specific taken from Table 5.1. All the videos are 20 seconds long at 4 frames per second (80 images), approximately 4-5 breathing cycles.

Figure 5.3 depict the tumor tracking performance. For comparison, we compute a baseline (in blue) that corresponds to the metric value without tracking. Videos 1 to 4 vary in amplitude without noise, videos 5 to 8 vary in

Patient	1	2	3	4	5	6	7
Tumor location (geometric)	S-A-L	S-A-L	S-A-R	I-A-L	S-A-R	S-P-L	I-P-R
Breathing cycle (mean) [s]	4.8	3	4.1	3.6	4.4	4.1	5.8
Tumor max. displacement [mm]	5.9	9.1	8.1	9.7	2.0	6.5	12.6
Tumor volume [cm^3]	5.5	5.6	12.0	3.1	3.1	8.0	5.7
Tumor sagittal area [cm^2]	2.7	3.0	7.4	3.3	1.8	4.7	4.1

Table 5.1: Summary of patient information. Tumor location is the geometric octant of where the tumor is with regards to the liver center of mass. The abbreviations correspond to Superior-Inferior, Anterior-Posterior, and Left-Right. Breathing cycle times were determined from real patient respiratory motion during 4DCT scans. Tumor displacements refer to the maximum motion presented in the video without registration. Gross tumor volumes/areas are estimated on the reference 3D/2D (sagittal) MR.

amplitude with Gaussian noise, and videos 9 to 12 vary in amplitude with Rician noise. Video 1 is the most representative as it has the default and control conditions. The Dice score summary as $mean \pm standarddeviation$ results of registration are 0.88 ± 0.06 , template matching 0.79 ± 0.12. The centroid distance obtained for the registration is $0.89 \pm 0.54mm$ versus template matching $1.71 \pm 3.81mm$. The Hausdorff distance obtained for registration is $3.23 \pm 1.35mm$ versus template matching $4.41 \pm 4.02mm$. Both algorithms' performance in terms of centroid distances is adequate for image-guided radiotherapy.

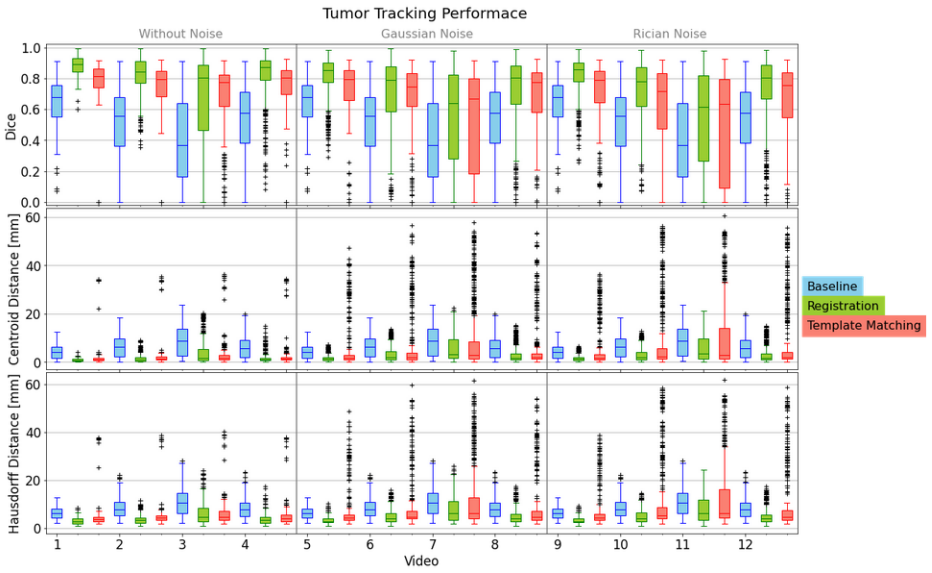


Figure 5.3: Video statistics of tumor tracking. The metrics are estimated per video and comprise all patients. Videos are generated as a full factorial experiment between breathing amplitude and added noise. Ascending numbers in groups of 4 videos correspond to breathing amplitude of 1, 1.5, 2, and random respectively. The metrics are shown from top to bottom as Dice, centroid distance, and Hausdorff distance. In blue, the baseline as the metrics computed without tracking. In green, the metrics determined with registration. In red, the metrics estimated with template matching.

All the tests were run on a workstation with 2 CPUs and 128 GB of RAM. Each CPU is an Intel(R) Xeon(R) Silver 4110 @ 2.10GHz, 8 cores, 16 threads. The computational time of the deformable registration algorithm time is on average 62.7 ms with a standard deviation of 42.3 ms. The maximum registration time is 242 ms. The computational time of the template matching algorithm is on average 16.7 ms with a standard deviation of 4.4 ms

5.6 Discussion and Conclusion

We validate the tumor motion between the real versus the simulated Cine-MRI. The centroid distance is the best metric to represent the motion, and its mean value distance of 0.78 mm indicates a high similarity between the real and the simulated data. We do not compare image intensities between the real and the simulated Cine-MRI because they come from sources acquired on different dates and under different conditions (pre-treatment - treatment, the field of view, alignment, among others). Furthermore, a direct comparison of pixel intensities or image similarity will only reveal how close they are in terms of signal, contrast, or even alignment but not how well the simulation model breathing and tumor motion.

Regarding our specific evaluation of tumor tracking algorithms, we identify in general that deformable image registration perform better. The template matching algorithm fails under noisy conditions and present several outliers. A breathing amplitude of 2.0 is an extreme condition and unrealistic. However, from the algorithms point of view is an interesting experiment. Both tracking algorithms fails to follow the tumor having wide ranges of Dice scores under this condition.

A limitation of our Cine-MRI simulator is that the breathing model uses a single respiratory cycle from the 4DCT scan. The breathing model overcomes this by composing transformations that are time interpolated. Time interpolation guarantees that different patterns arise due to asynchrony between the patient's breathing cycle time and sampling times. Furthermore, when we model with full 3D images and then create the 2D Saggital MR, we incorporate the desired out-of-plane motion, which is the main challenge for tracking algorithms. Overall, our goal is not to create a perfect breathing model but to facilitate challenging experiments to evaluate tumor tracking algorithms.

We designed a platform and a methodology to easily evaluate tracking algorithms on Cine-MR with ground truth segmentation. The video simulator does not require any training data and works only with pre-treatment images. The proposed methodology is the most automated way to evaluate tumor tracking algorithms with a ground truth. Our code is open source and available at https://github.com/josetascon/cinemri-simulation.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 764644. This paper only contains the author's views and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

Chapter 6

Tracking Multiple Organs

The work presented in this chapter is based on a paper that is to be submitted as: Tascón-Vidarte, J. D., Wahlstedt, I., Vogelius, I. R., & Darkner, S. (2022). Real-time tracking of multiple-organs for 2D Cine-MRI.

6.1 Abstract

Tracking on 2D Cine-MRI has focused primarily on tumors. In this paper, we evaluate multiple-organ tracking performance along with tumor tracking. We include the data of two liver patients and two lung patients within a retrospective clinical study that received stereotactic body radiation. Each patient has three treatment Cine-MRI sequences, for a total of 12 sequences. We compare four algorithms quantitatively using dice score, centroid distance, Hausdorff distance, and mean contour distance. Finally, we validate how well the multiple-organ tracking algorithms replicate convenient gating control signals for image-guided radiotherapy.
Keywords: *Real-Time, Tracking, Radiotherapy*

6.2 Introduction

The ultimate goal in image-guided radiotherapy is a fully controlled dose delivery to the tumor and avoid any delivery to the organs at risk [Murphy, 2004]. The introduction of MR-Linacs was a closer step to this goal [Paganelli et al., 2018]. The principle is to deliver radiation of what is seen with real-time Cine-MRI. One strategy that has come is beam gating [Crijns et al., 2011]. The strategy consists of stoping the dose delivery when the tumor is outside a boundary. Beam gating works well for patients during breath-hold. However, not all patients tolerate breath-hold for long periods, and they require free-breathing treatment [Høyer et al., 2012].

During free-breathing, the oncologists define a wider boundary. The beam gating strategy does not cover when other organs are close to the boundary region and when this occurs, the organs at risk receive increased radiation dose [Paganelli et al., 2018]. A beam gating strategy that considers multiple organ locations can be proposed to compensate for this. The first step to achieving such a strategy is to evaluate the performance of tracking algorithms with multiple organs.

The lung, the liver, and their associated tumors deform and move during treatment mainly caused by breathing motion [Murphy, 2004]. Tumor tracking is one of the main components of beam gating in the MR-Linac system. Tumor tracking is solved automatically using image analysis. Some proposed strategies for tumor tracking are based on template matching [Cervino et al., 2011, Shi et al., 2014], feature detection [Paganelli et al., 2015], optical-flow methods [Zachiu et al., 2015, Seregni et al., 2018], deformable image registration [Fast et al., 2017], segmentation [Gou et al., 2014], neural networks [Cervino et al., 2011, Yun et al., 2015, Friedrich et al., 2021] or modeling based [Garau et al., 2019]. Existing tumor tracking algorithms are accurate enough, but still, there is room for improvement [Friedrich et al., 2021].

Previously reported papers have evaluated tumor tracking during free-

breathing [Cervino et al., 2011, Gou et al., 2014, Shi et al., 2014, Paganelli et al., 2015, Zachiu et al., 2015, Fast et al., 2017, Seregni et al., 2018, Friedrich et al., 2021]. The breath-hold treatment is the most used respiratory motion management in practice [Feldman et al., 2019]. Breathing motion is more dramatic after breath-holding, and therefore the algorithms should perform well in this stricter scenario. Our study is the first to evaluate the tracking algorithms in this condition. Furthermore, most of the Cine-MRI sequences in former studies were acquired at 1.5T [Gou et al., 2014,Shi et al., 2014,Paganelli et al., 2015, Zachiu et al., 2015, Fast et al., 2017, Seregni et al., 2018] and Friedrich et al. [Friedrich et al., 2021] is the only study at 0.35T, i.e., inferior image quality due to lower signal to noise ratio. This study also uses images acquired at 0.35T.

There are studies with multiple algorithm evaluations of tumor tracking. The most comprehensive was Fast et al. [Fast et al., 2017] who compare four algorithms. The authors found that a B-splines registration and multiple template matching were the best performing algorithms. Recently, Friedrich et al. [Friedrich et al., 2021] compared tumor tracking between B-splines and U-net segmentation, the latter performed better. However, the B-splines algorithms presented by Fast et al. and Friedrich et al. did not achieve real-time operation at four frames per second. We exhibit here a B-spline algorithm working at one frame per second and alternatively a demons registration algorithm operating at eight frames per second. All the compared algorithms by Fast et al. and Friedrich et al. use ten templates images to work. Under normal conditions, the oncologist only delineates organs over a single image on treatment day, and therefore we use algorithms that fulfill this condition.

All the previous studies have only focused on tumor tracking, and to the best of our knowledge, this is the first implementation and evaluation of tracking with multiple organs in real-time.

6.3 Methods

Data

This study includes retrospective image data of patients treated at at Rigshospitalet (Copenhagen, Denmark) between June 2019 and January 2020. The patients provided informed consent and approval for the usage of their anonymized data for research purposes. The data consist of treatment Cine-MRI sequences from two (2) patients already treated with SBRT for metastases in the liver and two (2) patients treated SBRT in the lung. The patients were treated under breath-hold condition. An MRidian MR-Linac is used to acquire the Cine-MRI sequences at four (4) frames per second (MR @ 0.35T, bSSFP-Sagittal, resolution [256x256 pixels], spacing [1.5x1.5 mm]. We use three Cine-MRI sequences per patient (3 fractions) for a total of 12 Cine-MRI sequences. Each

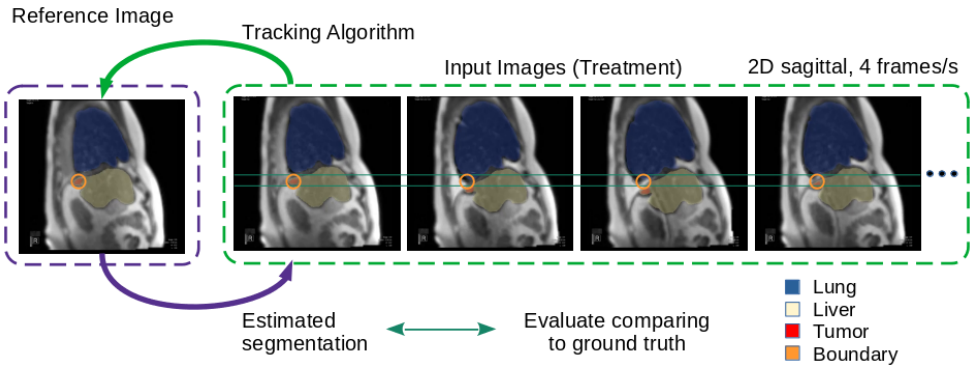


Figure 6.1: Pipeline used for multiple-organ tracking on 2D Cine-MRI. The patient corresponds to the second liver patient used to visualize the gating control signals.

sequence is 100 seconds long and contains breathing movement and breath-holds.

The field of view for all the patients covers the tumor, the lung, and the liver. The organs and the tumor were delineated on four random frames for each sequence when movement is occurring. For the second liver patient, who has a tumor in the upper position of the liver and close to the lung, one sequence is fully delineated to check the advantage of alternative gating control signals. This patient is shown in Figure 6.1.

Tracking System and Algorithms

The system pipeline is depicted in Figure 6.1. The algorithm inputs are the reference image, the reference organs masks (created from the delineation process), and the current input image. The output is the predicted organs masks corresponding to the current input image. We implement tracking algorithms based on deformable image registration and CNNs segmentation. The registration algorithms work similarly, where each input image from the Cine-MR is registered to the reference. The obtained transform is then applied to the reference organ masks and the predicted masks are obtained. The segmentation algorithms work as individual trackers functioning in parallel for each organ.

Demons registration (DEM)

We implement from scratch the diffeomorphic demons as a fast solution [Vercauteren et al., 2009]. The computational bottleneck of registration is the computation of the transformation and the similarity metric [Shams et al., 2010]. We primary focus on this stage to improve performance. Our reg-

istration algorithm is implemented on C++ and parallelized on CPU with OpenMP.

B-spline registration (BSP)

We implement B-spline registration algorithm [Rueckert et al., 2006] with cost function the sum of squared differences and optimizer LBFGS. The B-spline transform order is 3 and grid size 8 mm. The algorithm is implemented on C++ using the Insight Toolkit (ITK).

U-net segmentation (UNET)

The U-net algorithm [Ronneberger et al., 2015] is implemented in PyTorch. U-net is a supervised learning predictor and therefore only uses the current input image to predict the organs masks. The network is trained and validated patient-specific with a single image. To avoid over-fitting and improve the tracking performance we use image augmentation. The training hyper-parameters were: epochs = 1000, batch size = 1, and learning rate 10^{-3}. The augmentation parameters were: height shift = 20%, width shift = 10%, scale range = 20%, rotation range = 5^o, shear range = 10^o, gaussian noise = 20% of std. dev.

Siamese region proposal network for tracking (DRPN)

State of the art object tracking in computer vision has shown better performance with siamese region proposal networks [Li et al., 2018]. Object tracking in computer vision is based on bounding boxes. In our case we desire the organ mask and therefore we transform the reference (template) organ mask to fit the output bounding box. The algorithm is implemented on C++ using OpenCV's DaSiamRPN algorithm. The algorithm is pre-trained with Youtube-BB [Real et al., 2017].

Metrics

We used three metrics to evaluate the multiple-organ tracking algorithms. First, the Dice Similarity Coefficient, which serves to quantify the overlapping of the whole segmented structure. Second, the centroid distance, which quantifies the algorithm's performance to follow the organ center of mass. And third, the Hausdorff distance, which refers to the maximum distance between the organ contours, useful for detecting contour disagreement.

Gating Control Signals

Additional gating control signals are advantageous for patients with tumors located closely to organs at risk. An example of this kind of tumor is one placed

near the diaphragm for lung and liver patients. The second liver patient is an example of this condition (see Figure 6.1). We evaluate the conventional gating signal indicating that the tumor is inside the boundary, but we include two more, "liver inside the boundary" and "lung inside the boundary". The control signals are computed with logical operations using the organs masks. We compare qualitatively the algorithms' ability to mirror the ground truth signals.

6.4 Results

All the tests were run on a workstation with two CPUs, one GPU, and 128 GB of RAM. Each CPU is an Intel(R) Xeon(R) Silver 4110 @ 2.10GHz, 8 cores, 16 threads. The GPU is a Nvidia(R) GeForce(R) RTX 3090. The average computational time of the algorithms are DEM 112.7 ms (8.9 fps), BSP 746.8 ms (1.3 fps), UNET 13.5 ms (74.1 fps), DRPN 150.7 ms (6.6 fps).

Figure 6.2 depicts the algorithms tracking performance. UNET obtains the best median value performance for tracking the tumor with a 0.82 Dice score, 1.24mm centroid distance, and 2.5mm Hausdorff distance. For tracking the liver, DEM and UNET with 0.96 Dice score, DEM with centroid distance 0.63mm and DEM with Hausdorff distance 3.62mm. For tracking the lung, UNET with 0.97 Dice score, DEM and BSP with centroid distance 0.79mm and UNET with Hausdorff distance 3.62mm. The UNET algorithm is the best algorithm to track the tumor in terms of mean values for all the metrics. DEM algorithm shows a close behavior to UNET for the tumor. The registration algorithms DEM and BSP show better accuracy for tracking the organs with lower mean values and fewer outliers for centroid and Hausdorff distance. The reason is that UNET is optimized only for Dice, it occasionally generates some false positive structures, and therefore, greater values in Hausdorff distance appear. Instead, the registration algorithms perform consistently at a coarse scale for this field of view, where the liver and the lung are wider and well-defined structures. DRPN algorithm shows inferior performance compared to the others. However, it is acceptable considering its design is based only on bounding boxes, and its training did not include this specific dataset.

Figure 6.3 shows the algorithms' ability to replicate the gating control signals. DEM and BSP algorithms fail to detect some events for the "tumor inside the boundary" and "lung inside the boundary" signals. UNET is the best to replicate the gating control signals. DRPN mirrors perfectly the signal "tumor inside the boundary" but fails more with the remaining signals related to the organs. This is explained because DRPN design is based on bounding boxes and not fully deformable objects. The organs are expected to show more deformations in the sequence.

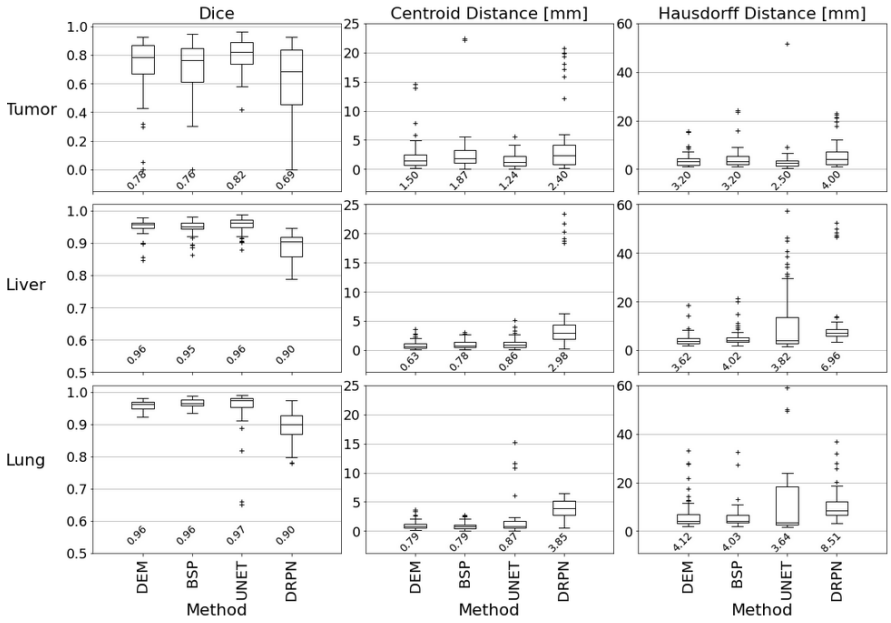


Figure 6.2: Results of multiple-organ tracking algorithms on Cine-MRI (sagittal-bSSFP) sequences of liver and lung patients. The median value is reported below each box plot.

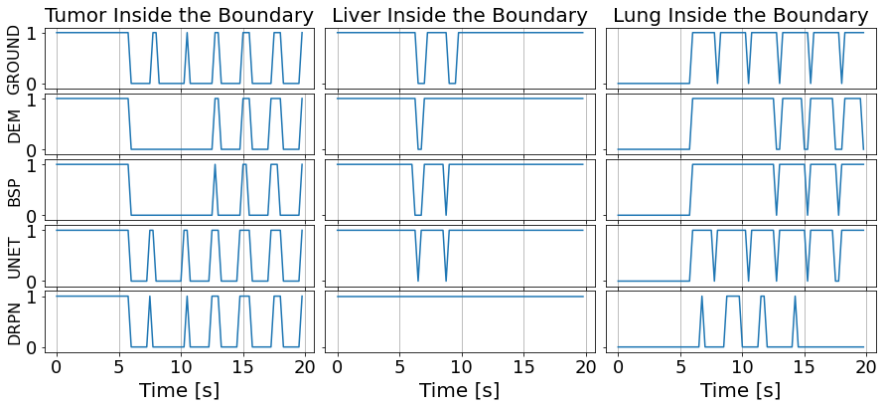


Figure 6.3: Convenient gating control signals available with multiple-organ tracking. Comparison of ground truth vs algorithms.

6.5 Conclusion

We demonstrate the implementation of successful real-time multiple-organ tracking algorithms. Deformable image registration DEM, BSP, and UNET segmentation are the most accurate methods in terms of dice, centroid distance, and contour distances. Furthermore, we evaluate the algorithms' ability to mirror gating control signals. Tracking with UNET shows superior perfor-

mance to replicate gating control signals. The "lung inside the boundary" signal could be employed during free-breathing treatment to reduce the lung dose delivery.

As opposed to the existing tracking studies, our algorithmic evaluation is made with multiple organs. Finally, we demonstrate the advantage of tracking multiple organs to replicate the new gating control signals that are available.

Acknowledgments

This project has received funding from the European Union's program "anonymous". This paper only contains the author's views and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

Compliance with ethical standards

This is a retrospective study without biological data and patients signed informed consent for use of the medical image data for research. The patients were treated according to standard practice in the department, in line with the principles of the Declaration of Helsinki.

Chapter 7

Real-time Tumor Tracking

The work presented in this chapter is based on a paper that is to be submitted as: Tascón-Vidarte, J. D., Wahlstedt, I., Jomier, J., Erleben, K., Vogelius, I. R., & Darkner, S. (2022). Real-time algorithms for tracking tumor and organs in 2D Cine-MRI.

7.1 Abstract

Background and purpose: MR-guided radiotherapy improves intra-fraction motion management using automatic dose delivery control based on real-time tumor tracking. We evaluate the accuracy and performance of nine (9) tracking algorithms on tumors and organs in simulation and treatment 2D Cine-MRIs used for high precision radiotherapy.
Methods: The tracking algorithms cover three strategies: deformable image registration (DIR), object tracking, and image segmentation. The simulation dataset represents free-breathing in seven liver patients. We generated 84 synthetic Cine-MRIs with the underlined tumor and organ delineations. The treatment dataset captured motion during breath-hold treatment of five liver patients and five lung patients. We included 30 Cine-MRIs with manual delineations of the tumor and the organ. We evaluated accuracy using the Dice score, centroid distance, Hausdorff distance, and mean contour distance.
Results: The best algorithms performed consistently well for simulation and treatment data. Tracking tumors on lung and liver patients offers no statistical significant differences regarding accuracy ($p > 0.05$). The best-performing algorithms achieved results close to inter-observer variability for treatment data.
Conclusion: UNET, TMDEM and CSRT were the best-performing algorithms overall for tumor tracking. DIR algorithms (DEM and BSP) perform better and are more stable for organ tracking.
Keywords: *Tracking, Cine-MRI, Real-time, Image-guided Radiotherapy*

7.2 Introduction

MR-guided radiotherapy is a milestone in external beam radiotherapy [Olsen et al., 2015, Raaymakers et al., 2017] and feasible with an MR-linac. This device combines magnetic resonance imaging with a linear accelerator. MR-linacs are becoming a choice for imaging, planning, and treatment of targets in soft tissue that move as a result of respiration, such as in the lung or liver [Paganelli et al., 2018, Dhont et al., 2020]. Image-guided radiotherapy with MR-linacs has proven to be versatile and adaptable for inter-fraction [Kontaxis et al., 2017] and intra-fraction [Paganelli et al., 2015] motion. Intra-fraction motion management allows tumor visualization and dose delivery control with the potential to reduce organs-at-risk exposure [Crijns et al., 2011]. Intra-fraction motion management is only feasible with automatic organ and tumor tracking techniques.

Tumor tracking can be solved automatically using image analysis. Some proposed strategies for tumor tracking are based on template matching [Cervino et al., 2011, Tryggestad et al., 2013, Shi et al., 2014, Menten et al., 2018, Fast et al., 2017], feature detection [Paganelli et al., 2015, Mazur et al., 2016, Fast

et al., 2017], optical-flow methods [Zachiu et al., 2015, Seregni et al., 2018], deformable image registration [Fast et al., 2017, Friedrich et al., 2021], modeling based [Garau et al., 2019], segmentation with variational methods [Gou et al., 2014], or segmentation with neural networks [Cervino et al., 2011, Yun et al., 2015, Friedrich et al., 2021]. Current tracking systems used in clinical practice may fail to track unexpected movements and have difficulty in tracking motion in the out-of-plane direction [Paganelli et al., 2018]. This leaves significant room for improvement in tumor tracking.

We observe a need to compare the performance of several automated algorithms under variant and challenging conditions such as fast or significant respiratory movement and noisy image data. We explore two ways to incorporate those conditions. The first way is with a simulation that allows variable motion and noise. The second way is to evaluate tumor tracking under fast movement in actual treatment. After breath-hold, the patients can exhibit faster breathing motion, and thus tracking becomes very difficult. Breath-hold treatment is the most used respiratory motion management in practice [Feldman et al., 2019]. Our study is the first to evaluate the tracking algorithms on breathing motion during breath-hold treatment. In contrast, the previously reported studies evaluate tumor tracking under free-breathing conditions [Cervino et al., 2011, Shi et al., 2014, Paganelli et al., 2015, Yun et al., 2015, Zachiu et al., 2015, Fast et al., 2017, Seregni et al., 2018, Garau et al., 2019, Friedrich et al., 2021].

Regarding the algorithms, only a few studies cover multiple algorithm evaluations of tumor tracking [Cervino et al., 2011, Fast et al., 2017, Friedrich et al., 2021]. The most comprehensive was Fast et al. [Fast et al., 2017] who compared four algorithms. The authors found that the two best-performing algorithms are deformable image registration (DIR) with B-splines and multiple template matching (MTM). Recently, Friedrich et al. [Friedrich et al., 2021] compared tumor tracking between DIR B-splines and UNET segmentation, the latter performed better. A drawback in both Fast et al. [Fast et al., 2017], and Friedrich et al. [Friedrich et al., 2021] is that all the compared algorithms use ten templates images to work. Under normal conditions, the oncologist only delineates organs and the tumor over a single image on treatment day, and therefore we use algorithms that fulfill this condition in our study. Increased delineations will require longer waiting times for patients on the couch. Having a single reference image motivates the inclusion of other automatic tracking algorithms.

A sub-branch of computer vision has been focussing on object tracking algorithms [Wu et al., 2013b, Ciaparrone et al., 2020]. The most successful algorithms belong to the class tracking-by-detection solved with correlation filters [Babenko et al., 2009, Henriques et al., 2012] or with deep learning [Bertinetto et al., 2016]. We are the first to include them in an evaluation with application to tumor and organ tracking in radiation therapy.

In summary, our contribution is a comprehensive evaluation of tumor

tracking including state-of-the-art medical imaging and computer vision algorithms. We also incorporate organ tracking in our comparisons. Our study covers the most challenging conditions for tracking, low image quality, and fast/significant breathing motion. We include liver and lung patients because they offer different imaging contrast and cover the most altered locations during breathing motion. We summarize the evaluation conditions of our study compared to previous work in Table 7.1.

Author	Breathing	Cine-MRI	FPS	MR Field	View	Organ	# Patients	# Videos	Algorithms
[Cervino et al., 2011]	Free, slow & force exhale, irregular	-	4	3T	Sagittal	Lung	5 *	16	TM, ANN
[Tryggestad et al., 2013]	Free	bSSFP	4	1.5T	Sag/Cor	Lung (2) and pancreas (1)	3	3	TM
[Shi et al., 2014]	Free	bSSFP	0.4	1.5T	Sagittal	Lung	12	12	TM
[Gou et al., 2014]	-	bSSFP	5	1.5T	Sag/Cor	Pancreas (stomach and liver)	2	5	Segmentation: EDRG, Level set, hGReS
[Paganelli et al., 2015]	Free	bSSFP	3.3	1.5T	Sagittal	Liver	30 **	30	SIFT
[Zachiu et al., 2015]	Free	-	12	1.5T	Sagittal	Liver and Kidneys	2 *	2	Optical Flow ([Horn and Schunck, 1981])
[Yun et al., 2015]	Free	bSSFP	4	0.5T †	Sagittal	Lung	4	4	PCNN
[Mazur et al., 2016]	-	bSSFP	4	0.35T	Sagittal	Several [a]	19	19	SIFT
[Fast et al., 2017]	Free	bSSFP	4	1.5T	Sag/Cor	Lung	6	22	MTM, PCNN, SIFT, DIR
[Menten et al., 2018]	Free	T1 [b]	-	1.5T	Coronal	Lung	14 ***	14	MTM
[Seregni et al., 2018]	Free	bSSFP	3.3	1.5T	Sagittal	Liver	30 **	30	TM + Optical flow ([Farnebäck, 2003])
[Garau et al., 2019]	-	bSSFP	3	-	Sag/Cor	Lung	2	2	DIR + PCA, offline
[Friedrich et al., 2021]	Free	bSSFP	4	0.35T	Sagittal	Liver	3	3	DIR, UNET
Ours	Free, Breath Hold	bSSFP	4	0.35T	Sagittal	Liver, Lung	7 Sim,10 ‡	84 Sim,30	9 §

Table 7.1: Summary of evaluation conditions from studies that include tumor tracking algorithms. Abbreviations: bSSFP = Balanced Steady-State Free Precession, TM = Template matching. MTM = TM with multiple templates. ANN = Artificial Neuronal Network, PCNN Pulse Coupled Neuronal Network. SIFT = scale-invariant feature transform. DIR Deformable Image Registration (B-spline transformation).

* Healthy
** 25 patients and 5 healthy
*** 8 patients and 6 healthy
[a] Stomach (5), lung (4), liver (4), adrenal glands (2), pancreas (2), spleen (1), mediastinum (1)
[b] Artificial 2D Cine-MRI from 4DMRI
† 3T pseudo 0.5T
‡ Sim refers to patients used to create simulation Cine-MRI. Only liver patients.
§ Algorithms: DEM, BSP, TM, TMDEM, UNET, MIL, KCF, CSRT, DRPN, see section 7.3

7.3 Materials and Methods

We compare the accuracy and computational time of nine (9) tracking algorithms on tumors and organs using simulation and treatment 2D Cine-MRIs.

Data

This study includes retrospective image data. All the patients were treated at Rigshospitalet (Copenhagen, Denmark) between April 2019 and January 2020. The patients provided informed consent and approval for the usage of their anonymized data for research purposes.

We incorporate 84 simulated Cine-MRIs of free-breathing motion from seven (7) liver patients (12 videos per patient). The advantage of using synthetic Cine-MRIs is that we generate a simulated sequence with a ground truth segmentation. Our previous paper [Tascón-Vidarte et al., 2021] describes the dataset and the simulation process (see Appendix A. Simulation Description).

The treatment data includes Cine-MRI sequences from five (5) patients treated with Stereotactic Body Radiotherapy (SBRT) for metastases in the liver and five (5) patients treated with SBRT in the lung. The patients were treated in breath-hold. An MRidian MR-Linac (ViewRay Inc., Mountain View, CA) is used to acquire the Cine-MRI sequences at four frames per second (MR @ 0.35T, bSSFP-Sagittal 2D, resolution [256x256 pixels], spacing [1.5x1.5 mm]). The liver patients received radiation doses in 3 fractions and the lung patients from 5 to 8 fractions. We use three Cine-MRI sequences per patient (3 fractions) for a total of 30 Cine-MRI sequences. The lung patients' fractions are randomly selected. From each fraction, we extracted 100 seconds of Cine-MRI sequence time. We only considered the frames with breathing motion after breath-hold for evaluation within this period.

Two imaging experts followed the same protocol to manually delineate the tumor and the organ. The observers segmented the organ and the tumor on four frames for each treatment Cine-MRI sequence. The first frame is at the breath-hold or end-inspiration position, and the other three are randomly selected when movement occurs after breath-hold. From here, we refer to the first frame as the reference image and its delineation as the reference tumor mask (or organ). Two liver patients showed low contrast in the tumor area. Therefore, we selected a surrogate structure for tracking and delineation (similar to [Paganelli et al., 2015]). The surrogates are anatomical features near the tumor. We analyzed inter-observer variability with the metrics described in section 7.3.

Tracking Algorithms

A tracking algorithm uses the reference image and the reference tumor mask (or organ) to estimate the new tumor position at every new frame. An identical process applies to the organ at risk.

We implemented nine (9) tracking algorithms based on deformable image registration (DIR), object tracking, and image segmentation. For more details see Appendix B. Algorithms. The recommended constraint of real-time tumor tracking is 0.5 s (2 fps) [Keall et al., 2006]. The algorithms presented here aim

for lower computational times than the typical acquisition time of 250 ms (4 fps) of commercial MR-linacs [Olsen et al., 2015, Raaymakers et al., 2017].

Demons registration (DEM)

We implemented the diffeomorphic demons as a fast solution of DIR [Vercauteren et al., 2009]. The computational bottleneck of registration is the computation of the transformation-interpolation process [Shams et al., 2010]. We primary focus on this stage to improve performance. We use a multi-resolution framework with three pyramid levels to improve convergence. Our registration algorithm is implemented in C++ and parallelized on CPU with OpenMP.

B-spline registration (BSP)

We implemented a B-spline DIR algorithm [Rueckert et al., 2006]. The cost function is the sum of squared differences, and the optimizer is LBFGS. The B-spline transform order is three (3) and grid size $8mm$. The algorithm is implemented on C++ using the Insight Toolkit (ITK) [ITK, 2021].

Template Matching (TM)

We implemented TM similar to Cerviño et al. [Cervino et al., 2011]. The metric for comparison is cross-correlation. First, we extract the template bounding box from the tumor or the organ mask, adding 10 pixels per side. Finally, we select a search window of 20 pixels around the target and run the metric in the search region. This algorithm is an in-house implementation in C++.

Template Matching with Demons Registration (TMDEM)

We propose a new algorithm based on template matching for global tracking combined with demons deformable image registration for local refinement. The TM works the same as described earlier, and the demons algorithm works with two pyramid levels.

U-net segmentation (UNET)

U-net is a supervised learning segmentation method and therefore only uses the current input image intensities to predict the organs masks [Ronneberger et al., 2015]. We train and validate the network patient-specific with a single image. In the simulation, we test the training using one image (UNET) versus using ten images (UNET10), the latter similar to Fiedriech et al. [Friedrich et al., 2021]. The algorithms are implemented in PyTorch.

To avoid over-fitting and improve the tracking performance we use image augmentation. The training hyper-parameters of UNET were: epochs = 1000,

batch size = 1, and learning rate 10^{-3}. The augmentation parameters were: height shift = 20%, width shift = 10%, scale range = 20%, rotation range = 5^o, shear range = 10^o, gaussian noise = 20% of standard deviation. UNET10 was trained with the same parameters but with epochs = 150.

Multiple Instance Learning (MIL)

The MIL algorithm train a discriminative classifier in an online manner to separate the object from the background [Babenko et al., 2009]. For each new frame, the classifier estimates the object location expressed as a probability function, and then the classifier is updated. This algorithm and all the ones described next are object tracking algorithms implemented in C++ using OpenCV [OpenCV, 2021] with their default parameters.

Kernelized Correlation Filter (KCF)

KCF is a dense sampling approach that exploits circulant matrices computed as kernels in the Fourier domain [Henriques et al., 2012]. The algorithm calculates an initial kernel model from the reference frame, and then each new frame detection is used to update the model.

Discriminative Correlation Filter with Channel and Spatial Reliability (CSRT)

CSRT algorithm incorporates channel and spatial corrections during the bounding box update [Lukezic et al., 2017]. For our case, Cine-MRI sequences are single-channel images (grayscale) and, therefore, channel correction is not relevant.

Siamese region proposal network for tracking (DRPN)

State of the art object tracking in computer vision has shown better performance with siamese region proposal networks [Li et al., 2018]. The algorithm is implemented on C++ using OpenCV's DaSiamRPN algorithm [Zhu et al., 2018]. The algorithm is pre-trained with Youtube-BB [Real et al., 2017].

Metrics for Algorithm Evaluation

We used four metrics to evaluate the tumor tracking algorithms with the Cine-MRIs. The metrics are: Dice similarity coefficient, centroid distance, Hausdorff distance and mean contour distance. The metrics are detailed in Appendix C. Metrics.

We depict in each comparison plot a baseline (called BASE) that corresponds to the metrics values considering the tumor always at breath-hold position. For treatment, we include the interobserver performance (called OBS).

We compare the algorithms' accuracy statistically in simulation and treatment. We run multiple hypothesis tests with Kruskal–Wallis method [Corder and Foreman, 2011], followed by pairwise analysis using Dunn test [Dunn, 1964] with Benjamini–Hochberg correction [Benjamini and Hochberg, 1995]. For treatment, we used Kolmogorov-Smirnov test [Massey Jr, 1951] to compare performance over lung versus liver patients.

7.4 Results

All the tests are run on a workstation with two CPUs, one GPU, and 128 GB of RAM. Each CPU is an Intel(R) Xeon(R) Silver 4110 @ 2.10GHz, 8 cores, 16 threads. The GPU is a Nvidia(R) GeForce(R) RTX 3090. The UNET and DRPN algorithms run on GPU, and the remaining algorithms (DEM, BSP, TM, TMDEM, MIL, KCF, CSRT) run on CPU.

The average computational time (and corresponding frames per second, fps) of the algorithms are: KCF 1.6 ms (620.1 fps), UNET 4.3 ms (232.5 fps), TM 18.3 ms (54.8 fps), TMDEM 18.8 (53.3 fps), CSRT 20.9 ms (47.8 fps), DRPN 56.9 ms (17.6 fps), MIL 62.2 ms (16.1 fps), DEM 112.7 ms (8.9 fps), BSP 746.8 ms (1.3 fps). The average training time of UNET10 is 2.2 min, and UNET is 7.5 min.

Simulation Data

We collect the results for all the generated Cine-MRIs and all patients in Figure 7.1. Thus, the plot covers all the simulation conditions, i.e., 12 videos. Individual video results are illustrated in Appendix G. Simulation Extended Results. The four best-performing algorithms for tumor tracking based on median values considering all the metrics are DEM, TMDEM, TM, and UNET10. The four best-performing algorithms for tracking the liver are BSP, TM, TMDEM, and UNET10.

We compare in simulation UNET10 and UNET. There are statistically significant differences in all the metrics between UNET10 and UNET ($p < 0.001$). UNET10 performed better due to the more extensive training data. UNET is, however, relatively close to UNET10 in median values accuracy ($\text{Dice}_{P_{5\%}}$ 0.52 and 0.5 respectively).

Treatment Data

Visual quality

Figure 7.2 illustrates the generated contours of tumor tracking algorithms on one liver patient (Figure 7.2a) and one lung patient (Figure 7.2b). For these particular patients, KCF and MIL algorithms perform the poorest qualitatively. The other algorithms require quantitative comparisons for assessment.

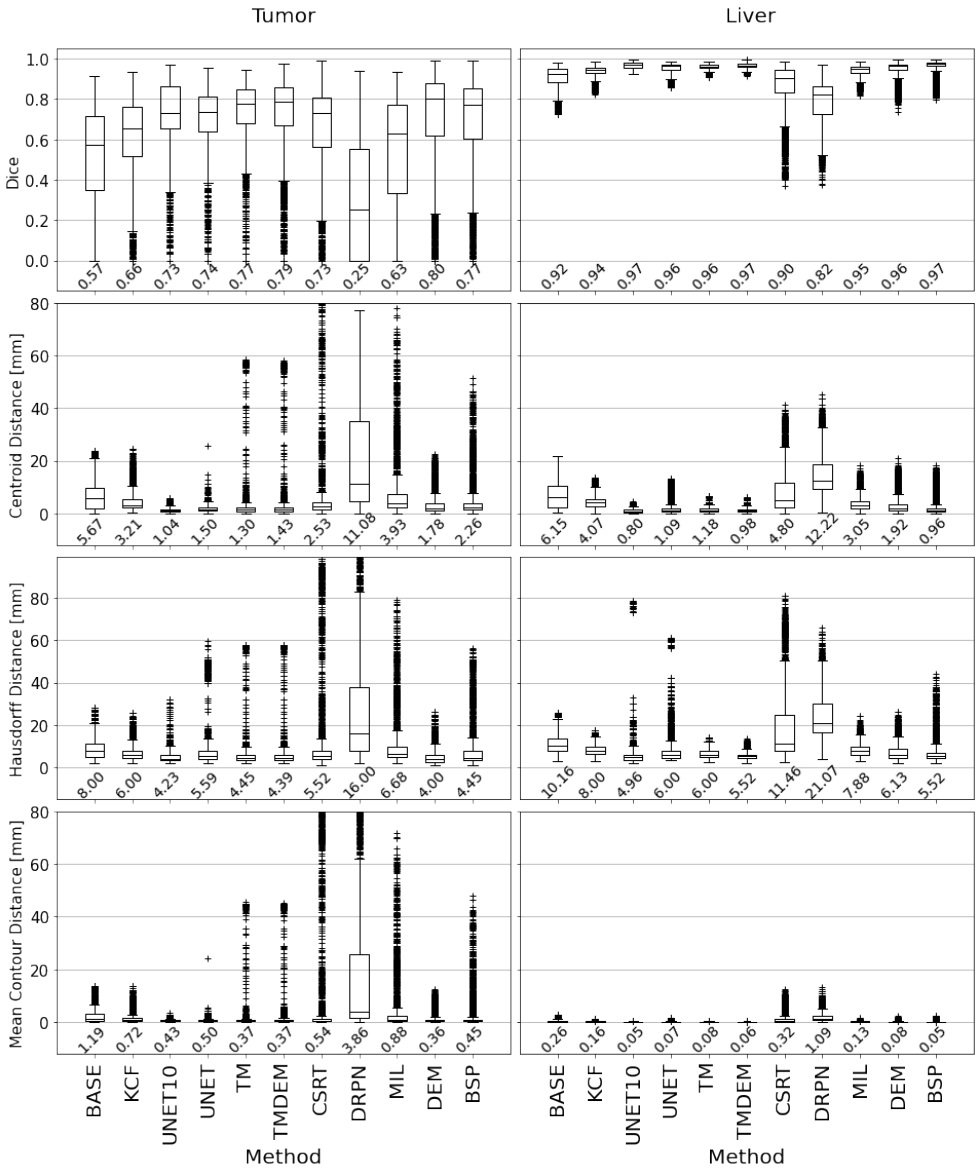
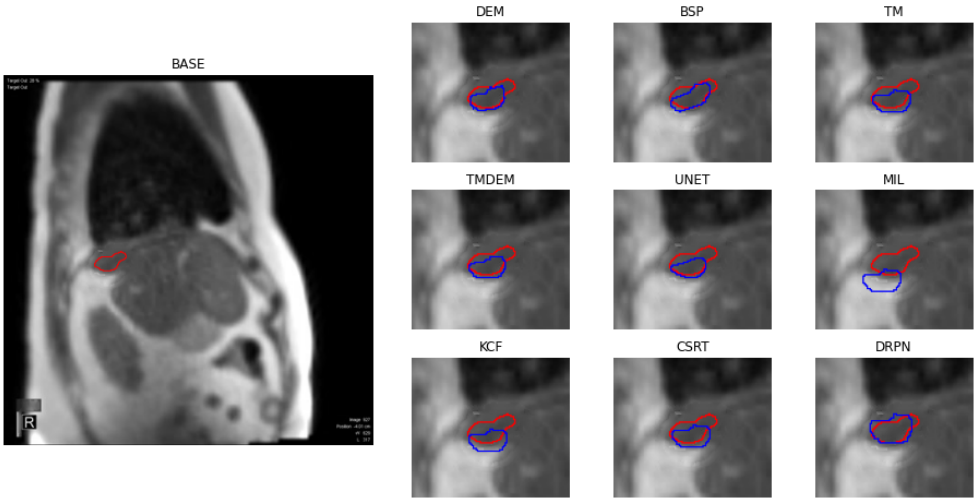


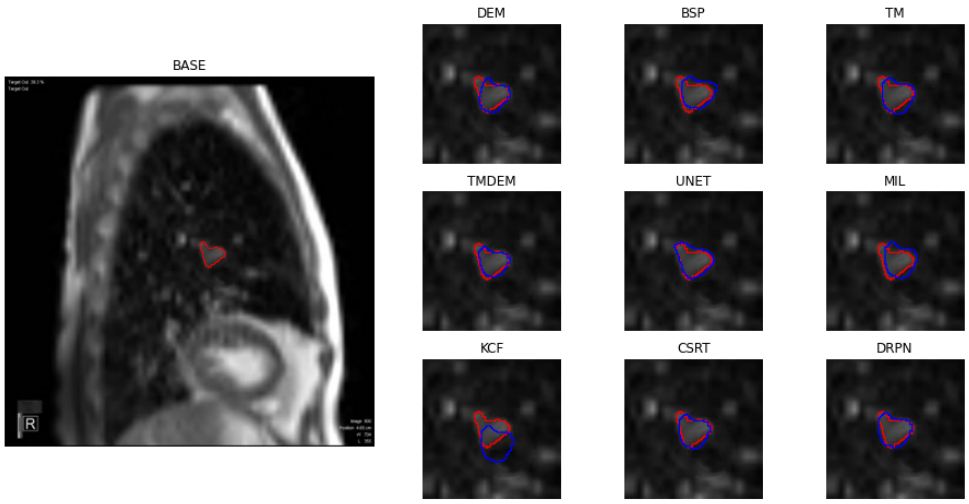
Figure 7.1: Results of tracking algorithms with liver on simulated Cine-MRI. The simulation data resemble free-breathing with varying motion amplitudes and image noise. We calculate the metrics with all the patients. Below each box plot is the median value.

Interobserver analysis

Figure 7.3 depicts the interobserver analysis. Comparing the interobserver performance for tumor delineations of liver and lung patients shows only statistically significant differences for the centroid distance ($p < 0.05$).



(a) Liver patient



(b) Lung patient

Figure 7.2: Visual comparison of manual (red) and automatically generated contours from tracking algorithms (blue). The larger image on the left is the reference image at breath-hold position. The smaller images on the right correspond to a random sample during breathing motion after breath-hold.

Algorithms performance

The quantitative results of the algorithms are depicted in Figure 7.4 and 7.5. We report the best-performing algorithms based on median values. Regarding liver patients, the four best-performing algorithms for tumor tracking are UNET, TMDEM, TM, CSRT. For tracking the liver, the best algorithms are DEM, BSP, UNET, TMDEM. CSRT algorithm has fewer outliers, low

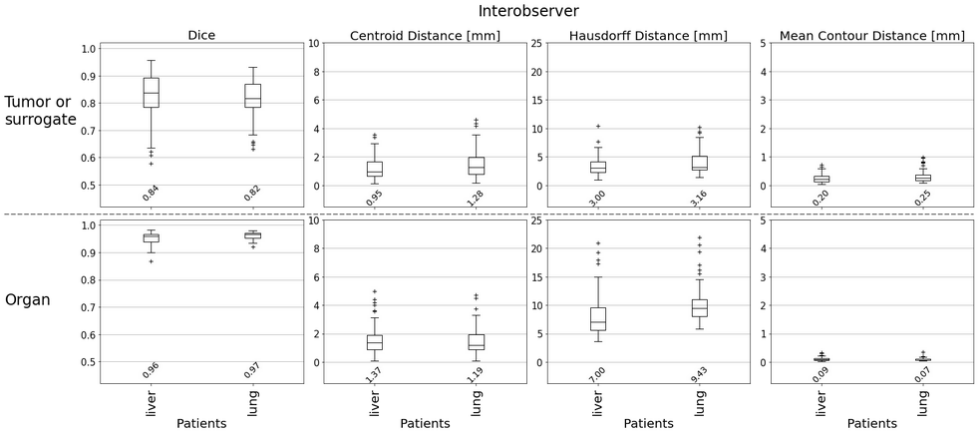


Figure 7.3: Results of inter-observer variability analysis. Top row plots correspond to the metrics computed for the tumor or the surrogate selected for tracking. Bottom row plots correspond to the organ results, i.e., lung for lung patients and liver for liver patients. The median value is reported below each box plot.

interquartile distance in Dice (IQR: 0.15), and is very close to the other algorithms performance. Regarding lung patients, the four best-performing algorithms for tumor tracking are UNET, DRPN, CSRT, TMDEM. For tracking the lung, the best algorithms are UNET, BSP, DEM, TMDEM.

We compare the best-performing algorithms on each metric with the interobserver results. Regarding the liver patients, the observers have a better Dice score ($p < 0.05$), and there are no statistically significant differences for centroid distance, Hausdorff distance, and mean contour distance ($p > 0.05$). Regarding the lung patients, the observers have worse Hausdorff distance ($p < 0.05$), and there are no statistically significant differences for Dice, centroid distance, and mean contour distance ($p > 0.05$).

7.5 Discussion

Most of the previously reported tracking evaluations were made in a 1.5T MRI scanner [Gou et al., 2014, Shi et al., 2014, Paganelli et al., 2015, Zachiu et al., 2015, Fast et al., 2017, Seregni et al., 2018] (see Table 7.1). The only reported studies with an MRI scanner with a lower field (0.35T) are Mazur et al. [Mazur et al., 2016], and Friedrich et al. [Friedrich et al., 2021]. This study also uses images acquired at 0.35T. An MR image from a scanner with a lower field offers a lower signal-to-noise ratio, representing a more challenging condition for tumor tracking.

Regarding the simulation conditions, Figure 7.1 depicts the accumulated results of all the variable conditions. For a general picture of how the algo-

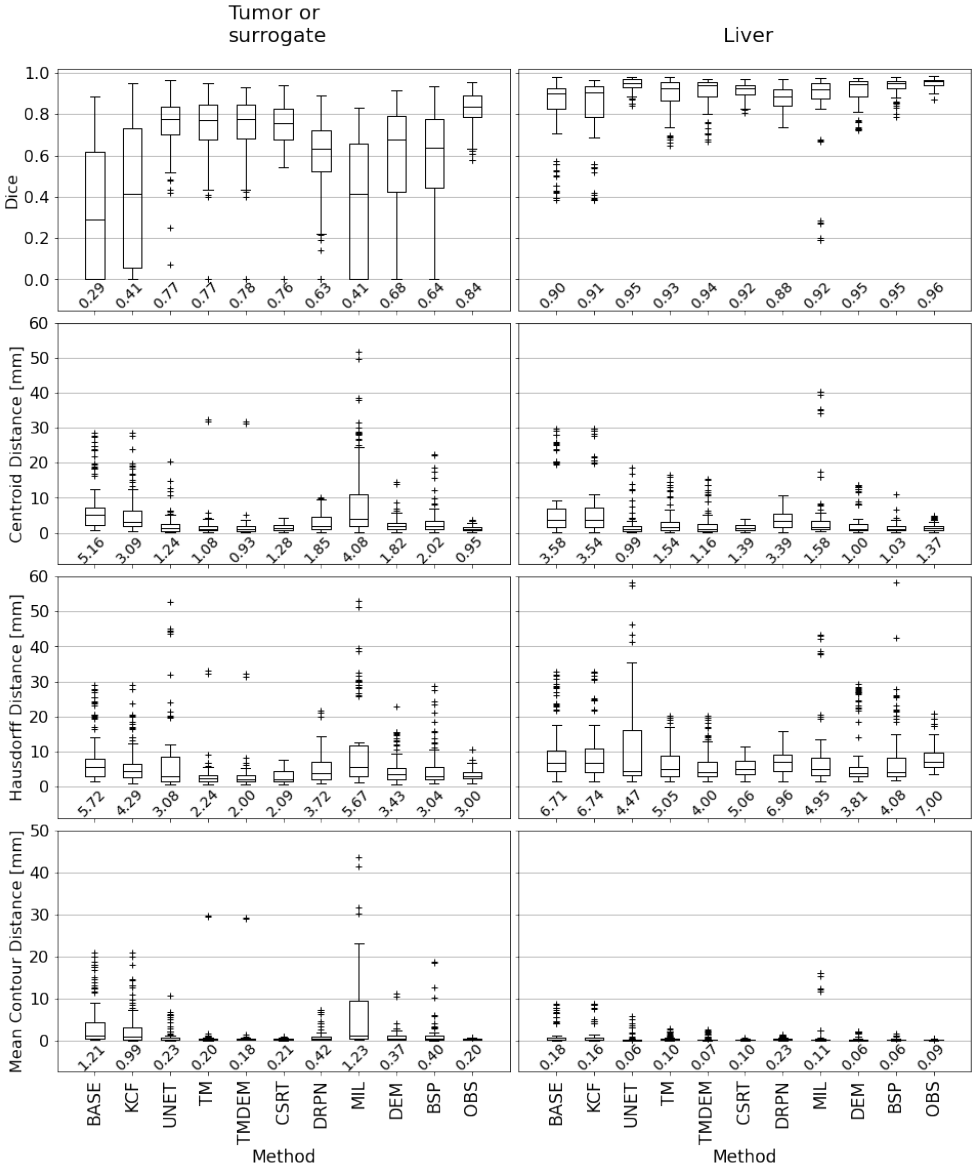


Figure 7.4: Results of tracking algorithms for the treatment dataset of liver patients. Cine-MRI (sagittal-bSSFP) sequences capture breathing motion during breath-hold treatment. The median value is reported below each box plot.

rithms perform with variable motion and noise, see Supplementary Material Appendix G. Simulation Extended Results. We observe better performance of DIR algorithms (DEM and BSP). DIR algorithms are more suitable for noisy conditions. The reason is that regularization maintains consistency at the global level. The drawback is that for large motions, the limited optimization steps to achieve real-time becomes a constraint. The TM, TMDEM,

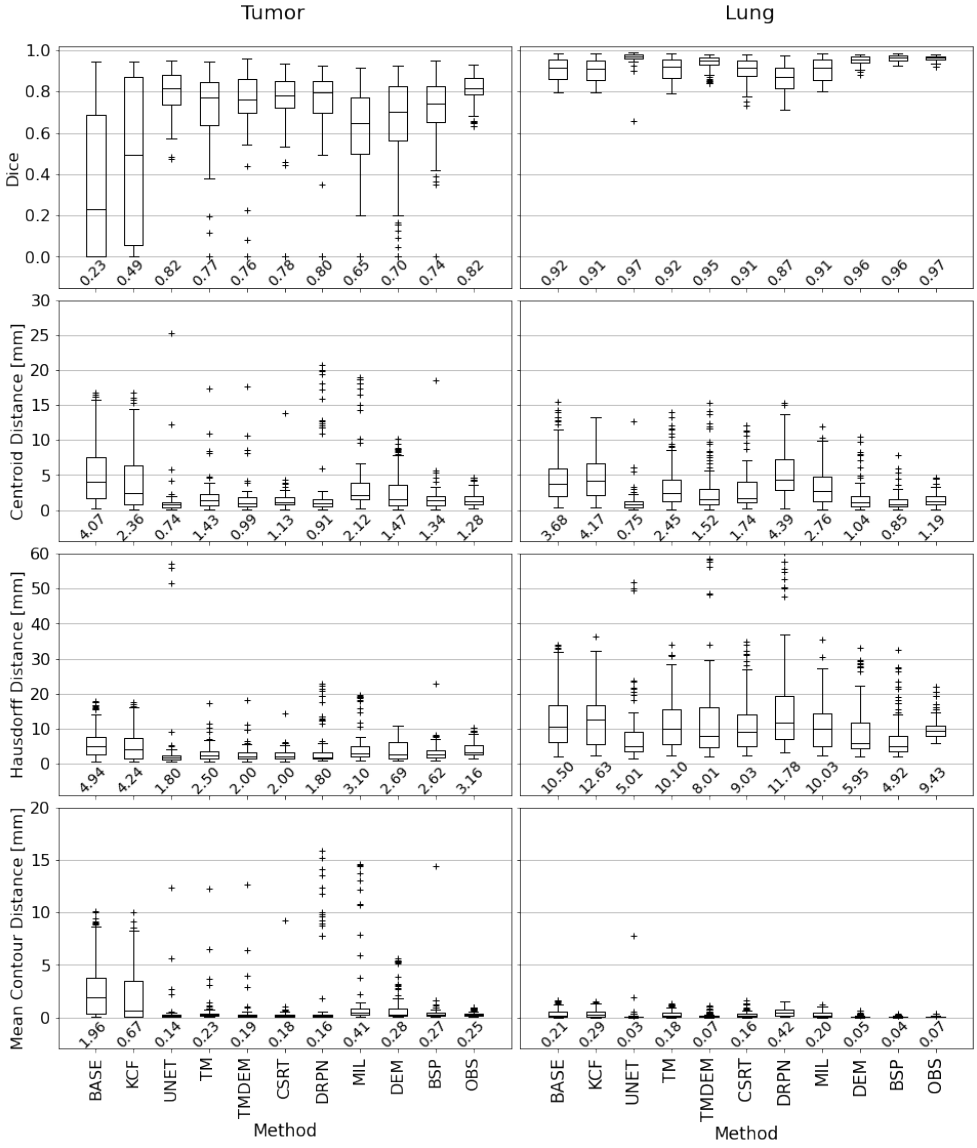


Figure 7.5: Results of tracking algorithms for the treatment dataset of lung patients. Cine-MRI (sagittal-bSSFP) sequences capture breathing motion during breath-hold treatment. The median value is reported below each box plot.

UNET10, and UNET algorithms show slightly lower accuracy to DEM and BSP. However, these four algorithms perform better with the variable motion amplitudes.

The interobserver analysis showed no statistically differences in tumor delineation of liver patients versus lung patients ($p > 0.05$). Consistently, the best algorithms perform similarly in both scenarios ($p > 0.05$). Comparing the interobserver performance with the best algorithms, observers show only

inferior performance in terms of Hausdorff distance for lung patients. The reason is that Hausdorff distance is a metric for detecting outliers in contour distance, and any failure is heavily considered.

Regarding tumor tracking on the treatment data, we identify that UNET, TMDEM, and CSRT algorithms perform consistently better. A disadvantage of the UNET algorithm is the training time. The required time of training for in-room treatment could be excessive. In this sense, a good alternative is TMDEM, a ready-to-use algorithm for tumor tracking with good accuracy and time performance.

Deformable image registration is a suitable algorithm to solve unexpected deformations from out-of-plane motion. For instance, we observe the better accuracy of DEM and BSP for tracking organs that deform more than tumors where the structure is geometrically simpler. The BSP algorithm has slightly superior accuracy than DEM in simulation and treatment but does not fulfill the real-time constraint to work on MR-linacs at 250 ms (4 fps).

To the best of our knowledge, our DEM algorithm is the first real-time deformable registration algorithm to achieve less than 250 ms of performance, i.e., 112.7 ms (8.9 fps). Our BSP algorithm performs at 746.8 ms (1.3 fps). In comparison, the BSP method presented by Fast et al. run at 500 ms (2 fps), and the method presented by Friedrich et al. run at 1300 ms (0.77 fps). New MR-Linacs are evolving with higher acquisitions rates [Paganelli et al., 2018], and therefore, faster and robust tracking algorithms are relevant.

TM and TMDEM algorithms perform consistently well for simulation and treatment data. TMDEM represents an improvement due to the extra step to deform the local area. Computational time does not increase significantly because the registration is made on a small area and with a reduced optimization setup compared to the full-scale DEM algorithm.

MIL and KCF algorithms were inadequate for tumor and organ tracking. CSRT offers tracking with fewer outliers for liver and lung patients, low interquartile distance in Dice (IQR: 0.15), and is very close to the best algorithms performance for the treatment data. This is because CSRT updates the bounding box scale and keeps track of the foreground and background histograms.

DRPN performs better for the treatment data compared to simulation. Its performance is acceptable considering that DRPN was trained with a set of outdoors and indoors objects and not specifically with tissue or organs. A domain adaptation training [Wang and Deng, 2018] can improve DRPN performance with medical images. Retraining DRPN is out of the scope of this evaluation and is worth considering in further studies.

Object tracking in computer vision is a rapidly evolving field with several new proposals per year [Ciaparrone et al., 2020]. As we showed with our evaluation, algorithms such as CSRT and DRPN a worth considering for tumor tracking.

7.6 Conclusion

We extensively evaluated tracking algorithms for tumors and organs on 2D sagittal Cine-MRIs. The evaluation included multiple liver patients and variant conditions with simulation data and the most challenging conditions for lung and liver patients with treatment data. Our study includes only automated tumor tracking algorithms that use a single reference. The algorithms with superior accuracy and time performance are UNET, TMDEM, DEM, and CSRT. All the algorithms code is open source and available at https://github.com/josetascon/tracking-tissue.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 764644. This paper only contains the author's views and the Research Executive Agency and the Commission are not responsible for any use that may be made of the information it contains.

7.7 Appendix

Appendix A. Simulation Description

We use our patient-specific Cine-MRI simulator [Tascón-Vidarte et al., 2021]. The system is capable to generate a Cine-MRI with simulated ground truth contours of the desired organs using pre-treatment images. The study validated that the simulation represent breathing motion. The input images are a 4DCT scan and an MR with organ contours. The video simulator has the following input parameters: video time, frames per second, breathing cycle time, breathing amplitude, and additive noise.

For our study we create 12 Cine-MRIs videos per patient for a total of 84 Cine-MRI sequences. The input parameters values are: video time = $80s$, frames per second = 4, breathing cycle is patient-specific obtained from 4DCT respiratory signal (range $3 - 5.8s$). The breathing amplitude is varied with values 1.0, 1.5, 2.0, or random (range $1.0 - 2.0$). The noise is varied between none, Gaussian and Rician. Gaussian and Rician noise applied in all the tests are equivalent to 20% of added noise. Videos 1 to 4 vary in amplitude without noise, videos 5 to 8 vary in amplitude with Gaussian noise, and videos 9 to 12 vary in amplitude with Rician noise. Video 1 is the most representative as it has the default and control conditions. Table 7.2 describe the variable parameters.

Video	Breathing Amplitude	Noise
1	1.0	None
2	1.5	None
3	2.0	None
4	Random	None
5	1.0	Gaussian
6	1.5	Gaussian
7	2.0	Gaussian
8	Random	Gaussian
9	1.0	Rician
10	1.5	Rician
11	2.0	Rician
12	Random	Rician

Table 7.2: Cine-MRI videos and their corresponding breathing amplitude and noise parameters

Appendix B. Algorithms

The tracking algorithms cover three strategies: deformable image registration (DIR), object tracking, and image segmentation. The DIR algorithms find a transformation between the reference and each input image from the Cine-MRI. The resulting transform is then applied to the reference tumor mask to obtain the predicted mask. Computer vision proposals solved object tracking with a group of algorithms called tracking by detection. Here a classifier model locates the object, and then a model update is produced for each frame. A bounding box represents the object's location within the image. In our case, we desire the tumor mask, and therefore we transform the reference tumor mask to fit the output bounding box. The segmentation algorithms work as individual trackers running in parallel for the organ and the tumor. The segmentation algorithms are pre-trained first and afterward only use the input image intensities to predict the masks.

All the algorithms code is open source and available at `https://github.com/josetascon/tracking-tissue`. We extended the details of the first four algorithms for reproducibility purposes and because they provide the background of our proposed method TMDEM.

Demons registration (DEM)

The demons algorithm is a DIR algorithm based in optical flow [Thirion, 1998]. The diffeomorphic demons algorithm is an extension of the original algorithm with a constraint to the transformation to be diffeomorphic [Vercauteren et al., 2009].

A registration problem is defined as: Having a fixed image $I_0(x)$ and a

moving image $I_1(x)$ find the spatial transformation $\varphi(x)$ between them. The transformation is typically a deformation vector field defined as:

$$\varphi_x : \varphi(x) = \mathbf{x} + u(\mathbf{x}) \tag{7.1}$$

Where $\mathbf{x}$ are the spatial coordinates (in $\mathbb{R}^d$) and $u(\mathbf{x})$ are the displacement vectors.

A registration algorithm solves a cost function E with a similarity metric $E_{\mathcal{M}}$ and regularization $E_{\mathcal{R}}$ components. An optimization routine will find the transformation $\varphi(x)$ parameters. The mathematical expression is:

$$\arg\min_{\varphi(x)} E = E_{\mathcal{M}}(I_0, I_1 \circ \varphi(x)) + E_{\mathcal{R}}(\varphi(x)) \tag{7.2}$$

Where:

$$x_j \, \epsilon \, \Omega_j, \; \Omega_j \subset \mathbb{R}^d, \; j = 0, 1$$

$$I_j : \Omega_j \to \mathbb{R}, \; \varphi(x) : \Omega_0 \to \Omega_1$$

The demons algorithm can be considered as an approximation of a second order gradient descent on the sum of square of intensity differences (similarity metric) criterion [Pennec et al., 1999] with Gaussian regularization. The deformation field for demons algorithm has an update:

$$F(\nabla E(\varphi_x)) \equiv \frac{(I_0 - I_1(\varphi_x))\nabla I_1(\varphi_x)}{\|\nabla I_1(\varphi_x)\|^2 + (I_0 - I_1(\varphi_x))^2} \tag{7.3}$$

The iterative process to update the transformation $\varphi(x)$ is defined as:

$$\begin{aligned} compute: &\quad u(x) = F(\nabla E(\varphi_x)) \\ fluid\,update: &\quad u(x) \leftarrow G^{\sigma_f} * u(x) \\ diffeomorphic: &\quad \varphi(x)' \leftarrow exp(u(x)) \\ diffusion\,update: &\quad \varphi(x) \leftarrow G^{\sigma_d} * \varphi(x)' \end{aligned} \tag{7.4}$$

Where G refers to a Gaussian kernel. The expression $exp()$ refers to the intrinsic update on the Lie group of diffeomorphisms. For more details please refer to Vercauten et al. [Vercauteren et al., 2009].

We select a multi-resolution framework with three levels (8,4,2). We limit to iterations to (120,100,800) for each level to achieve real-time performance. The diffusion update has smoothing Gaussian sigma of 2.5 for each iteration. We do not apply the fluid update.

B-spline registration (BSP)

The B-splines transformation approximate the dense displacement $u(\mathbf{x})$ of equation 7.1 as:

$$u(\mathbf{x}) = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3}\beta_x\beta_y\beta_z c_{i,j,k} \tag{7.5}$$

Where c are control points that parametrize the transformation. The B-spline transform order is three (3) and grid size $8mm$. The similarity metric of our B-splines algorithm is the sum of squared differences defined as:

$$E_{\mathcal{M}} := SSD = \frac{1}{N}\sum_{x\epsilon\Omega_0}^{n}(I_0(x) - I_1 \circ \varphi(x))^2 \tag{7.6}$$

We use an explicit B-splines regularization [Tustison et al., 2013] implemented with ITK [ITK, 2021]. The optimizer is LBFGS. A multi-resolution framework is selected with two levels (8,4) and corresponding smoothing sigmas (1,0). We limit to 30 iterations on each level to achieve real-time performance.

Template Matching (TM)

Template matching consists in defining a template and a comparing metric to determine the best location of the template in the input image by searching viable locations within the search image. The metric for comparison is cross-correlation defined as:

$$CC(I_0, I_1, \varphi(x)) = \frac{\left(\sum_{x\epsilon\Omega_0}^{n}(I_0(x) - \bar{I}_0)\cdot(I_1\circ\varphi(x) - \bar{I}_1)\right)^2}{\sum_{x\epsilon\Omega_0}^{n}(I_0(x) - \bar{I}_0)^2 \cdot \sum_{x\epsilon\Omega_0}^{n}(I_1\circ\varphi(x) - \bar{I}_1)^2} \tag{7.7}$$

First, we extract the template bounding box with coordinates $\mathbf{x}_0$ from the tumor or the organ mask, adding 10 pixels per side. Next, we select a search window of 20 pixels around the target and run the metric in the search region. The minimum value is selected. The search will find the coordinates $\mathbf{z} = \mathbf{x}_0 + (i,j)^T$. The search process is expressed as:

$$\arg\min_{\mathbf{z}} W = CC(I_0^{box}, I_1^{box}, \mathbf{x}_0 + (i,j)^T) \quad for\ \ i = -20,..,20; j = -20,..,20 \tag{7.8}$$

Where $I^{box} \subset I$, is the image subset extracted from the larger image with the bounding box. After finding the bounding box coordinates in the input image a mask with the tumor or organ is created with the estimated location.

Template Matching with Demons Algorithm (TMDEM)

We propose a new algorithm based on template matching for global tracking combined with demons deformable image registration for local refinement.

The TM works as described in equation 7.8. After finding the bounding box location we extract the subset images $I_0^{box}, I_1^{box(\mathbf{z})}$. We run the demons algorithm of equation 7.4 using the subsets images. The found transformation is applied to the mask located at the estimated bounding box. The demons algorithm works with two pyramid levels (2,1) and limited iterations per level (100,80).

Appendix C. Metrics

We use four metrics to compare the accuracy of the tracking algorithms. First, the Dice Similarity Coefficient, which serves to quantify the whole segmented structure. Second, the centroid distance, which quantifies the algorithm's ability to follow the tumor's center of mass (COM). Third, the Hausdorff distance, which provides a measurement for detecting significant contour disagreement. And fourth, the mean contour distance also known as mean distance to agreement, which offers an average measurement of the organ contours distance. We apply the same metrics for the inter-observer analysis. Each metric is described below.

The Dice Similarity Coeffient (DSC) is defined as:

$$DSC(X,Y) \equiv \frac{2|X \cap Y|}{|X| + |Y|} \tag{7.9}$$

where X is a control segmentation structure and Y is the estimated structure.

The centroid distance is defined as:

$$d_{com}(X,Y) \equiv \|X_{com} - Y_{com}\| \tag{7.10}$$

where X_{com} is the control structure center of mass and Y_{com} the estimated structure center of mass.

The Hausdorff distance is defined as:

$$d_H(X,Y) \equiv \max \left\{ \max_{x \in A_X} \left(\min_{y \in B_Y} \|(x-y)\| \right), \max_{y \in B_Y} \left(\min_{x \in A_X} \|(y-x)\| \right) \right\} \tag{7.11}$$

where x is a point in the contour A_X related to structure X, and y is a point in the estimated contour B_Y related to structure Y.

The mean contour distance is defined as:

$$d_H(X,Y) \equiv \max \left\{ \underset{x \in A_X}{mean} \left(\min_{y \in B_Y} \|(x-y)\| \right), \underset{y \in B_Y}{mean} \left(\min_{x \in A_X} \|(y-x)\| \right) \right\} \tag{7.12}$$

Appendix D. Quantitative Results

Region	Method	Dice		Centroid Distance [mm]		Hausdorff distance [mm]		Mean contour distance [mm]	
		Median	$P_{5\%}$	Median	$P_{95\%}$	Median	$P_{95\%}$	Median	$P_{95\%}$
Tumor	BASE	0.57	0.02	5.67	16.41	8.00	18.24	1.19	7.27
	KCF	0.66	0.13	3.21	12.45	6.00	14.30	0.72	4.69
	UNET10	0.73	**0.52**	**1.04**	**3.03**	**4.23**	9.01	0.43	**0.98**
	UNET	0.71	**0.52**	1.65	**4.15**	5.59	46.82	0.53	2.05
	TM	0.77	0.45	**1.30**	4.19	4.45	**8.86**	**0.37**	**1.38**
	TMDEM	**0.79**	0.23	1.43	4.46	4.39	**8.86**	**0.37**	2.45
	CSRT	0.73	0.00	2.53	87.15	5.52	88.57	0.54	72.78
	DRPN	0.25	0.00	11.08	62.39	16.00	65.49	3.86	51.76
	MIL	0.63	0.00	3.93	28.69	6.68	32.14	0.88	20.11
	DEM	**0.80**	0.14	1.78	11.04	**4.00**	12.96	**0.36**	4.08
	BSP	0.77	0.03	2.26	13.71	4.45	23.46	0.45	7.52
Liver	BASE	0.92	0.80	6.15	16.31	10.16	18.65	0.26	1.38
	KCF	0.94	0.90	4.07	8.04	8.00	13.35	0.16	0.40
	UNET10	**0.97**	**0.93**	**0.80**	**2.16**	**4.96**	**9.69**	**0.05**	**0.16**
	UNET	0.96	0.92	1.33	3.42	6.00	16.03	0.07	0.24
	TM	0.96	**0.93**	1.18	2.97	6.00	10.52	0.08	0.22
	TMDEM	**0.97**	**0.93**	0.98	**2.73**	**5.52**	**9.69**	0.06	**0.18**
	CSRT	0.90	0.60	4.80	23.73	11.46	44.67	0.32	4.62
	DRPN	0.82	0.59	12.22	30.57	21.07	42.20	1.09	5.25
	MIL	0.95	0.90	3.05	8.21	7.88	13.03	0.13	0.41
	DEM	0.96	0.89	1.92	9.21	6.13	14.03	0.08	0.46
	BSP	**0.97**	0.89	**0.96**	8.62	**5.52**	20.84	**0.05**	0.63

Table 7.3: Simulation quantitative results. Algorithms' accuracy in terms of median values and percentile for each metric. The two best performing algorithms in bold.

Region	Method	Dice		Centroid Distance [mm]		Hausdorff distance [mm]		Mean contour distance [mm]	
		Median	$P_{5\%}$	Median	$P_{95\%}$	Median	$P_{95\%}$	Median	$P_{95\%}$
Tumor or surrogate	BASE	0.29	0.00	5.16	24.68	5.72	25.57	1.21	18.00
	KCF	0.41	0.00	3.09	19.07	4.29	20.07	0.99	13.76
	UNET	0.77	0.48	1.24	8.68	3.08	59.70	0.23	5.71
	TM	0.77	0.45	1.08	4.22	2.24	6.41	**0.20**	0.95
	TMDEM	**0.78**	0.45	**0.93**	**3.61**	**2.00**	**5.77**	**0.18**	**0.70**
	CSRT	0.76	**0.58**	1.28	3.96	**2.09**	7.01	0.21	0.79
	DRPN	0.63	0.06	1.85	8.14	3.72	11.87	0.42	5.59
	MIL	0.41	0.00	4.08	30.81	5.67	32.06	1.23	22.70
	DEM	0.68	0.11	1.82	6.06	3.43	14.82	0.37	2.87
	BSP	0.64	0.00	2.02	14.13	3.04	19.85	0.40	5.91
	OBS	**0.84**	**0.63**	**0.95**	**2.91**	3.00	**6.34**	**0.20**	**0.59**
Liver	BASE	0.90	0.42	3.58	25.65	6.71	29.04	0.18	7.08
	KCF	0.91	0.41	3.54	27.98	6.74	30.58	0.16	7.71
	UNET	**0.95**	**0.87**	**0.99**	10.75	4.47	81.80	**0.06**	3.04
	TM	0.93	0.70	1.54	13.32	5.05	16.90	0.10	2.05
	TMDEM	0.94	0.72	1.16	11.95	**4.00**	16.90	0.07	1.72
	CSRT	0.92	0.85	1.39	**3.17**	5.06	**9.25**	0.10	**0.32**
	DRPN	0.88	0.76	3.39	10.19	6.96	**14.38**	0.23	1.04
	MIL	0.92	0.28	1.58	34.56	4.95	37.97	0.11	11.93
	DEM	**0.95**	0.74	**1.00**	12.98	**3.81**	27.51	**0.06**	1.98
	BSP	**0.95**	0.85	1.03	**3.55**	4.08	25.69	**0.06**	0.44
	OBS	**0.96**	**0.91**	1.37	3.99	7.00	17.29	0.09	**0.24**

Table 7.4: Treatment quantitative results for liver patients. Algorithms' accuracy in terms of median values and percentile for each metric. The two best performing algorithms in bold.

Region	Method	Dice		Centroid Distance [mm]		Hausdorff distance [mm]		Mean contour distance [mm]	
		Median	$P_{5\%}$	Median	$P_{95\%}$	Median	$P_{95\%}$	Median	$P_{95\%}$
Tumor	BASE	0.23	0.00	4.07	15.61	4.94	16.61	1.96	8.99
	KCF	0.49	0.00	2.36	14.08	4.24	15.50	0.67	8.00
	UNET	**0.82**	**0.61**	**0.74**	**2.12**	**1.80**	32.43	**0.14**	**0.48**
	TM	0.77	0.42	1.43	4.70	2.50	6.92	0.23	1.29
	TMDEM	0.76	0.48	0.99	4.02	**2.00**	**6.21**	0.19	0.97
	CSRT	0.78	0.45	1.13	4.09	**2.00**	**6.07**	0.18	0.91
	DRPN	0.80	0.00	**0.91**	17.63	1.80	19.43	**0.16**	12.07
	MIL	0.65	0.00	2.12	17.73	3.10	18.55	0.41	12.67
	DEM	0.70	0.00	1.47	8.62	2.69	9.56	0.28	5.29
	BSP	0.74	0.42	1.34	**3.71**	2.62	7.40	0.27	1.03
	OBS	**0.82**	**0.66**	1.28	4.18	3.16	8.52	0.25	**0.84**
Lung	BASE	0.92	0.82	3.68	12.99	10.50	31.36	0.21	1.38
	KCF	0.91	0.82	4.17	12.20	12.63	30.39	0.29	1.29
	UNET	**0.97**	**0.94**	**0.75**	**3.22**	**5.01**	49.84	**0.03**	0.37
	TM	0.92	0.82	2.45	11.20	10.10	29.77	0.18	1.15
	TMDEM	0.95	0.85	1.52	11.84	8.01	52.85	0.07	0.79
	CSRT	0.91	0.80	1.74	10.19	9.03	30.40	0.16	1.24
	DRPN	0.87	0.76	4.39	11.94	11.78	53.86	0.42	1.27
	MIL	0.91	0.82	2.76	7.05	10.03	25.83	0.20	0.92
	DEM	0.96	0.92	1.04	7.48	5.95	28.07	0.05	0.35
	BSP	0.96	**0.94**	**0.85**	**3.31**	**4.92**	**22.69**	**0.04**	**0.25**
	OBS	**0.97**	**0.94**	1.19	3.32	9.43	**17.12**	0.07	**0.16**

Table 7.5: Treatment quantitative results for lung patients. Algorithms' accuracy in terms of median values and percentile for each metric. The two best performing algorithms in bold.

Appendix F. Statistical Analysis

We used the Kolmogorov-Smirnov statistic to compare two samples [Massey Jr, 1951]. The null hypothesis is that two independent samples originate from the same continuous distribution. If the p-values are lower than 0.05, we discard the null hypothesis. We compare UNET10 versus UNET in simulation, interobserver with liver versus lung patients, and the algorithms' accuracy with liver versus lung patients.

UNET10 versus UNET in simulation

		Dice	Centroid Distance	Hausdorff distance	Mean contour distanc
Tumor	Statistic	0.16	0.22	0.22	0.17
	P-value	3.06e-78	7.63e-142	3.43e-144	9.60e-86
Liver	Statistic	0.28	0.18	0.28	0.28
	P-value	3.89e-228	7.38e-90	1.84e-221	1.13e-235

Table 7.6: Comparison of UNET10 and UNET in simulation. Statistic and p-value using Kolmogorov-Smirnov test.

Interobserver liver versus lung patients

	Dice	Centroid Distance	Hausdorff distance	Mean contour distance
Statistic	0.16	0.18	0.17	0.17
P-value	0.10	**0.04**	0.07	0.05

Table 7.7: Interobserver comparison for liver and lung patients. Statistic and p-value using Kolmogorov-Smirnov test. P-values lower than 0.05 in bold.

Algorithms versus themselves with liver and lung patients

Comparing all the algorithms in simulation and treatment (liver and lung patients)

In order to find the statistical differences among the algorithms we run multiple hypothesis tests with Kruskal–Wallis method [Corder and Foreman, 2011], followed by pairwise analysis using Dunn test [Dunn, 1964] with Benjamini–Hochberg correction [Benjamini and Hochberg, 1995].

Regarding the pairwise comparison, since we have 11 independent samples (9 algorithms, BASE and OBS), we show the results as a significance plot or a 11×11 matrix with a heatmap for easy interpretation. The plot depict if the p-values are lower than 0.001, 0.01, 0.05 or greater than 0.05 in dark blue, blue, light blue and red respectively. The results are shown in Figure 7.6, 7.7 and 7.8 for simulation, liver patients and lung patients respectively.

	Dice	Centroid Distance	Hausdorff distance	Mean contour distance
BASE	0.23	0.08	0.12	0.05
KCF	0.08	0.05	0.16	0.05
UNET	0.05	**1.90e-05**	**1.90e-05**	**3.19e-03**
TM	0.51	0.51	0.87	0.4
TMDIR	0.87	0.64	0.51	0.99
CSRT	0.12	0.4	0.16	0.12
DRPN	**2.24e-08**	**1.03e-03**	**5.61e-04**	**5.72e-08**
MIL	**3.43e-07**	**2.99e-04**	**2.99e-04**	**3.93e-05**
DIR	0.23	0.51	0.12	0.31
BSP	**1.83e-03**	**3.19e-03**	0.08	**5.61e-04**
OBS	0.1	**0.04**	0.07	0.05

Table 7.8: Comparison of algorithms' accuracy with themselves for liver and lung patients. P-values using Kolmogorov-Smirnov test. P-values lower than 0.05 in bold.

		Dice	Centroid Distance	Hausdorff distance	Mean contour distance
Tumor	Statistic	1.50e+04	2.41e+04	1.56e+04	1.82e+04
Tumor	P-value	1.80e-78	2.45e-53	8.11e-42	5.60e-72
Liver	Statistic	3.55e+04	3.75e+04	3.07e+04	3.60e+04
Liver	P-value	6.70e-52	6.94e-31	2.33e-32	8.80e-48

Table 7.9: Kruskal–Wallis statistic of simulation. The p-values lower than 0.001 indicate that there are statistical differences among the algorithms.

		Dice	Centroid Distance	Hausdorff distance	Mean contour distance
Tumor*	Statistic	3.35e+02	2.46e+02	1.37e+02	2.53e+02
Tumor*	P-value	7.71e-66	3.83e-47	1.81e-24	1.46e-48
Liver	Statistic	2.44e+02	1.72e+02	7.74e+01	1.36e+02
Liver	P-value	9.18e-47	1.00e-31	1.59e-12	2.32e-24

Table 7.10: Kruskal–Wallis statistic of liver patients. The p-values lower than 0.001 indicate that there are statistical differences among the algorithms. * Tumor or surrogate.

		Dice	Centroid Distance	Hausdorff distance	Mean contour distance
Tumor	Statistic	1.63e+02	1.61e+02	1.29e+02	1.36e+02
Tumor	P-value	6.20e-30	1.87e-29	7.02e-23	2.62e-24
Lung	Statistic	3.70e+02	2.77e+02	1.04e+02	2.94e+02
Lung	P-value	2.25e-73	9.12e-54	1.00e-17	2.43e-57

Table 7.11: Kruskal–Wallis statistic of lung patients. The p-values lower than 0.05 indicate that there are statistical differences among the algorithms.

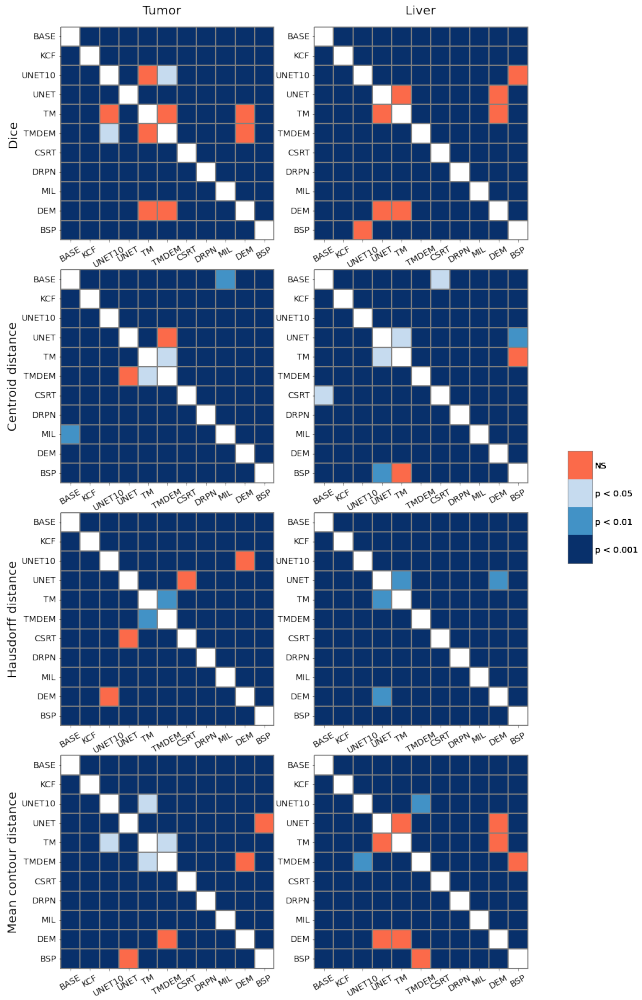


Figure 7.6: Significance plot with pairwise comparisons of algorithms and reference (BASE) for simulation. We depict in red color that there are no statistical significant differences between the samples ($p > 0.05$). Statistical significant differences are shown in three blue colors, defined as: dark blue $p < 0.001$, blue $p < 0.01$, and light blue $p < 0.05$.

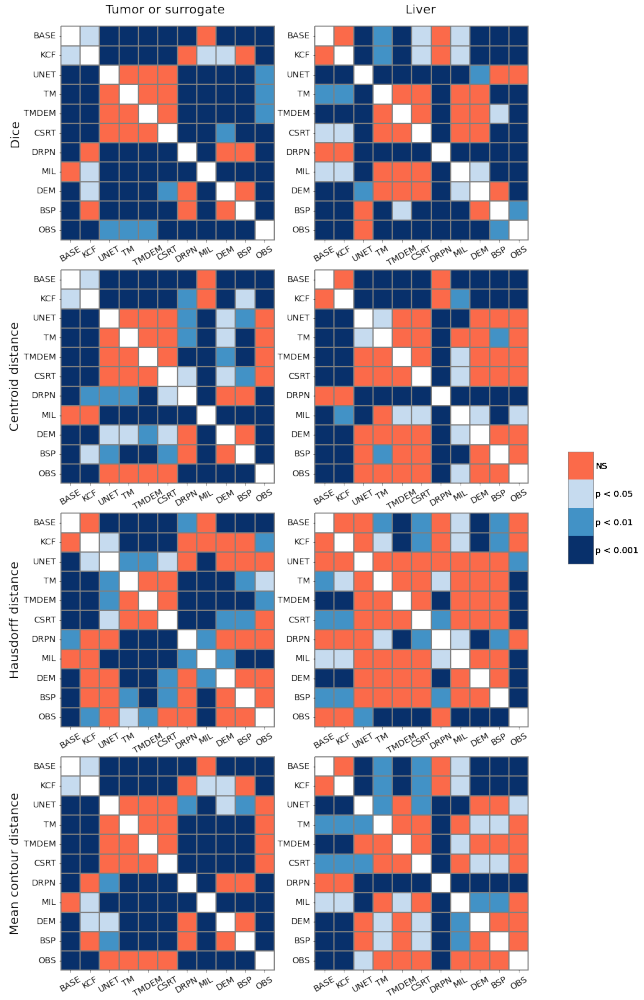


Figure 7.7: Significance plot with pairwise comparisons of algorithms, reference (BASE), and interobservers (OBS) for the treatment liver patients. We depict in red color that there are no statistical significant differences between the samples ($p > 0.05$). Statistical significant differences are shown in three blue colors, defined as: dark blue $p < 0.001$, blue $p < 0.01$, and light blue $p < 0.05$.

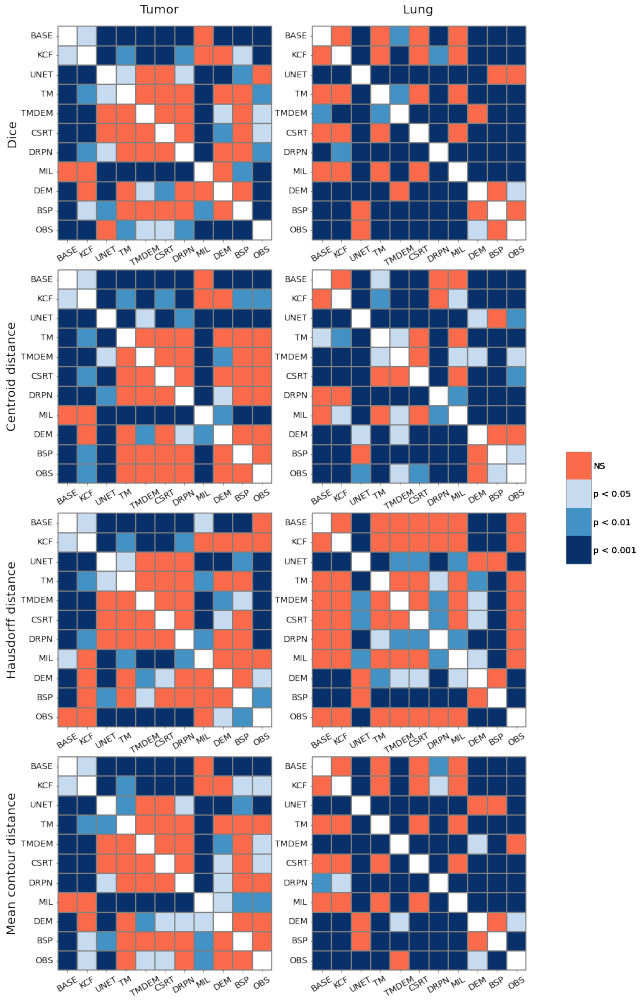


Figure 7.8: Significance plot with pairwise comparisons of algorithms, reference (BASE), and interobservers (OBS) for the treatment lung patients. We depict in red color that there are no statistical significant differences between the samples ($p > 0.05$). Statistical significant differences are shown in three blue colors, defined as: dark blue $p < 0.001$, blue $p < 0.01$, and light blue $p < 0.05$.

Appendix G. Simulation Extended Results

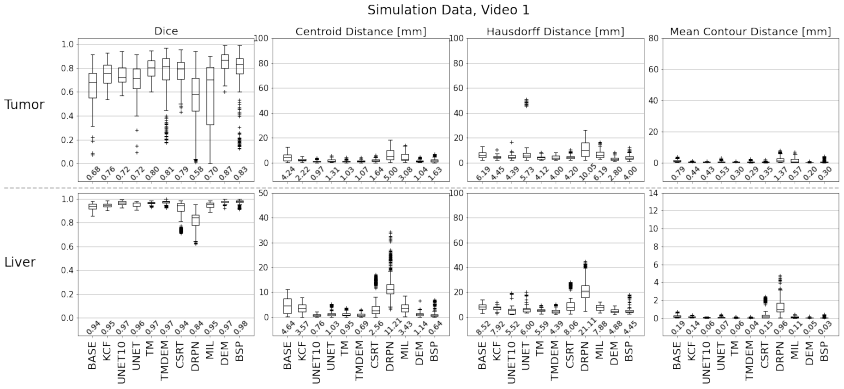


Figure 7.9: Algorithms accuracy with simulation video 1. Simulation of free-breathing motion without varying motion amplitude and without noise.

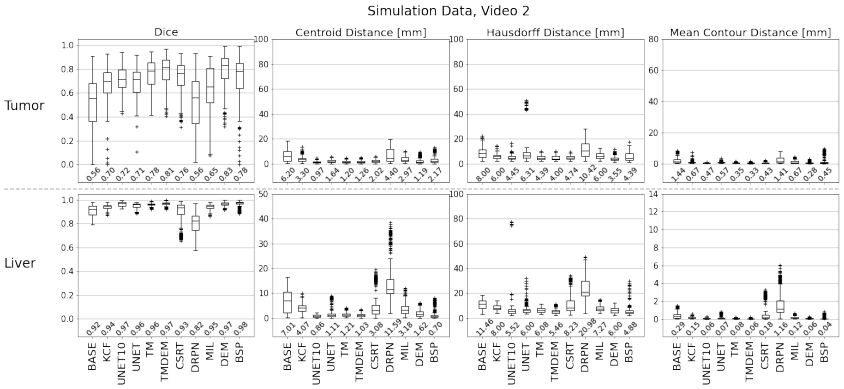


Figure 7.10: Algorithms accuracy with simulation video 2. Simulation of free-breathing motion with motion amplitude 1.5 times and without noise.

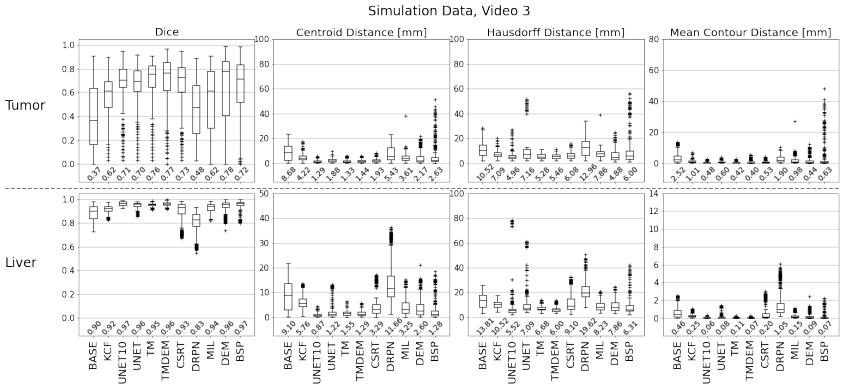


Figure 7.11: Algorithms accuracy with simulation video 3. Simulation of free-breathing motion with motion amplitude 2.0 times and without noise.

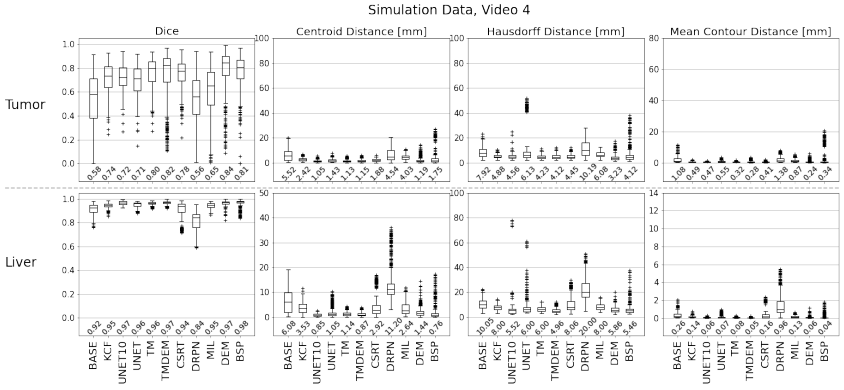


Figure 7.12: Algorithms accuracy with simulation video 4. Simulation of free-breathing motion with random motion amplitude (range from 1.0 to 2.0 per breathing cycle) and without noise.

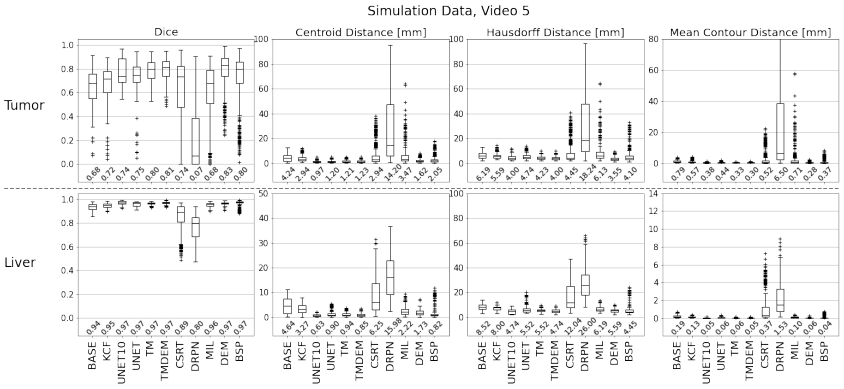


Figure 7.13: Algorithms accuracy with simulation video 5. Simulation of free-breathing motion without varying motion amplitude and adding Gaussian noise.

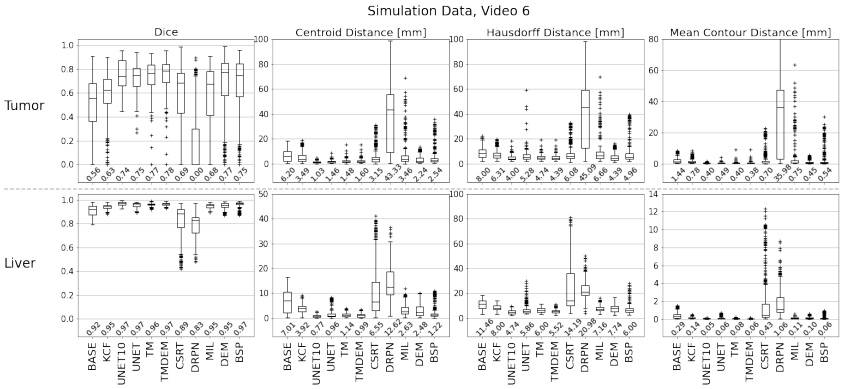


Figure 7.14: Algorithms accuracy with simulation video 6. Simulation of free-breathing motion with motion amplitude 1.5 times and adding Gaussian noise.

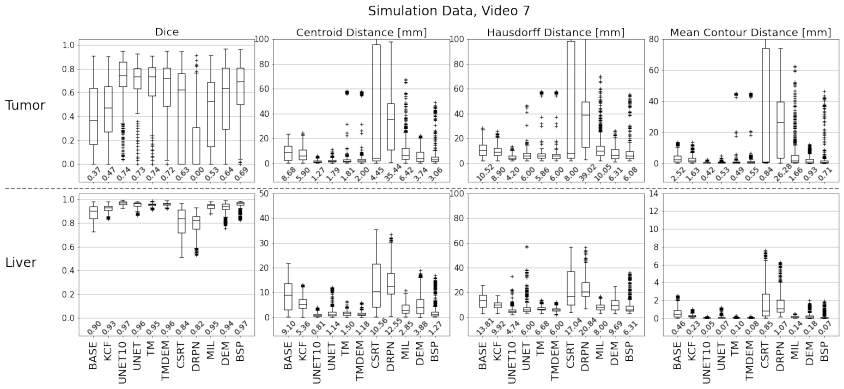


Figure 7.15: Algorithms accuracy with simulation video 7. Simulation of free-breathing motion with motion amplitude 2.0 times and adding Gaussian noise.

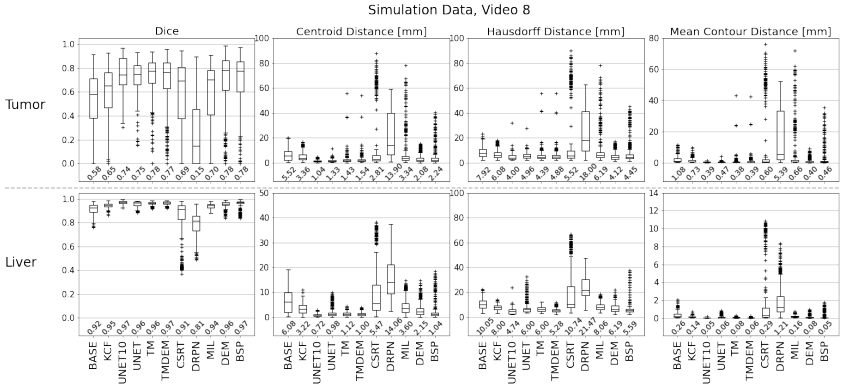


Figure 7.16: Algorithms accuracy with simulation video 8. Simulation of free-breathing motion with random motion amplitude (range from 1.0 to 2.0 per breathing cycle) and adding Gaussian noise.

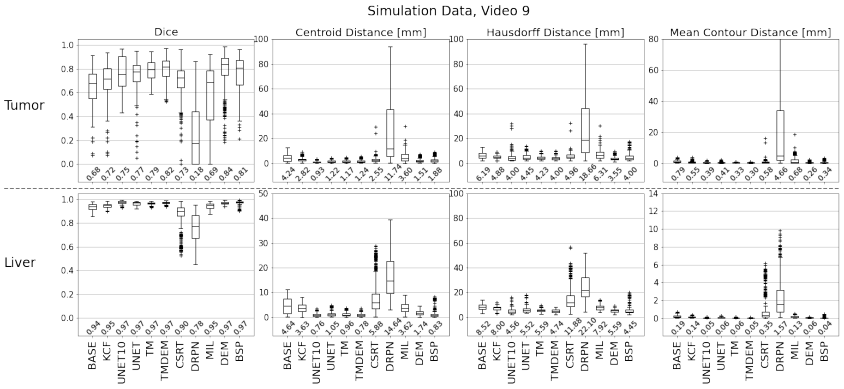


Figure 7.17: Algorithms accuracy with simulation video 9. Simulation of free-breathing motion without varying motion amplitude and adding Rician noise.

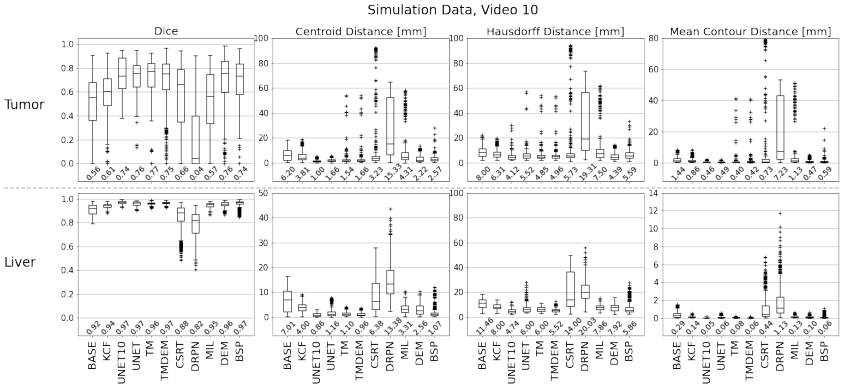


Figure 7.18: Algorithms accuracy with simulation video 10. Simulation of free-breathing motion with motion amplitude 1.5 times and adding Rician noise.

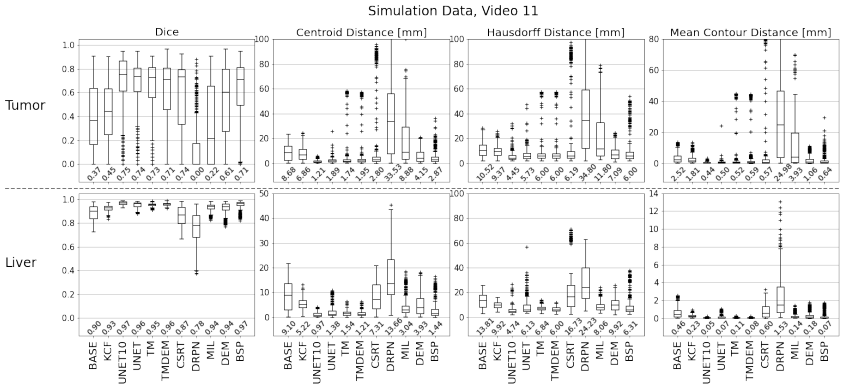


Figure 7.19: Algorithms accuracy with simulation video 11. Simulation of free-breathing motion with motion amplitude 2.0 times and adding Rician noise.

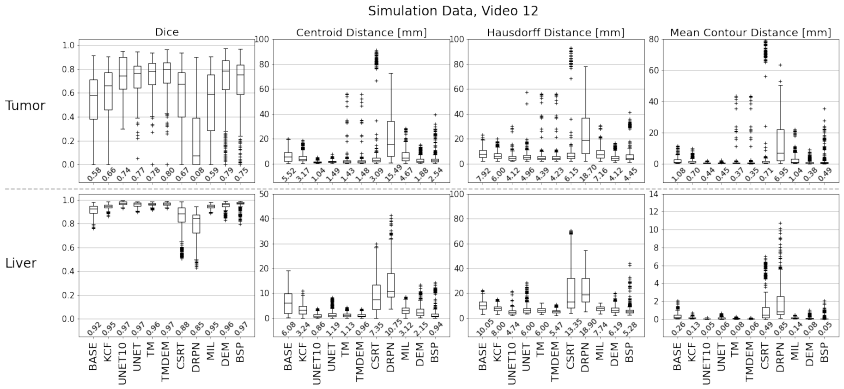


Figure 7.20: Algorithms accuracy with simulation video 12. Simulation of free-breathing motion with random motion amplitude (range from 1.0 to 2.0 per breathing cycle) and adding Rician noise.

Chapter 8

Summary and Discussion

8.1 Summary

This thesis presented six novel aspects related to real-time deformable image registration and tumor tracking.

A high performance library for medical image registration (Chapter 4). The library includes a real-time deformable image registration algorithm that is reported to be the first DIR algorithm at a full imaging scale to accomplish the real-time performance under the constraint of four frames per second for tumor tracking. The algorithm is proven and tested over multiple conditions and scenarios in radiotherapy (Chapter 5, 7, 6). The code is open source and publicly available at `https://github.com/josetascon/imart`.

A pipeline for simulation of 2D Cine-MRI with the underlined delineations of organs and tumors (Chapter 5). The advantage of using synthetic Cine-MRIs is that we generate a simulated sequence with a ground truth segmentation and control the breathing motion. Unfortunately, medical imaging data is limited, and ground truth is manually intensive. However, pre-treatment images are typically available, and the proposed system takes full advantage of this. The code is open source and publicly available at `https://github.com/josetascon/cinemri-simulation`.

A comprehensive evaluation of real-time tumor tracking algorithms (Chapter 7). The evaluation incorporates the most challenging conditions for tumor tracking with simulation and treatment Cine-MRIs. We summarized the evaluation conditions of our study compared to previous work in Table 7.1. The simulation allows variable motion and added noise. The treatment data was acquired during motion after breath-holding and with a low magnetic field MR device (0.35T), i.e., a lower signal-to-noise ratio. Additionally, our study includes only automated tumor tracking algorithms that use a single reference image as the normal condition in clinical practice. In contrast, previous studies [Fast et al., 2017, Friedrich et al., 2021] use ten templates images that suppose longer waiting times for patients on the couch during treatment. Finally,

we included state-of-the-art medical imaging and computer vision algorithms in our comparison. All the evaluated algorithms' code is publicly available at https://github.com/josetascon/tracking-tissue.

A UNET algorithm trained with a single image. Although UNET is a well-established algorithm in medical imaging [Ronneberger et al., 2015], this approach to train with a single image using augmentations is novel. We used the simulation dataset to compare the accuracy with UNET10, a conventional UNET trained with ten images (similar to [Friedrich et al., 2021]). The results showed that UNET is relatively close to UNET10 in terms of median values accuracy ($\text{Dice}_{P_{5\%}}$ 0.52 and 0.5 respectively). The proposed training strategy fulfills the automation requirement of the clinical setting without a drastic loss of accuracy.

A novel algorithm for tumor tracking named TMDEM. We proposed an algorithm based on template matching for global tracking combined with demons deformable image registration for local refinement. The template matching algorithm had acceptable accuracy for tumor tracking, and therefore, we aim to improve it to support out-of-plane deformations during breathing motion. The result is one of the most accurate algorithms for tumor tracking in Chapter 7. The average computational time is 18.8 ms (53.3 fps), corresponding to the nine algorithms' fourth place and an adequate frame rate for real-time tracking in radiotherapy ($> 4fps$).

The first evaluation of multiple-organ tracking (Chapter 6). Previous studies in radiotherapy with MR-Linacs have focused only on tumor tracking [Cervino et al., 2011, Gou et al., 2014, Shi et al., 2014, Paganelli et al., 2015, Yun et al., 2015, Zachiu et al., 2015, Fast et al., 2017, Seregni et al., 2018, Garau et al., 2019, Friedrich et al., 2021]. To the best of our knowledge, our study is the first implementation and evaluation of tracking with multiple organs in real-time. We exploit multiple tracking to proposed gating control signals useful for image-guided radiotherapy and beam control. The signals are very convenient to improve dose delivery and reduce organs-at-risk for tumors located close to the edge of an organ (e.g., liver or lung tumors close to the diaphragm).

8.2 Discussion and Conclusion

This thesis focused on achieving real-time deformable image registration. For this purpose, we selected tumor tracking in Cine-MRI as the application. The comprehensive review of medical image registration in Chapter 2 determined our goal to improve the performance of an intensity-based DIR with a variational approach.

A clear advantage of a variational approach is that it generalizes better for different fields of view. For example, if a patient has a condition where the tracking is more convenient to perform in the coronal plane (e.g. [Tryggestad

et al., 2013, Gou et al., 2014, Fast et al., 2017, Garau et al., 2019]), the DIR variational approach is guaranteed to work. In contrast, if the deep learning approach is trained only with sagittal images, it is not guaranteed that the algorithm would perform well.

We did not evaluate deep learning algorithms for real-time deformable image registration. Considering the tracking application that we wanted to solve, the deep learning approach would require a significant amount of data unavailable in our project. Furthermore, there is a high risk of falling due to a lack of generalization of the deformable transformations. One reason is that the tumor shapes may vary significantly from patient to patient. Another reason is that the field of view in the thorax of a training dataset will not be sufficient to cover multiple tracking locations.

The DaSiamRPN (DRPN) algorithm [Li et al., 2018] evaluated in Chapter 7 is the best example of why deep learning approaches fail to generalize. The algorithm was designed as a general object tracking algorithm and trained with thousands of images of everyday indoor and outdoor objects. Nevertheless, we demonstrated that an object tracking algorithm with a variational approach like CSRT [Lukezic et al., 2017] generalizes better after being tested with organs. This situation supports our decision to improve the performance of variational approaches for DIR rather than selecting a deep learning approach.

A difficulty with an intensity-based DIR algorithm is that the accuracy in low contrast organs such as the liver has been sparsely studied. The finite element model-based DIR proposed by Brock et al. [Brock et al., 2005] is the only algorithm to be extensively evaluated in CT with low contrast organs [Brock et al., 2006, Velec et al., 2012, Velec et al., 2017]. The MIDRAS study [Brock et al., 2010] compared several algorithms for the abdomen and the liver with only one 4DCT scan. Finite element model-based DIR methods seem to perform slightly better than intensity-based algorithms for the liver. The drawback is that one patient is not representative of the general performance of the algorithms. An advantage of intensity-based DIR is that it computes registration without prior information or preprocessing. In contrast to the finite element model-based DIR that requires organ segmentation, surface mesh conversion, boundary conditions, and manual parametrization of material properties [Brock et al., 2005]. In Chapter 3 we demonstrated that an intensity-based DIR algorithm is accurate within the desired resolution for radiotherapy [Brock et al., 2017]. These results encouraged our real-time implementation of an intensity-based algorithm.

We initiated the tumor tracking application, considering how to evaluate the algorithms. Thus, we formulated a study that includes simulation and treatment data. We designed a platform and a methodology to easily evaluate tracking algorithms on synthetic Cine-MRI with simulated ground truth segmentations. The video simulator does not require any training data and works only with pre-treatment images. A limitation of our Cine-MRI simulator is that the breathing model uses a single respiratory cycle from the 4DCT

scan. However, the breathing model overcomes this by composing transformations that are time interpolated. Time interpolation guarantees that different patterns arise due to asynchrony between the patient's breathing cycle and sampling times. Furthermore, since we modeled with full 3D images and then created the 2D Saggital MR, we incorporated the desired out-of-plane motion, which is the main challenge for tracking algorithms. Overall, our goal is not to create a perfect breathing model but to facilitate challenging experiments to evaluate tumor tracking algorithms.

Regarding the treatment data used in Chapter 7, we included five liver and five lung patients. We used three Cine-MRI sequences per patient (3 fractions) for a total of 30 Cine-MRI sequences. In addition, two imaging experts performed manual delineations of the organs and the tumor using the same protocol. The dataset used in Chapter 6 is a subset of this dataset having two liver and two lung patients with a field of view that cover the lung and the liver.

Concerning the whole treatment data, we had two liver patients where the liver tumor was not visually identifiable for all the frames. A previous study has also used a surrogate in the livers to evaluate tumor tracking [Paganelli et al., 2015]. Furthermore, the observers selected random frames for ground truth delineations. Random selection reduces the load in the delineation process. A previous study has also selected frames randomly to evaluate tracking [Mazur et al., 2016]. To reduce the impact of these drawbacks in the treatment data, we analyzed and contrasted the algorithms between the simulation data and the treatment data and found consistencies in the performance results.

Our studies (Chapter 6,7) only considered Cine-MRIs acquired in the sagittal plane for two reasons. First, breathing movement and, consequently, tumor motion mainly occurs in two axes, Anterior-Posterior and Superior-Inferior [Seppenwoolde et al., 2002, Kitamura et al., 2003]. Second, most of the patients treated in MR-Linacs occur in the sagittal plane [Paganelli et al., 2018] and is the favored orientation in the literature to capture the significant motion directions.

In terms of tumor tracking algorithms, we offered a comprehensive evaluation with the most significant number of algorithms (see Table 7.1). This is a noteworthy contribution due to the lack of publicly available datasets or benchmarks for tumor tracking with Cine-MRIs. Furthermore, the evaluated algorithms themselves introduce some novelty. For instance, we are the first to evaluate tumor tracking on Cine-MRIs with a collection of object tracking algorithms from the computer vision field. Moreover, our developed and proposed high-performing DIR algorithms obtain notable results.

To the best of our knowledge, our high-performing demons algorithm (DEM) is the first real-time deformable registration method at full imaging scale to achieve less than 250 ms of performance, i.e., 112.7 ms (8.9 fps). We implemented a B-spline algorithm (BSP) that performs at 746.8 ms (1.3 fps).

In comparison, the BSP method presented by Fast et al. [Fast et al., 2017] ran at 500 ms (2 fps), and the method presented by Friedrich et al. [Friedrich et al., 2021] ran at 1300 ms (0.77 fps). New MR-Linacs are evolving with higher acquisitions rates [Paganelli et al., 2018], and therefore, faster and robust tracking algorithms are relevant. In this sense, our novel TMDEM algorithm achieved an even better computational time of 18.8 ms (53.3 fps) with adequate accuracy.

The template matching (TM) and TMDEM algorithms perform consistently well for simulation and treatment data. TMDEM represents an improvement to TM due to the extra step to deform the local area. The computational time does not increase significantly because the registration is made on a small area with a reduced optimization setup compared to the full-scale DEM algorithm.

Regarding the tumor tracking with simulation data, we observed better performance of DIR algorithms (DEM and BSP). DIR algorithms are more suitable for noisy conditions. The reason is that regularization maintains consistency at the global level. The drawback is that the limited optimization steps to achieve real-time become a constraint for significant respiratory motion. The TM, TMDEM, and UNET algorithms show slightly lower accuracy than DEM and BSP. However, these four algorithms perform better with the variable motion amplitudes.

Regarding tumor tracking on the treatment data, we identify that UNET, TMDEM, and CSRT algorithms perform consistently better. However, a disadvantage of the UNET algorithm is the training time. The required time of training for in-room treatment could be excessive. In this sense, an alternative is our novel TMDEM, a ready-to-use algorithm for tumor tracking with adequate accuracy and time performance.

In conclusion, we successfully designed, implemented, and tested a real-time deformable image registration algorithm. We proved and applied the algorithm to the tumor tracking problem in the context of image-guided radiotherapy. We covered multiple and challenging conditions for a proper evaluation and compared several algorithms. A DIR algorithm for tumor and organ tracking has acceptable results in terms of accuracy for radiotherapy. We proposed a novel TMDEM algorithm that combines template matching and deformable registration algorithms to balance performance and accuracy. The results indicate that this algorithm is a prime choice for tumor tracking. All the code produced in this thesis is publicly available to ensure its usage and further deployment to other applications.

8.3 Future Prospects

The natural extension of our work is to deploy the high-performance DIR algorithm, the proposed tracking algorithms, or the developed framework to

other applications. For instance, real-time tracking with ultrasound images [Cifor et al., 2013, De Luca et al., 2015, De Luca et al., 2018] or localization in image-guided surgical procedures [Rivaz et al., 2014, Reaungamornrat et al., 2016, Alam et al., 2018] are suitable applications.

The performance of DIR can be further improved with a sampling strategy during optimization. For instance, ITK ([ITK, 2021]) and some derived frameworks (e.g. [ANTs, 2021, Elastix, 2021]) support two sampling strategies: regular and random sampling. The regular sampling becomes convenient to reduce the time but may create convergence problems to local minima. A random sampling strategy enables the use of a stochastic gradient descent [Klein et al., 2009b]. The disadvantage is that this optimization strategy requires more iterations to converge. Nonetheless, a proper study of convergence and sampling strategy similar to [Klein et al., 2007] should be performed.

The design of our IMART library allows us to test the high-performing algorithms with alternative devices. This versatility comes from the OpenCL and OpenMP support. Other devices such as FPGAs represent an excellent alternative for testing [Waidyasooriya et al., 2018]. Furthermore, an extension to support multiple processing devices (multi-node) simultaneously is attractive. The distributed memory design adopted by CLAIRE is a good example and the groundwork of how to extend a high-performance DIR algorithm to multiple devices [Mang et al., 2019].

A missing component of our work is to evaluate the performance of a deep learning algorithm, e.g., [Yang et al., 2017b, Balakrishnan et al., 2019, de Vos et al., 2019]. As mentioned earlier, generalization should be considered as the main driving factor in the design. Our proposed simulator or augmentation with Generative Adversarial Networks ([Kazeminia et al., 2020]) can increase and improve the training dataset for this purpose.

Tumor tracking with a variational approach offered adequate performance and accuracy. A missing component is an object predictor based on positions and the current velocity. A filter approach can incorporate a motion model and integrate the current measurement [Chui et al., 2017]. The advantage of breathing motion is that it has a cyclic movement pattern. In other tracking problems, the movement has to be considered all over the field of view. Including this constraint in the tracking strategy can generate more accurate algorithms.

Bibliography

[Aitken and Hawkins, 2015] Aitken, K. and Hawkins, M. (2015). Stereotactic body radiotherapy for liver metastases. *Clinical oncology*, 27(5):307–315.

[Al-Ward et al., 2018] Al-Ward, S., Wronski, M., Ahmad, S. B., Myrehaug, S., Chu, W., Sahgal, A., and Keller, B. M. (2018). The radiobiological impact of motion tracking of liver, pancreas and kidney sbrt tumors in a mr-linac. *Physics in Medicine & Biology*, 63(21):215022.

[Alam et al., 2018] Alam, F., Rahman, S. U., Ullah, S., and Gulati, K. (2018). Medical image registration in image guided surgery: Issues, challenges and research opportunities. *Biocybernetics and Biomedical Engineering*, 38(1):71–89.

[ANTs, 2021] ANTs (2021). Advanced normalization tools. Web Page: http://stnava.github.io/ANTs/. [Consulted in June 2021].

[Arsigny et al., 2006] Arsigny, V., Commowick, O., Pennec, X., and Ayache, N. (2006). A log-euclidean framework for statistics on diffeomorphisms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 924–931. Springer.

[Ashburner, 2007] Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113.

[Ashburner and Miller, 2015] Ashburner, J. and Miller, M. (2015). Diffeomorphic image registration. Elsevier.

[Avants et al., 2008] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41.

[Babenko et al., 2009] Babenko, B., Yang, M.-H., and Belongie, S. (2009). Visual tracking with online multiple instance learning. In *2009 IEEE Conference on computer vision and Pattern Recognition*, pages 983–990. IEEE.

[Balakrishnan et al., 2019] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2019). Voxelmorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800.

[Beddar et al., 2008] Beddar, A. S., Briere, T. M., Balter, P., Pan, T., Tolani, N., Ng, C., Szklaruk, J., and Krishnan, S. (2008). 4d-ct imaging with synchronized intravenous contrast injection to improve delineation of liver tumors for treatment planning. *Radiotherapy and Oncology*, 87(3):445–448.

[Beddar et al., 2007] Beddar, A. S., Kainz, K., Briere, T. M., Tsunashima, Y., Pan, T., Prado, K., Mohan, R., Gillin, M., and Krishnan, S. (2007). Correlation between internal fiducial tumor motion and external marker motion for liver tumors imaged with 4d-ct. *International Journal of Radiation Oncology* Biology* Physics*, 67(2):630–638.

[Beg et al., 2005] Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157.

[Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

[Bertholet et al., 2016] Bertholet, J., Worm, E. S., Fledelius, W., Høyer, M., and Poulsen, P. R. (2016). Time-resolved intrafraction target translations and rotations during stereotactic liver radiation therapy: implications for marker-based localization accuracy. *International Journal of Radiation Oncology* Biology* Physics*, 95(2):802–809.

[Bertinetto et al., 2016] Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., and Torr, P. H. (2016). Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer.

[BITE, 2021] BITE (2021). Brain images of tumors for evaluation database. Web Page: http://nist.mni.mcgill.ca/?page_id=672. [Consulted in June 2021].

[Boldea et al., 2008] Boldea, V., Sharp, G. C., Jiang, S. B., and Sarrut, D. (2008). 4d-ct lung motion estimation with deformable registration: quantification of motion nonlinearity and hysteresis. *Medical physics*, 35(3):1008–1018.

[Brock et al., 2005] Brock, K., Sharpe, M., Dawson, L., Kim, S., and Jaffray, D. (2005). Accuracy of finite element model-based multi-organ deformable image registration. *Medical physics*, 32(6Part1):1647–1659.

[Brock et al., 2010] Brock, K. K., Consortium, D. R. A., et al. (2010). Results of a multi-institution deformable registration accuracy study (midras). *International Journal of Radiation Oncology* Biology* Physics*, 76(2):583–596.

[Brock et al., 2006] Brock, K. K., Dawson, L. A., Sharpe, M. B., Moseley, D. J., and Jaffray, D. A. (2006). Feasibility of a novel deformable image registration technique to facilitate classification, targeting, and monitoring of tumor and normal tissue. *International Journal of Radiation Oncology* Biology* Physics*, 64(4):1245–1254.

[Brock et al., 2017] Brock, K. K., Mutic, S., McNutt, T. R., Li, H., and Kessler, M. L. (2017). Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the aapm radiation therapy committee task group no. 132. *Medical physics*, 44(7):e43–e76.

[Brook et al., 2012] Brook, O. R., Gourtsoyianni, S., Brook, A., Mahadevan, A., Wilcox, C., and Raptopoulos, V. (2012). Spectral ct with metal artifacts reduction software for improvement of tumor visibility in the vicinity of gold fiducial markers. *Radiology*, 263(3):696–705.

[Brunn et al., 2021] Brunn, M., Himthani, N., Biros, G., Mehl, M., and Mang, A. (2021). Fast gpu 3d diffeomorphic image registration. *Journal of Parallel and Distributed Computing*, 149:149–162.

[Caillet et al., 2017] Caillet, V., Booth, J. T., and Keall, P. (2017). Igrt and motion management during lung sbrt delivery. *Physica Medica*, 44:113–122.

[Castillo et al., 2009a] Castillo, E., Castillo, R., Martinez, J., Shenoy, M., and Guerrero, T. (2009a). Four-dimensional deformable image registration using trajectory modeling. *Physics in Medicine & Biology*, 55(1):305.

[Castillo et al., 2013] Castillo, R., Castillo, E., Fuentes, D., Ahmad, M., Wood, A. M., Ludwig, M. S., and Guerrero, T. (2013). A reference dataset for deformable image registration spatial accuracy evaluation using the copdgene study archive. *Physics in Medicine & Biology*, 58(9):2861.

[Castillo et al., 2009b] Castillo, R., Castillo, E., Guerra, R., Johnson, V. E., McPhail, T., Garg, A. K., and Guerrero, T. (2009b). A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine & Biology*, 54(7):1849.

[Cervino et al., 2011] Cervino, L. I., Du, J., and Jiang, S. B. (2011). Mri-guided tumor tracking in lung cancer radiotherapy. *Physics in Medicine & Biology*, 56(13):3773.

[Cheng et al., 2018] Cheng, X., Zhang, L., and Zheng, Y. (2018). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):248–252.

[Christensen and Johnson, 2001] Christensen, G. E. and Johnson, H. J. (2001). Consistent image registration. *IEEE transactions on medical imaging*, 20(7):568–582.

[Chui et al., 2017] Chui, C. K., Chen, G., et al. (2017). *Kalman filtering*. Springer.

[Ciaparrone et al., 2020] Ciaparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R., and Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381:61–88.

[Cifor et al., 2013] Cifor, A., Risser, L., Chung, D., Anderson, E. M., and Schnabel, J. A. (2013). Hybrid feature-based diffeomorphic registration for tumor tracking in 2-d liver ultrasound images. *IEEE transactions on medical imaging*, 32(9):1647–1656.

[CLAIRE, 2021] CLAIRE (2021). Constrained large deformation diffeomorphic image registration. Web Page: https://github.com/andreasmang/claire. [Consulted in June 2021].

[Corder and Foreman, 2011] Corder, G. W. and Foreman, D. I. (2011). Non-parametric statistics for non-statisticians.

[Crijns et al., 2011] Crijns, S., Kok, J., Lagendijk, J., and Raaymakers, B. (2011). Towards mri-guided linear accelerator control: gating on an mri accelerator. *Physics in Medicine & Biology*, 56(15):4815.

[Crum et al., 2006] Crum, W. R., Camara, O., and Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging*, 25(11):1451–1461.

[Crum et al., 2004] Crum, W. R., Hartkens, T., and Hill, D. (2004). Non-rigid image registration: theory and practice. *The British journal of radiology*, 77(suppl_2):S140–S153.

[De Luca et al., 2018] De Luca, V., Banerjee, J., Hallack, A., Kondo, S., Makhinya, M., Nouri, D., Royer, L., Cifor, A., Dardenne, G., Goksel, O., et al. (2018). Evaluation of 2d and 3d ultrasound tracking algorithms and impact on ultrasound-guided liver radiotherapy margins. *Medical physics*, 45(11):4986–5003.

[De Luca et al., 2015] De Luca, V., Benz, T., Kondo, S., König, L., Lübke, D., Rothlübbers, S., Somphone, O., Allaire, S., Bell, M. L., Chung, D., et al. (2015). The 2014 liver ultrasound tracking benchmark. *Physics in Medicine & Biology*, 60(14):5571.

[de Senneville et al., 2016] de Senneville, B. D., Zachiu, C., Ries, M., and Moonen, C. (2016). Evolution: an edge-based variational method for non-rigid multi-modal image registration. *Physics in Medicine & Biology*, 61(20):7377.

[de Vos et al., 2019] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Medical image analysis*, 52:128–143.

[Dennis and Moré, 1977] Dennis, Jr, J. E. and Moré, J. J. (1977). Quasi-newton methods, motivation and theory. *SIAM review*, 19(1):46–89.

[Dhont et al., 2020] Dhont, J., Harden, S., Chee, L., Aitken, K., Hanna, G., and Bertholet, J. (2020). Image-guided radiotherapy to manage respiratory motion: lung and liver. *Clinical oncology*, 32(12):792–804.

[dir-lab, 2021] dir-lab (2021). The deformable image registration laboratory. Web Page: https://www.dir-lab.com/. [Consulted in June 2021].

[Dunn, 1964] Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252.

[Eklund et al., 2013] Eklund, A., Dufort, P., Forsberg, D., and LaConte, S. M. (2013). Medical image processing on the gpu–past, present and future. *Medical image analysis*, 17(8):1073–1094.

[Ekström et al., 2021] Ekström, S., Pilia, M., Kullberg, J., Ahlström, H., Strand, R., and Malmberg, F. (2021). Faster dense deformable image registration by utilizing both cpu and gpu. *Journal of Medical Imaging*, 8(1):014002.

[Elastix, 2021] Elastix (2021). A toolbox for rigid and nonrigid registration of images. Web Page: http://elastix.isi.uu.nl/. [Consulted in June 2021].

[Farnebäck, 2003] Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer.

[Fast et al., 2017] Fast, M. F., Eiben, B., Menten, M. J., Wetscherek, A., Hawkes, D. J., McClelland, J. R., and Oelfke, U. (2017). Tumour auto-contouring on 2d cine mri for locally advanced lung cancer: A comparative study. *Radiotherapy and Oncology*, 125(3):485–491.

[Feldman et al., 2019] Feldman, A. M., Modh, A., Glide-Hurst, C., Chetty, I. J., and Movsas, B. (2019). Real-time magnetic resonance-guided liver stereotactic body radiation therapy: an institutional report using a magnetic resonance-linac system. *Cureus*, 11(9).

[Fluck et al., 2011] Fluck, O., Vetter, C., Wein, W., Kamen, A., Preim, B., and Westermann, R. (2011). A survey of medical image registration on graphics hardware. *Computer methods and programs in biomedicine*, 104(3):e45–e57.

[Friedrich et al., 2021] Friedrich, F., Hörner-Rieber, J., Renkamp, C. K., Klüter, S., Bachert, P., Ladd, M. E., and Knowles, B. R. (2021). Stability of conventional and machine learning-based tumor auto-segmentation techniques using undersampled dynamic radial bssfp acquisitions on a 0.35 t hybrid mr-linac system. *Medical Physics*, 48(2):587–596.

[Fu et al., 2019] Fu, Y., Wu, X., Thomas, A. M., Li, H. H., and Yang, D. (2019). Automatic large quantity landmark pairs detection in 4dct lung images. *Medical physics*, 46(10):4490–4501.

[Garau et al., 2019] Garau, N., Via, R., Meschini, G., Lee, D., Keall, P., Riboldi, M., Baroni, G., and Paganelli, C. (2019). A roi-based global motion model established on 4dct and 2d cine-mri data for mri-guidance in radiation therapy. *Physics in Medicine & Biology*, 64(4):045002.

[Goerres et al., 2002] Goerres, G. W., Kamel, E., Heidelberg, T.-N. H., Schwitter, M. R., Burger, C., and von Schulthess, G. K. (2002). Pet-ct image co-registration in the thorax: influence of respiration. *European journal of nuclear medicine and molecular imaging*, 29(3):351–360.

[Google/Benchmark, 2021] Google/Benchmark (2021). A microbenchmark support library. Web Page: https://github.com/google/benchmark. [Consulted in June 2021].

[Goshtasby, 2005] Goshtasby, A. A. (2005). *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons.

[Gou et al., 2014] Gou, S., Wu, J., Liu, F., Lee, P., Rapacchi, S., Hu, P., and Sheng, K. (2014). Feasibility of automated pancreas segmentation based on dynamic mri. *The British journal of radiology*, 87(1044):20140248.

[Gu et al., 2009] Gu, X., Pan, H., Liang, Y., Castillo, R., Yang, D., Choi, D., Castillo, E., Majumdar, A., Guerrero, T., and Jiang, S. B. (2009). Implementation and evaluation of various demons deformable image registration algorithms on a gpu. *Physics in Medicine & Biology*, 55(1):207.

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

[Guven et al., 2016] Guven, O., Eftekhar, A., Kindt, W., and Constandinou, T. G. (2016). Computationally efficient real-time interpolation algorithm for non-uniform sampled biosignals. *Healthcare technology letters*, 3(2):105–110.

[Haber and Modersitzki, 2006] Haber, E. and Modersitzki, J. (2006). Intensity gradient based registration and fusion of multi-modal images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 726–733. Springer.

[Hadamard, 1923] Hadamard, J. (1923). *Lectures on Cauchy's problem in linear partial differential equations*. Yale University Press.

[Hajnal and Hill, 2001] Hajnal, J. V. and Hill, D. L. (2001). *Medical image registration*. CRC press.

[Hardcastle et al., 2013] Hardcastle, N., Van Elmpt, W., De Ruysscher, D., Bzdusek, K., and Tomé, W. A. (2013). Accuracy of deformable image registration for contour propagation in adaptive lung radiotherapy. *Radiation Oncology*, 8(1):243.

[Harris et al., 2018] Harris, W., Wang, C., Yin, F.-F., Cai, J., and Ren, L. (2018). A novel method to generate on-board 4d mri using prior 4d mri and on-board kv projections from a conventional linac for target localization in liver sbrt. *Medical physics*, 45(7):3238–3245.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*, volume 2. Cambridge university press Cambridge, UK:.

[Haskins et al., 2020] Haskins, G., Kruger, U., and Yan, P. (2020). Deep learning in medical image registration: a survey. *Machine Vision and Applications*, 31(1):1–18.

[Hellier et al., 2003] Hellier, P., Barillot, C., Corouge, I., Gibaud, B., Le Goualher, G., Collins, D. L., Evans, A., Malandain, G., Ayache, N., Christensen, G. E., et al. (2003). Retrospective evaluation of intersubject brain registration. *IEEE transactions on medical imaging*, 22(9):1120–1130.

[Henriques et al., 2012] Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. (2012). Exploiting the circulant structure of tracking-by-detection with kernels. In *European conference on computer vision*, pages 702–715. Springer.

[Holden, 2008] Holden, M. (2008). A review of geometric transformations for nonrigid body registration. *IEEE transactions on medical imaging*, 27(1):111.

[Horn and Schunck, 1981] Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. In *Techniques and Applications of Image Understanding*, volume 281, pages 319–331. International Society for Optics and Photonics.

[Høyer et al., 2012] Høyer, M., Swaminath, A., Bydder, S., Lock, M., Romero, A. M., Kavanagh, B., Goodman, K. A., Okunieff, P., and Dawson, L. A. (2012). Radiotherapy for liver metastases: a review of evidence. *International Journal of Radiation Oncology* Biology* Physics*, 82(3):1047–1057.

[Hua et al., 2017] Hua, R., Pozo, J. M., Taylor, Z. A., and Frangi, A. F. (2017). Multiresolution extended free-form deformations (xffd) for non-rigid registration with discontinuous transforms. *Medical image analysis*, 36:113–122.

[ITK, 2021] ITK (2021). Insight segmentation and registration toolkit. Web Page: https://itk.org/. [Consulted in June 2021].

[Kazeminia et al., 2020] Kazeminia, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., and Mukhopadhyay, A. (2020). Gans for medical image analysis. *Artificial Intelligence in Medicine*, page 101938.

[Keall et al., 2006] Keall, P. J., Mageras, G. S., Balter, J. M., Emery, R. S., Forster, K. M., Jiang, S. B., Kapatoes, J. M., Low, D. A., Murphy, M. J., Murray, B. R., et al. (2006). The management of respiratory motion in radiation oncology report of aapm task group 76 a. *Medical physics*, 33(10):3874–3900.

[Kitamura et al., 2003] Kitamura, K., Shirato, H., Seppenwoolde, Y., Shimizu, T., Kodama, Y., Endo, H., Onimaru, R., Oda, M., Fujita, K., Shimizu, S., et al. (2003). Tumor location, cirrhosis, and surgical history contribute to tumor movement in the liver, as measured during stereotactic irradiation using a real-time tumor-tracking radiotherapy system. *International Journal of Radiation Oncology* Biology* Physics*, 56(1):221–228.

[Klein et al., 2009a] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., et al. (2009a). Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, 46(3):786–802.

[Klein et al., 2009b] Klein, S., Pluim, J. P., Staring, M., and Viergever, M. A. (2009b). Adaptive stochastic gradient descent optimisation for image registration. *International journal of computer vision*, 81(3):227.

[Klein et al., 2010] Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. (2010). Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205.

[Klein et al., 2007] Klein, S., Staring, M., and Pluim, J. P. (2007). Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE transactions on image processing*, 16(12):2879–2890.

[Kontaxis et al., 2017] Kontaxis, C., Bol, G., Stemkens, B., Glitzner, M., Prins, F., Kerkmeijer, L., Lagendijk, J., and Raaymakers, B. (2017). Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the mr-linac. *Physics in Medicine & Biology*, 62(18):7233.

[Lee et al., 1996] Lee, S., Wolberg, G., Chwa, K.-Y., and Shin, S. Y. (1996). Image metamorphosis with scattered feature constraints. *IEEE transactions on visualization and computer graphics*, 2(4):337–354.

[Lee et al., 1997] Lee, S., Wolberg, G., and Shin, S. Y. (1997). Scattered data interpolation with multilevel b-splines. *IEEE transactions on visualization and computer graphics*, 3(3):228–244.

[Li et al., 2018] Li, B., Yan, J., Wu, W., Zhu, Z., and Hu, X. (2018). High performance visual tracking with siamese region proposal network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8971–8980.

[Litjens et al., 2017] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

[Lorenzen et al., 2006] Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., and Joshi, S. (2006). Multi-modal image set registration and atlas formation. *Medical image analysis*, 10(3):440–451.

[Lorenzi et al., 2013] Lorenzi, M., Ayache, N., Frisoni, G. B., Pennec, X., (ADNI, A. D. N. I., et al. (2013). Lcc-demons: a robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81:470–483.

[Lukezic et al., 2017] Lukezic, A., Vojir, T., ˇCehovin Zajc, L., Matas, J., and Kristan, M. (2017). Discriminative correlation filter with channel and spatial reliability. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6309–6318.

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

[Mang et al., 2019] Mang, A., Gholami, A., Davatzikos, C., and Biros, G. (2019). Claire: A distributed-memory solver for constrained large deformation diffeomorphic image registration. *SIAM Journal on Scientific Computing*, 41(5):C548–C584.

[Marchesseau et al., 2017] Marchesseau, S., Chatelin, S., and Delingette, H. (2017). Nonlinear biomechanical model of the liver. In *Biomechanics of living organs*, pages 243–265. Elsevier.

[Massey Jr, 1951] Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

[Mattes et al., 2003] Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). Pet-ct image registration in the chest using free-form deformations. *IEEE transactions on medical imaging*, 22(1):120–128.

[Mattes et al., 2001] Mattes, D., Haynor, D. R., Vesselle, H., Lewellyn, T. K., and Eubank, W. (2001). Nonrigid multimodality image registration. In *Medical Imaging 2001: Image Processing*, volume 4322, pages 1609–1621. International Society for Optics and Photonics.

[Mazur et al., 2016] Mazur, T. R., Fischer-Valuck, B. W., Wang, Y., Yang, D., Mutic, S., and Li, H. H. (2016). Sift-based dense pixel tracking on 0.35 t cine-mr images acquired during image-guided radiation therapy with application to gating optimization. *Medical physics*, 43(1):279–293.

[McClelland et al., 2013] McClelland, J. R., Hawkes, D. J., Schaeffter, T., and King, A. P. (2013). Respiratory motion models: a review. *Medical image analysis*, 17(1):19–42.

[Menten et al., 2018] Menten, M. J., Fast, M. F., Wetscherek, A., Rank, C. M., Kachelrieß, M., Collins, D. J., Nill, S., and Oelfke, U. (2018). The impact of 2d cine mr imaging parameters on automated tumor and organ localization for mr-guided real-time adaptive radiotherapy. *Physics in Medicine & Biology*, 63(23):235005.

[Mercier et al., 2012] Mercier, L., Del Maestro, R. F., Petrecca, K., Araujo, D., Haegelen, C., and Collins, D. L. (2012). Online database of clinical mr and ultrasound images of brain tumors. *Medical physics*, 39(6Part1):3253–3261.

[Miao et al., 2016] Miao, S., Wang, Z. J., and Liao, R. (2016). A cnn regression approach for real-time 2d/3d registration. *IEEE transactions on medical imaging*, 35(5):1352–1363.

[Modersitzki, 2004] Modersitzki, J. (2004). *Numerical methods for image registration*. Oxford University Press on Demand.

[Mogadas et al., 2018] Mogadas, N., Sothmann, T., Knopp, T., Gauer, T., Petersen, C., and Werner, R. (2018). Influence of deformable image registration on 4d dose simulation for extracranial sbrt: A multi-registration framework study. *Radiotherapy and Oncology*, 127(2):225–232.

[Moré, 1978] Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

[Motegi et al., 2019] Motegi, K., Tachibana, H., Motegi, A., Hotta, K., Baba, H., and Akimoto, T. (2019). Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy. *Journal of applied clinical medical physics*, 20(1):229–236.

[Murphy et al., 2011] Murphy, K., Van Ginneken, B., Reinhardt, J. M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G. E., Garcia, V., et al. (2011). Evaluation of registration methods on thoracic ct: the empire10 challenge. *IEEE transactions on medical imaging*, 30(11):1901–1920.

[Murphy, 2004] Murphy, M. J. (2004). Tracking moving organs in real time. In *Seminars in radiation oncology*, volume 14, pages 91–100. Elsevier.

[Muyan-Ozcelik et al., 2008] Muyan-Ozcelik, P., Owens, J. D., Xia, J., and Samant, S. S. (2008). Fast deformable registration on the gpu: A cuda implementation of demons. In *2008 International Conference on Computational Sciences and Its Applications*, pages 223–233. IEEE.

[Nielsen et al., 2019] Nielsen, R. K., Darkner, S., and Feragen, A. (2019). Topaware: Topology-aware registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 364–372. Springer.

[NiftyReg, 2021] NiftyReg (2021). Open-source software for efficient medical image registration. Web Page: http://cmictig.cs.ucl.ac.uk/wiki/index.php/NiftyReg. [Consulted in June 2021].

[Noblet et al., 2005] Noblet, V., Heinrich, C., Heitz, F., and Armspach, J.-P. (2005). 3-d deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *IEEE Transactions on image processing*, 14(5):553–566.

[Nocedal and Wright, 2006] Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

[Oh et al., 2011] Oh, J., Martin, D., and Škrinjar, O. (2011). Gpu-based motion correction of contrast-enhanced liver mri scans: An opencl implementation. In *2011 IEEE international symposium on biomedical imaging: from nano to macro*, pages 783–786. IEEE.

[Oliveira and Tavares, 2014] Oliveira, F. P. and Tavares, J. M. R. (2014). Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, 17(2):73–93.

[Olsen et al., 2015] Olsen, J., Green, O., and Kashani, R. (2015). World's first applicaton of mr-guidance for radiotherapy. *Missouri medicine*, 112(5):358.

[OpenCV, 2021] OpenCV (2021). Open source computer vision library. Web Page: https://opencv.org/. [Consulted in June 2021].

[Paganelli et al., 2015] Paganelli, C., Seregni, M., Fattori, G., Summers, P., Bellomi, M., Baroni, G., and Riboldi, M. (2015). Magnetic resonance imaging–guided versus surrogate-based motion tracking in liver radiation therapy: a prospective comparative study. *International Journal of Radiation Oncology* Biology* Physics*, 91(4):840–848.

[Paganelli et al., 2018] Paganelli, C., Whelan, B., Peroni, M., Summers, P., Fast, M., van de Lindt, T., McClelland, J., Eiben, B., Keall, P., Lomax, T., et al. (2018). Mri-guidance for motion management in external beam radiotherapy: current status and future challenges. *Physics in Medicine & Biology*, 63(22):22TR03.

[Pandey et al., 2018] Pandey, P., Guy, P., Hodgson, A. J., and Abugharbieh, R. (2018). Fast and automatic bone segmentation and registration of 3d ultrasound to ct for the full pelvic anatomy: a comparative study. *International journal of computer assisted radiology and surgery*, 13(10):1515–1524.

[Park et al., 2012] Park, J. C., Park, S. H., Kim, J. H., Yoon, S. M., Song, S. Y., Liu, Z., Song, B., Kauweloa, K., Webster, M. J., Sandhu, A., et al. (2012). Liver motion during cone beam computed tomography guided stereotactic body radiation therapy. *Medical physics*, 39(10):6431–6442.

[Pennec et al., 1999] Pennec, X., Cachier, P., and Ayache, N. (1999). Understanding the "demon's algorithm": 3d non-rigid registration by gradient descent. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 597–605. Springer.

[Plastimatch, 2021] Plastimatch (2021). An open source software for high-performance volumetric registration of medical images. Web Page: http://plastimatch.org/. [Consulted in June 2021].

[Pluim et al., 2003] Pluim, J. P., Maintz, J. A., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004.

[Polzin et al., 2016] Polzin, T., Niethammer, M., Heinrich, M. P., Handels, H., and Modersitzki, J. (2016). Memory efficient lddmm for lung ct. In

International conference on medical image computing and computer-assisted intervention, pages 28–36. Springer.

[POPI-model, 2021] POPI-model (2021). A point-validated pixel-based breathing thorax model. Web Page: https://www.creatis.insa-lyon.fr/rio/popi-model_original_page. [Consulted in June 2021].

[Raaymakers et al., 2017] Raaymakers, B., Jürgenliemk-Schulz, I., Bol, G., Glitzner, M., Kotte, A., Van Asselen, B., De Boer, J., Bluemink, J., Hackett, S., Moerland, M., et al. (2017). First patients treated with a 1.5 t mri-linac: clinical proof of concept of a high-precision, high-field mri guided radiotherapy treatment. *Physics in Medicine & Biology*, 62(23):L41.

[Real et al., 2017] Real, E., Shlens, J., Mazzocchi, S., Pan, X., and Vanhoucke, V. (2017). Youtube-boundingboxes: A large high-precision human-annotated data set for object detection in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5296–5305.

[Reaungamornrat et al., 2016] Reaungamornrat, S., De Silva, T., Uneri, A., Vogt, S., Kleinszig, G., Khanna, A. J., Wolinsky, J.-P., Prince, J. L., and Siewerdsen, J. H. (2016). Mind demons: symmetric diffeomorphic deformable registration of mr and ct for image-guided spine surgery. *IEEE transactions on medical imaging*, 35(11):2413–2424.

[Rietzel and Chen, 2006] Rietzel, E. and Chen, G. T. (2006). Deformable registration of 4d computed tomography data. *Medical physics*, 33(11):4423–4430.

[RIRE, 2021] RIRE (2021). Retrospective image registration evaluation project. Web Page: https://rire.insight-journal.org/. [Consulted in June 2021].

[Rivaz et al., 2014] Rivaz, H., Chen, S. J.-S., and Collins, D. L. (2014). Automatic deformable mr-ultrasound registration for image-guided neurosurgery. *IEEE transactions on medical imaging*, 34(2):366–380.

[Rohlfing, 2011] Rohlfing, T. (2011). Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. *IEEE transactions on medical imaging*, 31(2):153–163.

[Romero et al., 2009] Romero, A. M., Zinkstok, R. T., Wunderink, W., van Os, R. M., Joosten, H., Seppenwoolde, Y., Nowak, P. J., Brandwijk, R. P., Verhoef, C., Ijzermans, J. N., et al. (2009). Stereotactic body radiation therapy for liver tumors: impact of daily setup corrections and day-to-day anatomic variations on dose in target and organs at risk. *International Journal of Radiation Oncology* Biology* Physics*, 75(4):1201–1208.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[Rosenman et al., 1998] Rosenman, J. G., Miller, E. P., and Cullip, T. J. (1998). Image registration: an essential part of radiation therapy treatment planning. *International Journal of Radiation Oncology* Biology* Physics*, 40(1):197–205.

[Roy et al., 2014] Roy, S., Carass, A., Jog, A., Prince, J. L., and Lee, J. (2014). Mr to ct registration of brains using image synthesis. In *Medical Imaging 2014: Image Processing*, volume 9034, page 903419. International Society for Optics and Photonics.

[Rueckert and Aljabar, 2010] Rueckert, D. and Aljabar, P. (2010). Nonrigid registration of medical images: Theory, methods, and applications [applications corner]. *IEEE Signal Processing Magazine*, 27(4):113–119.

[Rueckert et al., 2006] Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., and Hammers, A. (2006). Diffeomorphic registration using b-splines. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 702–709. Springer.

[Rueckert et al., 1999] Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721.

[Rupp et al., 2016] Rupp, K., Tillet, P., Rudolf, F., Weinbub, J., Morhammer, A., Grasser, T., Jungel, A., and Selberherr, S. (2016). Viennacl—linear algebra library for multi-and many-core architectures. *SIAM Journal on Scientific Computing*, 38(5):S412–S439.

[Samant et al., 2008] Samant, S. S., Xia, J., Muyan-Özçelik, P., and Owens, J. D. (2008). High performance computing for deformable image registration: Towards a new paradigm in adaptive radiotherapy. *Medical physics*, 35(8):3546–3553.

[Schonberger and Frahm, 2016] Schonberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113.

[Seppenwoolde et al., 2002] Seppenwoolde, Y., Shirato, H., Kitamura, K., Shimizu, S., Van Herk, M., Lebesque, J. V., and Miyasaka, K. (2002). Precise and real-time measurement of 3d tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 53(4):822–834.

[Seppenwoolde et al., 2011] Seppenwoolde, Y., Wunderink, W., Wunderink-van Veen, S., Storchi, P., Romero, A. M., and Heijmen, B. (2011). Treatment precision of image-guided liver sbrt using implanted fiducial markers depends on marker–tumour distance. *Physics in Medicine & Biology*, 56(17):5445.

[Seregni et al., 2018] Seregni, M., Paganelli, C., Summers, P., Bellomi, M., Baroni, G., and Riboldi, M. (2018). A hybrid image registration and matching framework for real-time motion tracking in mri-guided radiotherapy. *IEEE Transactions on Biomedical Engineering*, 65(1):131–139.

[Shams et al., 2007] Shams, R., Kennedy, R. A., Sadeghi, P., and Hartley, R. (2007). Gradient intensity-based registration of multi-modal images of the brain. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.

[Shams et al., 2010] Shams, R., Sadeghi, P., Kennedy, R. A., and Hartley, R. I. (2010). A survey of medical image registration on multicore and the gpu. *IEEE signal processing magazine*, 27(2):50–60.

[Sharp et al., 2007] Sharp, G., Kandasamy, N., Singh, H., and Folkert, M. (2007). Gpu-based streaming architectures for fast cone-beam ct image reconstruction and demons deformable registration. *Physics in Medicine & Biology*, 52(19):5771.

[Shi et al., 2014] Shi, X., Diwanji, T., Mooney, K. E., Lin, J., Feigenberg, S., D'Souza, W. D., and Mistry, N. N. (2014). Evaluation of template matching for tumor motion management with cine-mr images in lung cancer patients. *Medical physics*, 41(5):052304.

[Shimohigashi et al., 2017] Shimohigashi, Y., Toya, R., Saito, T., Ikeda, O., Maruyama, M., Yonemura, K., Nakaguchi, Y., Kai, Y., Yamashita, Y., Oya, N., et al. (2017). Tumor motion changes in stereotactic body radiotherapy for liver tumors: an evaluation based on four-dimensional cone-beam computed tomography and fiducial markers. *Radiation Oncology*, 12(1):1–8.

[Simonovsky et al., 2016] Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. (2016). A deep metric for multimodal registration. In *International conference on medical image computing and computer-assisted intervention*, pages 10–18. Springer.

[Sotiras et al., 2013] Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190.

[Stemkens et al., 2016] Stemkens, B., Tijssen, R. H., De Senneville, B. D., Lagendijk, J. J., and Van Den Berg, C. A. (2016). Image-driven, model-based 3d abdominal motion estimation for mr-guided radiotherapy. *Physics in Medicine & Biology*, 61(14):5335.

[Swaminath et al., 2015] Swaminath, A., Massey, C., Brierley, J. D., Dinniwell, R., Wong, R., Kim, J. J., Velec, M., Brock, K. K., and Dawson, L. A. (2015). Accumulated delivered dose response of stereotactic body radiation therapy for liver metastases. *International Journal of Radiation Oncology* Biology* Physics*, 93(3):639–648.

[Tascón-Vidarte et al., 2021] Tascón-Vidarte, J. D., Wahlstedt, I., Jomier, J., Erleben, K., Vogelius, I. R., and Darkner, S. (2021). Cine-mri simulation to evaluate tumor tracking. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 131–141. Springer.

[Telea, 2004] Telea, A. (2004). An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34.

[Thévenaz et al., 2000] Thévenaz, P., Blu, T., and Unser, M. (2000). Interpolation revisited [medical images application]. *IEEE Transactions on medical imaging*, 19(7):739–758.

[Thevenaz et al., 1998] Thevenaz, P., Ruttimann, U. E., and Unser, M. (1998). A pyramid approach to subpixel registration based on intensity. *IEEE transactions on image processing*, 7(1):27–41.

[Thirion, 1998] Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with maxwell's demons.

[Tryggestad et al., 2013] Tryggestad, E., Flammang, A., Hales, R., Herman, J., Lee, J., McNutt, T., Roland, T., Shea, S. M., and Wong, J. (2013). 4d tumor centroid tracking using orthogonal 2d dynamic mri: implications for radiotherapy planning. *Medical physics*, 40(9):091712.

[Tustison et al., 2013] Tustison, N. J. et al. (2013). Explicit b-spline regularization in diffeomorphic image registration. *Frontiers in neuroinformatics*, 7:39.

[Vandemeulebroucke et al., 2011] Vandemeulebroucke, J., Rit, S., Kybic, J., Clarysse, P., and Sarrut, D. (2011). Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs. *Medical physics*, 38(1):166–178.

[Velec et al., 2012] Velec, M., Moseley, J. L., Craig, T., Dawson, L. A., and Brock, K. K. (2012). Accumulated dose in liver stereotactic body radiotherapy: positioning, breathing, and deformation effects. *International Journal of Radiation Oncology* Biology* Physics*, 83(4):1132–1140.

[Velec et al., 2017] Velec, M., Moseley, J. L., Svensson, S., Hårdemark, B., Jaffray, D. A., and Brock, K. K. (2017). Validation of biomechanical deformable image registration in the abdomen, thorax, and pelvis in a commercial radiotherapy treatment planning system. *Medical physics*, 44(7):3407–3417.

[Vercauteren et al., 2009] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.

[Vickress et al., 2016] Vickress, J., Battista, J., Barnett, R., Morgan, J., and Yartsev, S. (2016). Automatic landmark generation for deformable image registration evaluation for 4d ct images of lung. *Physics in Medicine & Biology*, 61(20):7236.

[ViennaCL, 2021] ViennaCL (2021). Free open-source linear algebra library for computations on many-core architectures (gpus, mic) and multi-core cpus. Web Page: http://viennacl.sourceforge.net/. [Consulted in June 2021].

[Viergever et al., 2016] Viergever, M. A., Maintz, J. A., Klein, S., Murphy, K., Staring, M., and Pluim, J. P. (2016). A survey of medical image registration–under review.

[Waidyasooriya et al., 2018] Waidyasooriya, H. M., Hariyama, M., and Uchiyama, K. (2018). *Design of FPGA-based computing systems with OpenCL*. Springer.

[Wang and Deng, 2018] Wang, M. and Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153.

[Werner et al., 2014] Werner, R., Schmidt-Richberg, A., Handels, H., and Ehrhardt, J. (2014). Estimation of lung motion fields in 4d ct data by variational non-linear intensity-based registration: A comparison and evaluation study. *Physics in Medicine & Biology*, 59(15):4247.

[West et al., 1997] West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer Jr, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., et al. (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of computer assisted tomography*, 21(4):554–568.

[Wolthaus et al., 2006] Wolthaus, J. W., Schneider, C., Sonke, J.-J., van Herk, M., Belderbos, J. S., Rossi, M. M., Lebesque, J. V., and Damen, E. M. (2006). Mid-ventilation ct scan construction from four-dimensional respiration-correlated ct scans for radiotherapy planning of lung cancer patients. *International Journal of Radiation Oncology* Biology* Physics*, 65(5):1560–1571.

[Woods et al., 1992] Woods, R. P., Cherry, S. R., and Mazziotta, J. C. (1992). Rapid automated algorithm for aligning and reslicing pet images. *Journal of computer assisted tomography*, 16(4):620–633.

[Worm et al., 2016] Worm, E. S., Bertholet, J., Høyer, M., Fledelius, W., Hansen, A. T., Larsen, L. P., Nielsen, J. E., and Poulsen, P. R. (2016). Fiducial marker guided stereotactic liver radiotherapy: Is a time delay between marker implantation and planning ct needed? *Radiotherapy and Oncology*, 121(1):75–78.

[Worm et al., 2013] Worm, E. S., Høyer, M., Fledelius, W., Hansen, A. T., and Poulsen, P. R. (2013). Variations in magnitude and directionality of respiratory target motion throughout full treatment courses of stereotactic body radiotherapy for tumors in the liver. *Acta Oncologica*, 52(7):1437–1444.

[Wu et al., 2013a] Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., and Shen, D. (2013a). Unsupervised deep feature learning for deformable registration of mr brain images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–656. Springer.

[Wu et al., 2013b] Wu, Y., Lim, J., and Yang, M.-H. (2013b). Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2411–2418.

[Wunderink et al., 2008] Wunderink, W., Romero, A. M., De Kruijf, W., De Boer, H., Levendag, P., and Heijmen, B. (2008). Reduction of respiratory liver tumor motion by abdominal compression in stereotactic body frame, analyzed by tracking fiducial markers implanted in liver. *International Journal of Radiation Oncology* Biology* Physics*, 71(3):907–915.

[Wunderink et al., 2010] Wunderink, W., Romero, A. M., Seppenwoolde, Y., De Boer, H., Levendag, P., and Heijmen, B. (2010). Potentials and limitations of guiding liver stereotactic body radiation therapy set-up on liver-implanted fiducial markers. *International Journal of Radiation Oncology* Biology* Physics*, 77(5):1573–1583.

[Yang et al., 2008] Yang, D., Li, H., Low, D. A., Deasy, J. O., and El Naqa, I. (2008). A fast inverse consistent deformable image registration method based on symmetric optical flow computation. *Physics in Medicine & Biology*, 53(21):6143.

[Yang et al., 2017a] Yang, D., Zhang, M., Chang, X., Fu, Y., Liu, S., Li, H. H., Mutic, S., and Duan, Y. (2017a). A method to detect landmark pairs accurately between intra-patient volumetric medical images. *Medical physics*, 44(11):5859–5872.

[Yang et al., 2016] Yang, X., Kwitt, R., and Niethammer, M. (2016). Fast predictive image registration. In *Deep Learning and Data Labeling for Medical Applications*, pages 48–57. Springer.

[Yang et al., 2017b] Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017b). Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396.

[Yeo et al., 2009] Yeo, B. T., Sabuncu, M. R., Vercauteren, T., Ayache, N., Fischl, B., and Golland, P. (2009). Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE transactions on medical imaging*, 29(3):650–668.

[Younes, 2010] Younes, L. (2010). *Shapes and diffeomorphisms*, volume 171. Springer Science & Business Media.

[Yun et al., 2015] Yun, J., Yip, E., Gabos, Z., Wachowicz, K., Rathee, S., and Fallone, B. (2015). Neural-network based autocontouring algorithm for intrafractional lung-tumor tracking using linac-mr. *Medical physics*, 42(5):2296–2310.

[Zachiu et al., 2017] Zachiu, C., de Senneville, B. D., Tijssen, R. H., Kotte, A. N., Houweling, A. C., Kerkmeijer, L. G., Lagendijk, J. J., Moonen, C. T., and Ries, M. (2017). Non-rigid ct/cbct to cbct registration for online external beam radiotherapy guidance. *Physics in Medicine & Biology*, 63(1):015027.

[Zachiu et al., 2015] Zachiu, C., Papadakis, N., Ries, M., Moonen, C., and De Senneville, B. D. (2015). An improved optical flow tracking technique for real-time mr-guided beam therapies in moving organs. *Physics in Medicine & Biology*, 60(23):9003.

[Zhang, 2018] Zhang, J. (2018). Inverse-consistent deep networks for unsupervised deformable image registration. *arXiv preprint arXiv:1809.03443*.

[Zhu et al., 2018] Zhu, Z., Wang, Q., Li, B., Wu, W., Yan, J., and Hu, W. (2018). Distractor-aware siamese networks for visual object tracking. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 101–117.