Politecnico di Milano
Department of Bioengineering
PhD program in Bioengineering

METHODS AND ALGORITHMS FOR IMAGE GUIDED ADAPTIVE RADIO- AND HADRON-THERAPY

Doctoral Dissertation by
MARTA PERONI
738782

Tutor: Prof. A. Pedotti
Supervisor: Prof. G. Baroni
Chair of the Doctoral Program: Prof. M.G. Signorini

XXIV CICLO
2009-2011

"Não sou nada. Nunca serei nada. Não posso querer ser nada. À parte isso, tenho em mim todos os sonhos do mundo."

— F. Pessoa

"[...] nos esse quasi nanos gigantium humeris insidentes, ut possim plura eis et remotiora videre, non utique proprii visus acumine aut eminentia corporis, sed quia in altum subvehimur et extollimur magnitudine gigantea. "

"[...] we are like dwarves perched on the shoulders of giants, and thus we are able to see more and farther than the latter. And this is not at all because of the acuteness of our sight or the stature of our body, but because we are carried aloft and elevated by the magnitude of the giants"

— Bernardo de Chartres (in Giovanni di Salisbury, Metalogicon, III, 4)

"Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid." — A. Einstein

Dedicated to my parents and to Marco.
Ai miei genitori e a Marco.

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

Journal papers

- Murphy K., Van Ginneken B., Reinhardt J., Kabus S., Ding K., Deng X., Cao K., Du K., Christensen G., Garcia V., Vercauteren T., Ayache N., Commowick O., Malandain G., Glocker B., Paragios N., Navab N., Gorbunova V., Sporring J., De Bruijne M.,; Han X., Heinrich M., Schnabel J., Jenkinson M., Lorenz C., Modat M., McClelland J., Ourselin S., Muenzing S., Viergever M., De Nigris D., Collins L., Arbel T., Peroni M., Li R., Sharp G., Schmidt-Richberg A., Ehrhardt J., Werner R., Smeets D., Loeckx D., Song G., Tustison N., Avants B., Gee J., Staring M., Klein S., Stoel B., Urschler M., Werlberger M., Vandemeulebroucke J., Rit S., Sarrut D., Pluim J.,
  "Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge."
  IEEE Trans Med Imaging. 2011 May 31. [Epub ahead of print]

- M.F. Spadea, M. Peroni, E. Preve, M. Riboldi, G. Baroni, G.T.Y. Chen, G.C. Sharp
  "Uncertainties in lung motion prediction relying on external surrogate: a 4DCT study in regular vs. irregular breathers."
  Technol Cancer Res Treat. **9** 3:307-16 (2010)

Peer-reviewed conferences

- Paganelli C., Peroni M., Baroni G., Riboldi M.,
  "Validation of deformable registration in adaptive radiation therapy with scale invariant feature transform"
  IEEE International Symposium on Biomedical Imaging,2012, Barcelona, Spain

- Peroni M., Golland P., Sharp G.C., Baroni G.,
  "Ranking of Stopping Criteria for Log Domain Diffeomorphic Demons Application in Clinical Radiation Therapy"
  33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA

- G.C. Sharp, M. Peroni, R. Li, J. Shackleford and N. Kandasamy,
  "Evaluation of Plastimatch B-Spline Registration on the EMPIRE10 Data Set"
  Workshop Proceedings from the 13th International Conference of Medical Imaging Computing and Computer Assisted Intervention, Beijing (China), September 20th - 24th 2010, ISBN: 978-1453759394

- M. Peroni, J. Stancanello, A. Muacevic, M. Riboldi, M.F. Spadea, G.C. Sharp, G. Baroni,

"Feasibility Study of B-Spline Deformable Registration between CT and MRI Atlas for Trigeminal Neuralgia Application to Radiosurgery."
World Congress of medical physics and biomedical engineering 2009 - 7-12 September 2009. (vol. 25). ISBN/ISSN: 978-3-642-03897-6 / 1680-0737.

- M. Peroni, M.F. Spadea, M. Riboldi, G. Baroni, G.T.Y. Chen, G.C. Sharp,
"Validation of an automatic contour propagation method for lung cancer 4D adaptive radiation therapy"
IEEE International Symposium on Biomedical Imaging: From Nano to Macro,2009, Proceedings, p 1071-1074.

Conference contributions: talks and poster

Talks

- at the *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society,*
Contributed Paper: "Ranking of Stopping Criteria for Log Domain Diffeomorphic Demons Application in Clinical Radiation Therapy"
Aug 30th - Sept 3rd 2011, Boston, MA, USA

- at *World Congress of medical physics and biomedical engineering 2009.*,
Contributed Paper: *"Feasibility Study of B-Spline Deformable Registration between CT and MRI Atlas for Trigeminal Neuralgia Application to Radiosurgery"*
September 7-12th 2009, München, Deutschland

Posters

- at *11th biennal ESTRO - Conference of the European Society for Radiotherapy and Oncology 2011*,
Contributed Paper: *"GENERATION OF VIRTUAL DAILY PLANNING CT FOR ADAPTIVE RADIOTHERAPY IN HEAD AND NECK CANCER. Radiother Oncol. 99 (Suppl 1), S444"*
May 8-11th, 2011, London, UK

- at *XVIth International Conference on the Use of Computers in Radiation Therapy*,
Contributed Paper: *"QUANTITATIVE EVALUATION OF A DEFORMABLE REGISTRATION TOOLKIT."*
Contributed Paper: *"An Open Source Software Suite for Radiotherapy Image Processing."*
May 31st- June 4th 2010, Amsterdam, The Netherlands

- at the *52nd Annual Meeting of the American Association of Physicist in Medicine 2010*,
Contributed Paper: *"Automatic Contour Propagation between Planning Computed Tomography (CT) and Cone Beam CT (CBCT) Scan for In-Room Adaptive Planning: A Feasibility Study on Nasopharyngeal Cancer Patients. Med. Phys. 37, 3187 (2010)"*
July 18-22th 2010, Philadelphia, USA

- at *10th biennal ESTRO - Conference of the European Society for Radiotherapy and Oncology 2009*,
  Contributed Paper: *"Merging and non-rigid registration of CBCTs for large field adaptive radiotherapy in nasopharyngeal cancer. Radiother Oncol. 92 (Suppl 1), S164-S165. "*
  August 30th - September 3rd, 2009, Maastricht, The Netherlands

- at the *51st Annual Meeting of the American Association of Physicist in Medicine 2009*,
  Contributed Paper: *"C. Hancox, J. Seco, G.C. Sharp, M. Peroni, H. Paganetti, Dynamic Monte Carlo Dose Calculations for IMRT in Geant4, Med. Phys. 36, 2623 (2009), DOI:10.1118/1.3181921"*
  July 26-30th 2009, Anaheim, USA

The important thing is not to stop questioning.
Curiosity has its own reason for existing.
It is enough if one tries merely to comprehend a little of this mystery every day.
Never lose a holy curiosity.
— Einstein, 1879-1955

ACKNOWLEDGMENTS

First of all my unconditional thank goes to mum and dad. Your years of love and unwavering support shaped me into the person I am today. I cannot express in words my gratitude for the sacrifices you made for me. Without you I would never have been able to make it even nearly this far. Thanks to Marco and his re-shaping and re-doing Marta, for taking me up when I was down and never let me fall again, thanks, WE made it. Vedra, si anche tu sei parte del progetto e anche a te va il mio grazie.

I'd like to thank Prof. Pedotti, for having granted the possibility of working at TBM lab, and Prof. Signorini, for guiding us during the PhD program. None of this work would have been possible without the unconditional support of my advisor, Prof. Guido Baroni. I would like to thank you for, at times despite my best efforts, keeping me focused and on track. Your enthusiasm and ideas have been indispensable to me. Dr Sharp, Dr Seco, Prof. Golland, thank you for challenging me and helping make these years the most interesting and stimulating time of my life. At the same time, I want to thank Dr Riboldi and Dr Spadea for all the passion, support, dedication and energy they spent in this 3-years (oh my, it is really 3 years?) work. My thesis would have never be the same if they were not there asking question, supporting, teaching and sometimes also making fun of me: thank you for everything. A warm thank also to the Rocca Fundation for their support in connecting me to MIT and to the people at IEO (Dr. Alterio, Dr. Comi amongst others) , always trying to make my work clinically relevant and making me believe someone *could* actually benefit from my work.

Thanks to all of those who did bet non me: you see, I made it :) You know who you are, do not need to mention all the names.

Thanks to all of those who did NOT bet non me: you see, I made it :) you knew I would have been stubborn enough to do it :)

Grazie alle nonne, valvola di sfogo per i miei genitori quando la strada di questi tre anni s'è fatta dura. Thanks to aunties and uncles...thanks Gioanna, now it is your turn for all this stuff about studying at the university.

Thanks to all the CART lab/TBM lab, because you know what it means to do a PhD...Thanks to all the people I came across at MIT/Medical Computer Vision Lab...Thanks to all the Rocca-people...you know the great people you are!

Thanks to all the friends and people who stumbled on my work and/or myself...to cite a few in random order... Delia, ChiaraP, PaoloZ, Lorenzo, Maryide,

Elga, Alfonso, Daniele, Francesco, Erik il tonno Tonni, Giuseppe, Giulia, Tino-Giacomo-capitano, Lady Elena, Antonella, Carlo, Simona, Matteo, Morgan-etto, MarcoCobasIdealeAntonio e le sue apine, Antonio, Erika, Andrea, Costa che però si chiama Fabrizio, Francesca, Ardian, Alan, Elena, Floriano, Chiara, Roberto, Federico, Chiara, Marco, Daniela, Alberto, Elisa, Ernesto, Elisa, Rossy, Ema, Fede, Martina...e tutti gli altri!

Alessandro Gig-nano, Giulia, you know...you know you have a special place in this list. Thanks!

Thanks to "quelli del treno", Teo (+Barbara+Adelaide), Corrado, Fabio, Stefano, Samuela Samu Mu, Simone...Thanks to "quelli della piscina", Giuseppe the coach, Capitan-Gabri, Sonia, Silvia, Stefano, Stefano, Franco, Francesco, Francesca, Vale...per una piacevole distrazione mentale con un pò di relax....

Thanks to Don Michele, Don Federico, gli edu e i miei fanciulli....siete troppi non vi posso nominare tutti, ma voi sapete :)

Thanks to you who took the time to read this...I don't know if you feel under/over-mentioned, but be sure I'm thankful you stumble across my way and walk a short path at my side, so please, even though I'm pushing you away, keep doing it!

uh, by the way...I'll still be...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

RT Radiation Therapy

IGRT Image Guided Radiation Therapy

IGART Image Guided Adaptive Radiation Therapy

IMRT Intensity Modulated Radiation Therapy

EBRT External Beam Radiation Therapy

HN Head and Neck

HNCP Head and Neck Cancer Patient

GTV Gross Tumor Volume

PTV Planning Target Volume

CT Computed Tomography

CBCT Cone Beam Computed Tomography

MRI Magnetic Resonance Imaging

PET Positron Emission Tomography

CTsim Computed Tomography acquired at simulation

CTrepl Computed Tomography acquired at replanning

DIR Deformable Image Registration

BSDIR B-Spline Deformable Image Registration

LDDD Log Domain Diffeomorphic Demons DIR

SCV Stopping Condition Value

NMI Normalized Mutual Information

MSE Mean Squared Error

QU Quantity of Update

Jac Jacobian of the deformation

HE Harmonic Energy

DSC Dice Similarity Coefficient

COM Center Of Mass distance

RMSE Root Mean Squared Error

SIFT Scale Invariant Feature Transform

GW Gaussian Weighted fusion strategy

MV Majority Voting fusion strategy

AVG Averaging fusion strategy

FN Fixed Number selection strategy

TH THresholding selection strategy

ABSTRACT

In the modern clinical practice of Radiation Therapy (RT), the main goal of achieving a high local selectivity implies the necessity of accurately locate and monitor the lesion along the whole treatment. The concept of Image Guided Radiation Therapy (IGRT) has been introduced to underline the central role played by the imaging along the course of therapy. Both in the planning phase and in the delivery phase, the support of images has become fundamental for individuation and targeting of the tumor. Given the fractionated nature of the treatment, the basic idea is to compare the internal or external anatomy, acquired at each fraction during patient set-up routine, with images used in the planning phase. In case of discrepancies, the treatment plan is upgraded taking the new information into account. This procedure is often referred as Image Guided Adaptive Radiation Therapy (IGART).

While the delivery phase has been deeply investigated and new optimized and sophisticated techniques (e.g. Intensity Modulated Radiation Therapy, IMRT) have been developed to minimize damages to normal tissues, the treatment (re)planning phase still leaves some questions open. The whole burden of planning and updating is left on physicians and physicist, soaking up the most of their clinical time and, in certain cases, delaying treatment delivery. The motivation of my PhD project stems directly from these consideration, combined with the growing interest in non-rigid image registation techniques, progressively undisclosing its potential for RT applications.

The current strategy for adaptive External Beam Radiation Therapy (EBRT) is illustrated in Figure 0.1. The whole scheme can be divided into offline and online procedures. Offline phase includes initial images acquisition, structures of interest delineation and irradiation scheme definition. This is characterized by manual intervention as well as by time and computational resources uptake, lasting from one week to one month, depending on the clinical workload. Online procedures instead encompass one or more imaging sessions both for positioning and changes tracking prior to proper treatment delivery. Positioning modifications can be directly addressed by means of couch shifts and tilts, whereas non-rigid alterations of tissues geometry (e.g. tumor or organs at risk relative translation and shape changes) cannot yet be coped for without replanning and reoptimizing the whole treatment. These latter modifications are generally detected with a combination of macroscopic observations (e.g. immobilization devices not fitting anymore) and in-room imaging, such as double X-ray projections and Cone Beam Computed Tomography (CBCT). If changes are beyond an empirically established threshold, a completely manual replanning is done on a new Computed Tomography (CT), acquired sufficiently near the in-room imaging session. Therefore, in the classical strategy, no online adaptation of the treatment is foreseen but rather an offline manual update, mainly consisting of recontouring and dose reoptimization and/or recalculation.

My PhD project has been concentrating on automatization of the manual procedures both in the offline and online phases, aiming at developing fast and accurate automatic strategies for planning segmentation and virtual planning CT genera-

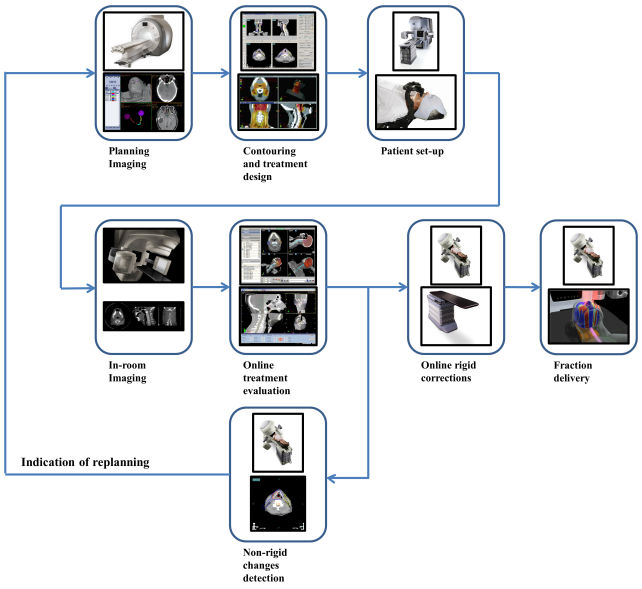


Figure 0.1: Current Clinical Workflow for Image Guided Adaptive Radiation Therapy. Offline phases include planning images acquisition as well as treatment (contours and beam geometry) definition. Patient setup adjustment as well as non-rigid changes detection happens online, right before daily fraction delivery. If significant modifications are detected, the daily treatment is suspended, while offline corrections are computed. This adaptation implies the acquisition of a new Computed Tomography.

tion. We concentrated on Head and Neck tumors, working on a database of total 45 patients, treated at European Institute of Oncology (Milan, Italy). All the patient underwent IMRT with a planned dose of 60-70Gy delivered in 2-2.12Gy per fraction, one fraction per day. Current protocol entails a simulation CT (CTsim, acquired on a GE Medical System Light Speed, Fairfield, CT) with or without contrast medium injection, verification CBCTs (On-Board Imager, OBI, Varian, CA) and replanning CTs (CTrepl) at 40Gy and 50Gy. CBCT and CTrepl are acquired generally one day apart, with 3 days distance as the worst case. Patient position reproducibility during imaging and treatment phases is guaranteed by customized head-neck and shoulders thermoplastic masks fixed onto the couch, bite block, and three radiopaque markers attached to the mask for initial alignment. Main indication for replanning is represented by the over dosage of OARs, as highlighted by Dose Volume Histograms on CTrepl, but no beam modification and/or plan re-optimization was employed in our population. Image resolution (for CTsim, CTrepl and CBCT) was 1mm in AP and LL on average and 3mm along SI direction. Brainstem, optical nerves, eyes, mandible, parotids and spine cord were contoured on CTsim of all the patients, whereas mandible, parotids and Gross Tumor Volumes (GTV) were selected as representative structures subjected to change during therapy.

The technological brick at the basis of my work is non rigid or Deformable Image Registration (DIR), thus the transformation of the different sets of image data into one geometric coordinate system. IR is also often called *data fusion*, to underline its capability of bringing a *floating* and a *reference* image not only into geometric align-

ment, but also literally to make them match. The number of degree of freedom of the transformation is defined a-priori and its parameters are calculated optimizing a cost function defining the similarity between the images involved in the process. The optimization problem in the case of non rigid registration becomes more complicated with the increase of the flexibility of the transformation, which generally speaking corresponds to an increase in the number of parameters. Transformation models control, either at a global or at a local level, how image features can be moved relative to another to improve the similarity and define interpolation rules between features where there is no usable information. This strategy has the potential to model deformations occurring at a different time points of the treatment (intra-subject registration) or between different patients (inter-subject registration), and is not restricted to same modality registration, but can be used to integrate information coming from different types of images, such as anatomical CT with functional Positron Emission Tomography (multi-modal registration).

A popular choice is an elastic non-parametric model, which deforms the image according to forces derived from the similarity measure. The image is considered just as an elastic solid and therefore it has an internal energy, which opposes to the external matching force. The deformation is applied until internal and external forces reach and equilibrium. As it is an elastic model, large and localized deformation are not allowed. Hence, a viscous fluid models have been developed, for example the demons algorithm. In this algorithm, the image is considered a set of isointensity contours, which are then pushed in the normal direction with a magnitude and direction derived from the instantaneous optical flow equation. Another popular choice is a class of parametric transformations based on the usage of splines as regularized interpolators. In particular, quadratic B-splines can be used to parameterize a vector field over the volume of interest. The B-spline coefficients are defined on control points, which lie on a coarse grid that covers the region and the deformation field is a smooth interpolation of the coefficient value.

During the course of my PhD program, Log-Domain Diffeomorphic Demons (LDDD) performances were studied. LDDD energy is optimized in a symmetric log domain by an efficient second-order minimization in which the deformation field ϕ is represented with a smooth and stationary velocity field v. The relationship between the two is defined as $\phi(x) = \exp(v)(x)$ which is a diffeomorphic deformation, progressively filtered by Gaussian Kernels so that its smoothness is conserved throughout the optimization. Main issue regarding this algorithm was the stopping condition or escape criteria at convergence. The most basic stopping condition is to terminate the algorithm after a predetermined number of iterations. Although very simple, this criteria is only weakly related to the actual convergence. To relate the stopping criteria with registration convergence, we analyze conditions based on image intensities and the deformation or velocity field. Criteria based on image intensities are generally aimed at observing the change in similarity between *floating* and *reference* image at current iteration *t* with respect to the previous time points *t-n*, whereas deformation or velocity field analysis focuses more on guaranteeing continuity and smoothness between neighbouring voxels. The study was carried out applying synthetic non-rigid deformation fields to phantom and patient CT. Analyzed trends of stopping conditions included registration metric, harmonic energy (i.e. degree of smoothness), jacobian (i.e. num-

ber of discontinuities) and quantity of update (i.e. speed of convergence) of the velocity field, calculated at each iteration. A set of possible reference values, to compare the current stopping condition value with, was also examined. Results indicate the minimum (maximum) of the last three stopping condition values with the minimum (maximum) of the three iterations before the current one as the best performing reference values. This kind of analysis was not necessary for Jacobian, which we already computed in the form of a percentage error. The four criteria were than ranked in terms of speed of convergence and stability. Both for patient and phantom, the most efficient condition is based on harmonic energy. Quantity of update ranked second in the phantom study, but demonstrated to be sensitive to image quality, while registration metric might be helped or obstacled by strong anchor points in the images (such as metal artifacts). Jacobian does not seem to robustly monitor the transformation evolution in a LDDD run. These results can easily be extended to other registration methods, by just substituting velocity with deformation field, in case the first is not easily accessible.

In addition, an efficient implementation of B-Splines Deformable Image Registration (BSDIR) algorithm was analyzed for use in a clinical environment. Using a grid alignment scheme, Plastimatch (www.plastimatch.org) can significantly accelerate the B-Spline interpolation and gradient computation, thus speeding up the registration process. BSDIR uses cubic B-spline curves to define a displacement field that maps the voxels in the *floating* image to those in a *reference* image. Since the vector field is defined in a parametric fashion (that is, in terms of coefficients provided by a set of control points), a cost function that quantifies the similarity between the images can be specified and registration can be posed as an optimization problem. The registration process then iteratively defines coefficients P, performs B-spline interpolation, evaluates the cost function C, calculates its derivative for each control point, and performs gradient-descent optimization to generate the next set of coefficients. Aligning the B-Splines grid to voxel grid allows us to precompute all relevant B-spline basis function products once instead of recomputing the evaluation for each individual tile of the control grid. The improved efficiency of the registration process is reported in terms of running time on a benchmark dataset. It takes 0.4 ~ 5.7 minutes to register to a pair of 3D CT benchmark images at full resolution. The method is accurate and certainly demonstrate the potential of clinical applicability. However, as all employed evaluation metrics are not conclusive in terms of local registration error, visual inspection is still needed in the clinical setting.

The performance of both LDDD and BSDIR were compared to one of the few FDA approved deformable registration softwares on the market, underlining critical issues and advantages of both approaches. Confirming mathematical properties and literature, BSDIR demonstrated to be more flexible, faster and less sensitive to initialization than Demons algorithm. Nonethless, this approach resulted more prone to discontinuous deformation field and the introduction of a regularization term was necessary.

Having this powerful instrument at our disposal, the problem of generating a robust segmentation of structures of interest of a planning CT was first addressed. The standard clinical procedure right now relies completely on manual contouring, which is a rather long and tedious operation, suffering also from a high degree of

inter- and intra-operator variability. An automatic segmentation method would be desirable to support and speed up clinical workflows.

Automatic segmentation can be coarsely divided into deformable or model-based approaches, and atlas-based methods. The first ones mainly rely on rely on local image features, such as edges and make use of the assumption that object boundaries are distinct. In order to cope for failures in regions where this assumption does not hold (e.g. soft tissues), prior information about the structures aspect has been introduced, but needs to be balanced in order not to limit the accuracy.

Several strategies for atlas based segmentation have been studied, which can be classified based on templates numerousness. In a single atlas strategy, the only template in the database is warped onto the new subject. The selection of single atlas can be done naively or rely on image quality, but can deeply affect the final segmentation outcome. In fact, its reliability increases exponentially, if the selected template is more similar to the new subject, as this would lower the ill-poseness of the registration problem. A popular choice for this algorithm is the employment of anthropomorphic phantoms, whose segmentation is guaranteed to be optimal. Besides being computational advantageous, this strategy cannot represent the anatomical variability to the fullest. A second possible strategy relies on the construction of a database of atlases, amongst which the most similar to the new patient is selected. In comparison to the previous approach, the main advantage is the ability of maximizing the probability of the chosen subject to be similar to the new patient, if the database is sufficiently varied. On the other side, the risk of compromising the results because of the failure of image registration and of a non sufficient rappresentation of the anatomical variability in the database are still present. An approach involving the construction of an average atlas might overcome this issue, despite a degradation of the information due to the co-registration of diverse model cases, shown as diffuse blur, getting more intense at structures borders.

In the present work, we developed a fully automatic multi-atlas based segmentation algorithm (Figure 0.2), for Head and Neck (HN) CT datasets. We considered the following structures: mandible, spine cord, eyes, optical nerves, parotid glands and brainstem. Generally speaking, we define atlas a CT image with its correspondent manual segmentation. Our standing point is to build a comprehensive database of patients (i.e. atlases), without establishing restrictive inclusions rules, but rather representing the general clinical variability. The database atlases are registered onto the patient to be segmented in a pairwise fashion and the resulting transformation is used to warp manual contours. Since the warped structures are all equally probable realizations of the anatomical structure of the new subject, we need to develop an efficient and robust strategy for final labelling estimation.

The first problem addressed was how to robustly clustering the atlases to select only the most appropriate for the segmentation of the test patient. The evaluated strategies rely on the pre-alignment of database subjects with a reference image (in my case an anthropomorphic phantom) by means of affine registration. The pre-alignment can occurr offline at the moment of database construction, while, at the moment of planning, just the new subject needs to be aligned to the common reference. Similarities in terms of Normalized Mutual Information (NMI) shared by each atlas and the CT scan to be segmented are than computed. Based on

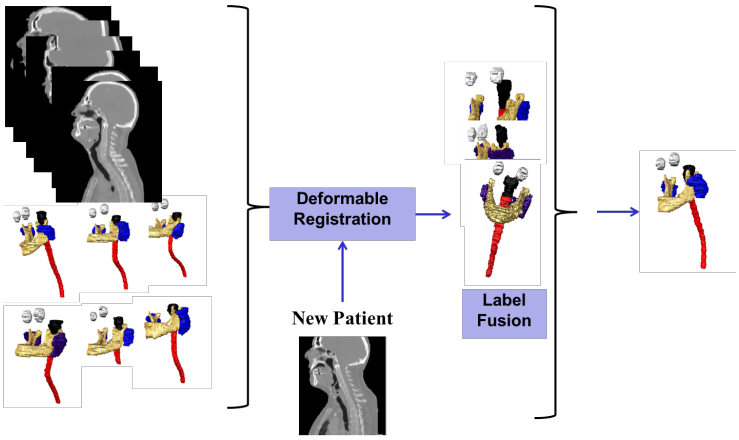


Figure 0.2: Atlas based segmentation workflow. A database of atlases (i.e. pre-segmented patients) are clustered according to their similarity to the new patient to be segmented after having aligned everything on a reference image. The selected template subjects are warped onto the new patient, applying the deformation field estimated by a non-rigid image registration. The anatomical variability is accounted for by the choice of more than one subject as atlas, but error introduced by deformable registration and label fusion should be carefully analyzed.

this ranking, the group of subjects to be actually used for image segmentation is than selected in a fixed number or NMI thresholding fashion. Fixing the number of subjects has the main advantage of providing the same anatomical variability no matter how the new subject looks like, while NMI thresholding guarantees the inclusion of just the most similar patients. Nonetheless, if the CT scan to be segmented is not well represented in the database, the first strategy could select sub-optimal patients, while the second method could choose just one atlas.

The selected atlases are than deformed onto the patient image by means of LDDD deformable registration and the corresponding segmented structures are propagated using the same deformation. The tentative segmentations are then recombined according to three different strategies (mean, majority voting and weighted fusion), summarized in Table 0.1, in order to select the most robust and accurate for RT planning. The main difference between mean and majority voting is to be found on the structures borders, where a clear majority may not be defined. Weighted fusion instead is a framework that assigns to each voxel a probability of being actually part of the labelled structure, combining a gaussian weighted difference between patient to be segmented and registered atlas images with an exponential distance transform map of the propagated labels. This method was first developed for Magnetic Resonance scans, in which the transition between the different structures is rather clear. In HN CT scans, instead, the Hounsfield Unit distribution is rather homogeneous in soft tissues, while bony structures are generally well contrasted. For these reasons, we relaxed the original hypothesis according to which the probability of belonging to the structures has to be strictly bigger than the one of not being part of it at each voxel. The reason for this assumption has to be found considering that on one side the low contrast in soft tissues causes not only uncertainity in image registration but also higher variability in contouring and on the other side over-segmentation is less critical than under-segmentation for

Table 0.1: Label Fusion strategies

Strategy	Formula
Average	$L(x) = \text{round}\left(\frac{\sum_s L_s(x)}{S}\right)$
Majority Voting	$L(x) = \begin{cases} 1 & \text{if } \sum_s L_s(x) > \frac{S}{2} \\ 0 & \text{in all other cases} \end{cases}$
Gaussian Weighted	$L(x) = \arg\max \sum p(L(x) = l \mid L_s, \Phi s) p(I(x) \mid I_s, \Phi s)$ where $p(L(x) = l \mid L_s, \Phi s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{1}{2\sigma^2}(I(x) - I_s(\phi_s(x)))^2}$ $p(I(x) \mid I_s, \Phi s) \propto \exp^{\rho D_s^l(\phi_s(x))}$

the foreseen RT application. Therefore, a scaling factor k was introduced, according to which $likelihood_1 \geqslant k \cdot likelihood_0$, where $likelihood_n$ is the probability of the voxel of assuming a value 1 (i.e. belonging to the real segmented structure) or 0 (i.e. not belonging to the real segmented structure) and $k = [0; 1]$. The factor k is taking into account the uncertainities connected with manual segmentation of the atlases in the database, as well as of deformable registration.

The full algorithm was tested with a leave-one-out strategy, which consists in selecting one patient at a time amongst all thirty-one CTsim of the patients in our database, thus serving as target patient, and use all the rest as atlases. Patients were acquired according to the standard clinical protocol designed for their patology. The automatic segmentation of each patient was than compared with the manual contours of the target patients by means of Dice Coefficient, which evaluates the percentage overlap between the volumes. We further compared the strategies by means of Wilcoxon rank test.

The percentile based selection and the fusion strategy based on Gaussian Weighted Fusion between all the propagated contours outperformed the other methods on this group of patients. In addition, a reduction in segmentation accuracy can be seen as the number of patients is reduced to selecting just the most similar one. Nonetheless, using only half of the atlases instead of the entire database, the results do not show a significant reduction of automatic segmentation accuracy. The algorithm accuracy is comparable to inter-observer variability for the considered structure, thus allowing the procedure, once optimized, to be adopted for clinical use.

Having generated a reliable segmentation of organ at risks, after the attending physician and physicist review treatment volume, analyze dose-to-volume relationships, checking for inhomogeneity and violation of the maximum and minimum dose contraints, and decide fractionation scheme, the patient can start the treatment. For HN cancer, chemotherapy is usually paired to irradiation, to maximize tumor control and minimize normal tissue damages. Only Twenty of the HN patients were included in this study. For each patient, we selected one or more cases

(CTsim, CBCT at 40Gy/50Gy and CTrepl at 40Gy/50Gy), raising the total number of cases to twenty-eight. Mandible, parotid glands and Gross Tumor Volumes (GTV) were retrospectively drawn on CTrepl by a radiation oncologist. CBCT images are acquired to monitor tissue changes, but the acquisition of a replanning CT (CTrepl) is still preferred, though some efforts have been dedicated to address CBCT technological limitations. The purpose of this work was to develop and validate an efficient and automatic strategy to generate on-line *virtualCT* scans for HN ART cancer treatment on the basis of a new CBCT volume, thus dropping the need of a new CTrepl.

To generate the *virtual*CT, a multistage B-Spline deformable registration between CTsim and CBCT was performed following rigid registration (Figure 0.3). For the rigid coarse alignment, we used the fast rigid registration routine based on mutual information offered in Slicer3D. Non rigid registration was performed using Plastimatch (www.plastimatch.org). Our hypothesis was that deformable registration ideally compensates for non-rigid modifications, detected by in-room CBCT scanning and call for a treatment replanning, when overcoming a predefined clinical threshold. We did use one combination of parameters for this study, to make our approach independent from the tuning of the registration parameters and thus clinically feasible. The contours propagated on CBCT from CTsim (i.e. *virtual* contours) and from CTrepl (i.e. CBCT surrogates contours) were compared in terms of Dice Coefficient (DSC), Center Of Mass (COM) distances and Root Mean Square Error (RMSE) between contour points. Taking into account the resolution of the images, a good registration is indicated by a residual of maximum 3mm. We estimated the deformation between CTsim versus CTrepl in terms of structures volume changes as well as COM shifts, whereas the absence of deformation between CBCT and CTrepl was verified by performing an additional deformable registration to estimate the residual vector field length distribution within contoured structures. Finally, we computed Spearman correlation coefficient R (p-value = 0.05) between the residual and estimated deformation and RMSE values on the contours, looking for inaccuracies in the validation method and polarization in the algorithm, respectively.

Median DSC was around 0.8 for mandible and parotid glands, but only 0.58 for GTV, due to the fairly homogeneous background and to its small volume. Median COM distance and RMSE were comparable with image resolution. Critical issues raised in some patients, such as 6 and 7, mainly because of either sub-optimal registration performances induced by external features like immobilization mask or to macroscopic modifications between CBCT and CTrepl (i.e. different jaw position). Only patients 1b, 7, 12, 15 and 16a show a RMSE equal or greater than 3 mm, localized at transition between bone and teeth, whereas the different mandible positioning causes validation method failure for patient 7. In fact, the residual median deformation between CBCT and CTrepl was between 0.13 mm and 1.8 mm in the mandible, for all the patients but patient 7. For this patient, the median deformation measured 3.80 mm, due to jaw displacement in the CTrepl scan compared to CTsim. For both parotid glands, median DSCs were 0.78 and 0.79 and median COMs distances were lower than 1 mm along LL and AP direction, and lower than 1.5 mm along SI direction. Looking at RMSE for both parotids, we notice very few cases, in which the distance was above the 3 mm threshold. At first sight, the

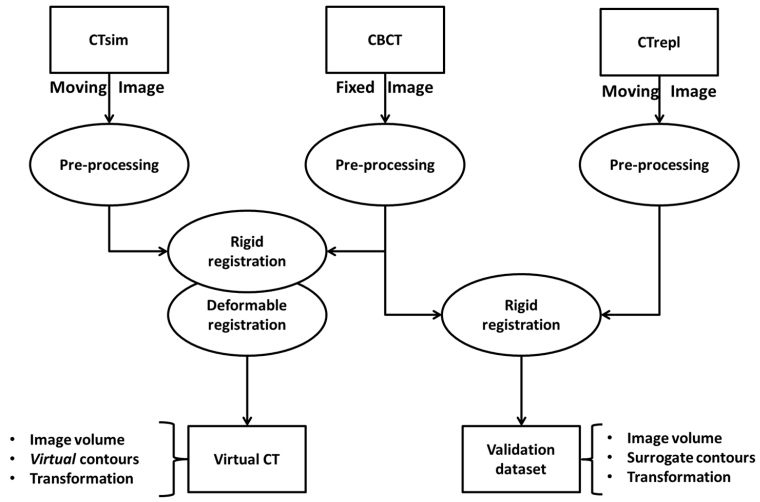


Figure 1

Figure 0.3: Workflow for generation of a *virtual*CT. B-Spline deformable registration was performed between CTsim and CBCT and compared with CTrepl, after having rigidly registered it on the same CBCT to compensate for coordinate system differences. Geometric validation was performed in terms of volume overlap, difference between center of masses and contour surfaces.

Table 0.2: Median and interquartile ranges for *virtual*CT and surrogate CBCT contours comparison with four different metrics (DSC, COM and RMSE)

	DSC [%]	COM [mm]	RMSE [mm]
Mandible	85 (5)	1.09 (1.71)	1.40 (0.67)
Left parotid	79 (6)	2.24 (2.23)	2.21 (0.52)
Right parotid	78 (6)	1.60 (2.34)	2.23 (0.40)
GTV	56 (16)	3.24 (3.30)	2.56 (1.35)

most critical structure appears to be GTV, for which median DSC was just 0.58 and RMSE violated the acceptance threshold in seven patients. Median results are summarized in Table 0.2.

The expectation of the mandible having the best results among all the re-contoured structures, because of its high contrast on the CT scan and the absence of non-rigid deformations, was confirmed by the quantitative results. With our strategy, we are able to compensate parotid glands shifts and shape differences, with accuracy comparable to the image resolution, allowing for a constant plan update between fractions, to avoid permanent side effects induced by irradiation, such as reduced salivary flow rate. The maximum of difference (around 1 cm) can be found at the frontal part of the parotids, where the separation between gland and masseter is not clear in CBCT, and in the lateral part of the contour, where the deformation is under-estimated because of errors induced by immobilization devices. For those patients in which the GTV was clearly contrasted the RMSE and COMs distances matched the axial voxel resolution of 1 mm (see for example patient 19) and DSC was greater than 0.7. Otherwise, deformable registration was

Table 0.3: Median and interquartile ranges of the distribution of accuracy of SIFT in terms of point-to-point distance between extracted on four couples of images (CTsim vs CBCT, CTsim_rig vs CBCT, *virtual*CT vs CBCT, CTrepl vs CBCT)

	median [mm]	50^{th}-25^{th} [mm]	75^{th}-50^{th} [mm]
CTsim vs CBCT	31.52	5.03	10.92
CTsim_rig vs CBCT	1.35	0.42	0.58
*virtual*CT vs CBCT	0.66	0.00	0.27
CTrepl vs CBCT	0.66	0.00	0.38

Table 0.4: Median synthetic non-rigid deformations for phantom study

	Right-Left [mm]	Anterior-Posterior [mm]	Superior-Inferior [mm]
1	0.06	0.52	0.02
2	0.06	0.52	1.51
3	0.62	0.52	1.41

not able to compensate for different shapes of GTV. In fact, for seven patients, the GTV was immerse in a rather uniform background, as it was already reported, and also the contoured lesion was very small, thus augmenting the probability of mis-positioning and of contouring differences as well as the negative correlation between structures volume and DSC value. Local tissue contrast enhancement, penalizing the sharp transitions, and mask refinement would limit inaccuracies in registration and further refine the deformation map. This latter can be used for cumulative dose evaluation as well as for treatment plan fast update. Dosimetric evaluation will be needed to compare virtualCT and CTrepl distributions and introduction of virtualCT concept into the clinic. As for the computational efficiency, the algorithm running times were on the order of 5 minutes with volumes of size 512x512x116 on a laptop mounting an Intel® Core™2 Duo 2.2GHz, with 2GB of RAM.

An alternative approach to evaluate DR performance is to compare the position of anatomic or external landmarks after registration has been performed. In clinical practice, most applications are based on the manually identification of anatomical landmarks, a time-consuming and operator-dependent procedure. Therefore, several algorithms have been developed to perform automatic or semi-automatic landmarks extraction and matching, with the goal of increasing the accuracy of detection and decreasing the cost in terms of time. First operators for the automatic extraction of points, such as Harris corner detectors, examined an image at a single scale, thus implicitly reducing the accuracy and stability already at the detection step. The invariance properties of a feature extraction method is very important, if we want to find salient interest points in images subjected to inter- or intra-patient variability. A method that can extract and match stable and trustworthy points at different scales between two images, is Scale Invariant Features Transform (SIFT). The features are detected as local extrema of a Difference Of Gaussian (DOG) pyramid and subsequently low contrasted (i.e. noisy) landmarks as well as overcrowed

edges descriptors are eliminated. To study the SIFT sensitivity to transformation, a phantom study was conducted, by applying different rigid transformations and, for the first time, non-rigid transformations. Finally, we apply SIFT to pairs of registered images of the replanned patients, with the goal of assessing DR performance. Selected deformations are summarized in Table 0.4.

Applying a rigid transformation to the image, the median for all five cases is below the voxel dimension, confirming the invariance of the method to this transformation. There are few outliers around 3mm, due to a wrong association of the keypoint in axial direction. In fact, the extraction of extrema from DoGs implies the calculation of a finite difference derivative, which concentrates the voxel information in its center. This implies uncertainty in the attribution of the axial coordinate to the keypoint, that justifies outliers values. SIFT can also cope with deformable vector field, without violating the upper bound on accuracy.

SIFT features extracted on original CBCT, CTsim rigidly registered on CBCT, *virtual*CT and CTrepl, were compared with the corresponding CBCT landmarks in terms of residual point distance and the accuracy of the associations is compared with other indices proposed in literature based on contours (DSC, COM and RMSE). The extracted features are located both on bony structures and in soft tissues and therefore provide a great alternative to manual point clicking. Nonetheless, the classical algorithm implementation relies on a descriptor based feature association, which does not consider any position relationship between the landmarks. The descriptor in fact is an histogram describing not only the properties of the voxel in which the feature is located, but also the neighbooring region. This lack of geometrical relative relationship can result in a wrong localization of the correspondencies between points. To overcome this limit, an hybrid SIFT - Iterative Closest Point algorithm should be tested.

Although much work is still needed for the definitive validation of the techniques developed during my PhD, we were able develop a robust and automatic segmentation method for CTsim, and to generate a *virtual*CT with its associated segmentation of OARs and GTV for ART purposes, based upon the information gained from the in-room CBCT scan and the application of a deformable image registration algorithm. We aimed at producing a reliable initial segmentation of CTsim and at sparing a CT scan (CTrepl) to the patient by modeling the inter-fractional deformations, thus avoiding any delay in the therapy as well as reducing the clinical workload. Computational time was reduced to the point that the strategies can be defined clinically feasible. Moreover, the gaussian weighting label fusion has the potential for becoming a competitive for robust CTsim structures planning. *virtual*CT contours are a representation of patient GTVs, parotids and mandible volumes at fixed treatment time points, as the calculated deformation is able to capture shape and position changes with an accuracy comparable to image resolution.

The developed approaches could therefore be appropriate for a clinical ART implementation, dropping the need of a CTrepl and improving the overall clinical workflow.

Part I

STANDING ON THE SHOULDERS OF GIANTS

1 INTRODUCTION

The object of this thesis is *Image Guided Adaptive Radiation and Hadron Therapy*, i.e. the development and optimization of clinically feasible strategies for the online update of radiation therapy treatment. In particular, we worked on automatization of the manual procedures both in the offline and online phases, aiming at developing fast and accurate automatic strategies for planning segmentation and *virtual* planning CT generation. The motivation for this work is to be found in the nature of the radiation therapy treatment itself, which involves multiple sessions and patient repositioning over several days. In this scenario, it is obvious that changes might be introduced that cannot be compensated for just by means of rigid transformations (i.e. by acting on treatment couch degrees of freedom). An important instrument becomes deformable registration, which is able to model all the non-rigid modifications. We will focus on Computed Tomography (CT) and Cone-Beam CT (CBCT) datasets, which are the most used in the clinic because of the relationship between tissue density (i.e. the information entailed into CT voxels) and beam energy deposit (i.e. dose delivery).

This dissertation is divided into five parts:

1. In chapter 2, we will introduce the base concept of modern radiation- and hadron-therapy, looking at radiobiological effects of current clinical protocols and at recently introduced technological innovations. We will describe in particular the use of imaging as support for physicians decision and guidance for beam planning and adaptation. The most widespread protocols for head and neck image guided adaptive radiation therapy will be introduced, using this as an example pathology for the main clinical need of technical instruments for treatment real-time monitoring and constant adaptation to the modified patients. Finally, the main issue raising in current clinical practice will be described, underlining the ones addressed in the present work.

2. In chapter 3, we will introduce the technological background of the present work and review its main components. The clinical applicability and current limitations of deformable registration will be discussed in details, tracing the path for the technical motivation beyond this thesis. We will describe the ill-poseness of the problem on the example of two main models of the non-rigid transformations (B-Splines and Log Domain Diffeomorphic Demons) and the issue raising when trying to assess the validity of the deformation and quantifying algorithm performances.

3. In chapter 4, possibile solutions for the limitations introduced in chapter 3 will be introduced. Having as main goal the enhancement of clinical applicability of deformable registration, we will study behaviour at convergence and suggest possible robust escape conditions on the example of Log Domain Diffeomorphic Demons CT-to-CT non-rigid registration. We will than propose an user-independent validation metric, studying its properties on

the example of CT-to-CT and CT-to-CBCT B-Spline registration. We further look into both Log Domain Diffeomorphic Demons and B-Splines, comparing their relative strengths and weaknesses.

4. In chapter 5.2, we will illustrate possible strategies for automatic segmentation at planning and online treatment adaptation on the example of head and neck district. The problem of generating a robust segmentation of structures of interest of a planning CT is first addressed, in response to the clinical need not only of less computational time, but also of generating a robust contour not affected by inter- and intra-operator variability for mandible, spine cord, eyes, optical nerves, parotid glands and brainstem. We developed a multi-atlas automatic segmentation strategy, relying on a database of CT images with its correspondent manual segmentation (i.e. atlases). Having generated a reliable segmentation of organ at risks, the patient can start the treatment, along which course image guided monitoring is foreseen to detect the modifications between diverse sessions. Here we will illustrate the developed and validated efficient and automatic strategy to generate on-line *virtualCT* scans to be used for head and neck adaptive cancer treatment.

5. In chapter 6, we will sum up results obtained in chapter 4 and 5.2 both for the technical components and clinical applications addressed in this work. The proposed innovative strategies and future improvements will be also reviewed.

2 IMAGE GUIDED ADAPTIVE THERAPY

Under the term Radiation Therapy (RT) we understand the delivery of medical ionizing radiation for cancer cure. It can be used as single treatment strategy, or after surgery and/or in combination with systemic therapies (i.e. chemotherapy).

The goal of RT is to achieve maximal therapeutic benefit, expressed in terms of probability of local tumor control with minimal side effects to surrounding structures. The radiation can be administered both placing sources into the patient (e.g. Brachytherapy) or, more often, as External Beam Radiation Therapy (EBRT). This second technique is used for about 60% of cancer patients and implies the delivery from an external device, i.e. a linear accelerator for photons or a cyclotron for particle therapy. In practice, this often equates to irradiate the tumor with a higher dose to improve local control whilst maintaining an acceptably low dose to other tissues, particularly those adjacent to the target.

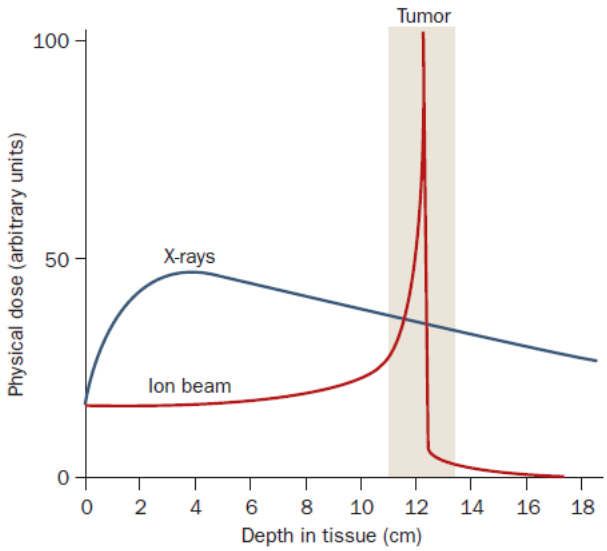
Sparing of normal tissues can be accomplished into two ways:

- Geometric avoidance, by directing multiple beams at the target, thus delivering a high dose at beam intersection (i.e. target) and a relatively lower dose outside of the intersection.
- Biological sparing, by fractionating the therapy over several weeks, irradiating daily. The tumor tissue lacks repair mechanisms to repair DNA damage from the radiation, whilst normal tissues can repair minor DNA damage. Therefore, by fractionating the treatment, normal tissues are provided time to repair, thus biologically sparing the normal tissue.

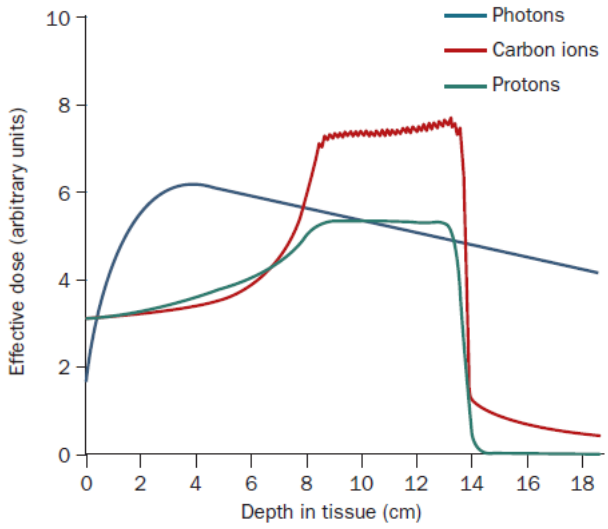
2.1 CONVENTIONAL RADIATION THERAPY

For cancer treatment, both photons and charged particles (e.g. electrons, protons or carbon ions) can be used. The main difference is given from the distribution of dose with respect to tissue depth (Figure 2.1). By contrast with X-rays, which show an exponential decrease in the radiation dose with increasing tissue depth, charged particles do not deposit most energy at the body's surface (Figure 2.1a), but rather just before they come to rest in tissue (i.e. in correspondance of the so-called Bragg Peak). The particles have very little energy beyond this peak, whose position is proportional to the energy of the particle. Since the Bragg Peak is narrow for particle at a given energy, to cover the whole tumor extent, a series of beams at different energies need to be superimposed, thus producing a Spread Out Bragg Peak (SOBP, Figure 2.1b). Variation in beam energy cause a shift in peak location, i.e. in delivery depth. Hadrontherapy is nonetheless mainly indicated for cancer inaccessible to surgery (e.g. eye, spine cord tumors) and/or are hard to treat with conventional RT.

For photons, a dose conformality similar to the one of charged particles can be achieved by irradiating the tissue from different directions, modulating beam in-

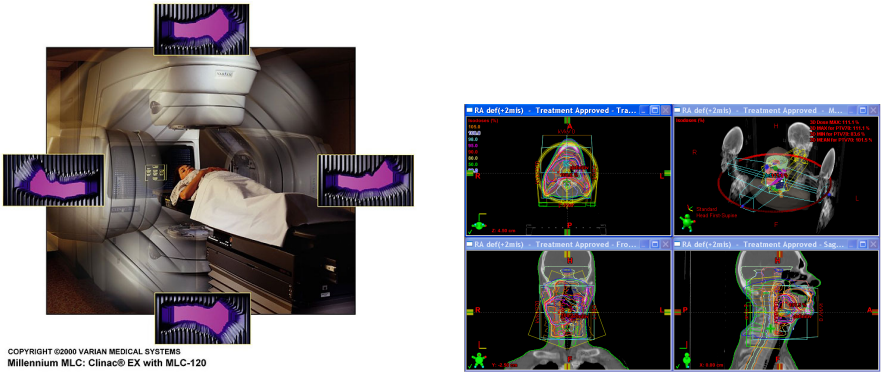


(a) Comparison between dose-depth distributions for X-rays and charged particles.



(b) Comparison between dose-depth distributions for X-rays and SOBP.

Figure 2.1: Dose-depth relationship for X-rays and particles. The Bragg Peak can achieve an almost punctual delivery of energy to the tissue in comparison to the X-rays distributions, which have a larger decay at patient surface (Panel a). To treat larger volumes, the Bragg Peak is enlarged by modulating energy in overlapping beams, generating the so-called SOBP (Panel b). The flatten plateau is achieved correcting the dose profile according to the radiobiological effect. [1]



(a) (b)

Figure 2.2: IMRT Treatment Planning on Varian Workstation. Panel (a) shows a Clinac with 120 Multi Leaf Collimator sculpting the beam according to tumor volume shape in the projection corresponding to the current angle of irradiation, while in Panel (b) the workstation view of the planned beams is illustrated.

tensities to vary delivery depth, and maximize dose at beams intersection. Intensity Modulated RT (IMRT) leads to excellent tumour control although a large volume of healthy tissues can be exposed to radiation. IMRT can include several directions individually shaped using a moving multileaf collimator, conforming the dose to tumor volume, and it is delivered by a medical linear accelerator mounted on a gantry and thus able to rotate around the patient (Figure 2.2).

Despite the extended time frame of fractionated radiotherapy (4–6 weeks), lesion and surrounding Organs At Risk (OAR) identification is carried out based on anatomical information derived from a single 3D anatomical Computed Tomography (CT) image data set acquired at the onset of treatment design and with the patient immobilized in treatment position. Planning can be divided into two different phases:

- Contouring phase, in which volumes are outlined on the CT scan by the attending radiation oncology.
- Physics optimization phase, in which the appropriate beam type (electron or photon), energy (e.g. 6 MeV, 12 MeV), geometry and collimation are selected.

Typical requirements of RT plan include homogeneity, conformity, avoidance, and simplicity [2, 3]. A conformity requirement is used to achieve the prescribed target dose, while minimizing the damage to OARs or healthy normal structure (i.e. avoidance requirement), evaluated by means of Dose-Volume Histograms (DVH), and to have uniform dose distributions on the target so that no portion of an organ is underdosed (i.e. *cold spots*) or overdosed (i.e. *hot spot*). These requirements can be enforced using lower and upper bounds on the dose, or approximated using penalization. Finally, simplicity requirements state that a treatment plan should be as simple as possible, to reduce the treatment time as well as implementation error.

Besides ensuring immobilization with proper casts and cushions, the patient is marked for repeated alignment with localization lasers in the treatment room to ensure correct macroscopic positioning of the patient at different fractions. There are two main sources of problems in the traditional system used for repositioning

Table 2.1: Inter- and Intra-fractional sources of errors on delineation of OAR and tumor volumes. [5]

Variations	Intra-fraction sources of errors		Inter-fraction sources of errors	
	Random	Systematic	Random	Systematic
in size	Physiological Process (e.g. breathing, peristalsis)	Physiological Process (e.g. heart beating, breathing)	Physiological Process (e.g. bladder filling, bowel gas)	Tumor reduction or swelling
in position relative to a fixed point in the patient				Weight loss or radiation edema
in patient position relative to treatment beam	Patient Movements		Daily set-up	Technical errors

in the clinical routine: on one side, the systems do not provide a fast and quantitative feedback to evaluate the quality of the positioning, and on the other side, each system requires the operator to be fully trained and expert in order the procedure to be successful. If we assume to have been accurate in positioning, still we would need to compensate for both rigid and non-rigid modifications, due both to physiological and non-physiological phaenomena. Intra-fractional (during a single treatment fraction) geometric change occurs during delivery due to breathing, cardiac motion, rectal peristalsis and bladder filling. Inter-fractional (day-to-day) geometric change occurs over the weeks of therapy, due to digestive processes, change of breathing patterns, difference in patient setup, and treatment response like growth or shrinkage of the tumor or nearby risk organs (e.g. parotids volume in head and neck treatment). A brief overview of principal modifications is shown in table 2.1. The sources of error can be divided into a systematic and random component. The systematic component is built up from all those errors committed in the preparation to the treatment, while the random component is mainly to be ascribed to the execution. van Herk [4] describes the effect of the random errors as blurring the dose distribution, whereas the systematic ones cause a shift in the cumulative dose distribution relative to the target (i.e. the highest dose region is moved to side from the target).

All of these changes are normally taken into account by population-based "uncertainty" margins around the target area, which may be excessive or conservative and are applied to the structures identified before the therapy begins, in order to safely irradiate tumor and ensure sparing of OARs.

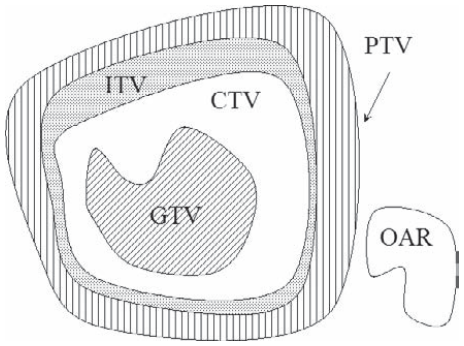


Figure 2.3: The different volumes defined for radiation therapy planning and their relation [3].

According to the International Commission on Radiation Units (ICRU) reports no. 50 and 62 [2, 3], the following volumes have to be defined when planning a treatment in any body district:

- GTV – Gross Target Volume: outlined on the planning scan/on medical images, it includes "*the gross palpable or visible/ demonstrable extent and location of malignant growth*" [2].
- CTV – Clinical Target Volume: it includes the GTV, the sub-clinical microscopic diseases and other areas (e. g. lymph nodes) that have to be eliminated because considered at risk. This can be defined by simply adding a fixed or variable margin to the GTV.
- ITV - Internal Target Volume: it includes CTV and an Internal Margin (IM). IM is designed to take into account the variations in the size and position of the CTV relative to the patient's reference frame (usually defined by the bony anatomy); that is, variations due to organ motions such as breathing and bladder or rectal contents [3].
- PTV – Planning Target Volume: this volume is defined to select the appropriate beam arrangements, taking into account all the possible geometrical variations [3] with IM. A Set-up Margin (SM) is also added to include the possible set-up errors or misalignment of beams.

Reducing set-up errors and understanding organ motion is therefore essential for shrinking the safety margins without compromising the tumor coverage. Both SM and IM are generally derived from experience or using values reported in literature, thus meaning that margins are site and patient specific, as their definition at the present mainly depend on motion magnitude.

2.2 IMAGE GUIDED RADIATION THERAPY

Besides laser coarse alignment, patient in-room position is monitored and corrected for at each fraction also with the use of X-rays projections and/or 3D image volumes acquisition. The monitoring through images does not only allow to detect global mis-positioning on treatment couch, but also modification induced by natural processes in the body and response of normal and target tissue to the treatment.

Broadly stated, IGRT involves any use of image to aid decisions in the radiotherapy process, such as what and how to treat, delineation of volumes, positioning, verification and monitoring of treatment outcomes.

Clinical imaging techniques can be loosely classified into anatomic methods, which measure the basic physical characteristics of tissue such as their density, and biological imaging techniques, which measure functional characteristics such as metabolism. Both of them are used to quantify and raise the accuracy of treatment delivery, but anatomical imaging plays a bigger role in contouring and planning, whether functional imaging is more popular for verification and follow-up of the treatment. In both cases, we can further divide image modalities in offline and in-room imaging, the latter grouping all those images that can be acquired while the patient is immobilized onto the treatment couch, just before, during or after the therapy. Biological imaging technique are generally offline techniques as they require injection of contrast media.

2.2.1 *Anatomical imaging in a radiation therapy clinic*

Offline anatomical image modalities include:

- CT, which allows 3D contouring, beams definition and analysis of the dose distribution decay, thanks to the linear relationship existing between tissue density and image gray levels (coded in terms of Hounsfield Units). The main advantages of CT are high spatial integrity and spatial resolution, very good delineation of bony structures. However, CT scan as well as other imaging modalities are subjected to artifacts due to breathing and other physiological movements, such as peristalsis [6, 5]. New protocols for CT scan have been developed in order to account and compensate for the presence of these movements, such as gated scanning or respiratory correlated CT acquisitons [7, 6].

- Magnetic Resonance Imaging (MRI), which has gained importance in particular in the treatments of Central Nervous System (CNS) and of the abdomen and pelvis, because of its ability to discriminate soft tissues. Also MRI scan is often fused with CT images to delineate the malignancy [8, 9]. Specialized scans, such as diffusion and perfusion MRI or functional MRI, enable physicians to make axonal and functional maps, which describe the structural connectivity of the brain and may allow a better definition of brain tumors and sparing of sensitive regions [10].

- Position Emission Tomography (PET), whose radioactive tracer comulation is used for malignancies identification and contouring. PET has a lower spatial resolution than CT images and contains no anatomical information about normal structures. The main weakness in this procedure is the usage of 18F-FGD (fluorine-18-labeled deoxyglucose) as tracer, whose injection causes discomfort to the patient and whose consumption can result evident also in some not pathological situations, such as brain or bladder. This problem was recently addressed thanks to the development of tracers which correlates with proliferation, rather than growth (e. g. FLT - fluorothymidine) . In addition, in

order to easily interpret the data, PET scan often needs to be fused with the correspondent CT images [11, 12]. The uncertainties of the fusion software are partly eliminated by the introduction of a PET/CT scanner, as the two scans are acquired one after the other, being the patient in the same position.

CT is and will be likely to remain the predominant volumetric imaging modality for radiotherapy. The role of CT in simulation is to provide a faithful representation of the object. When simulating treatment, the alignment of the patient can rely on skeletal alignment, often by means of a CT scout view, marking of the patient isocenter with a tattoo and/or overlaying radioopaque marker prior to the scan. An alternative approach is to identify the isocenter directly on the CT scan. In-room identification occurs after repositioning with external lasers, optoelectronic devices, and in-room imaging. For all approaches, careful attention to appropriate patient immobilization and recording the patient pose for reproduction of the set-up during treatment delivery are critical to accurate radiotherapy [13]. Eventual discrepancies in patient positioning are corrected acting on couch degrees of freedom [14]. An interesting comparison between external optoelectronic localization and kV in-room imaging is provided by Tagaste et al [15]. In dose calculations, after appropriate calibration, CT allows an appropriate CT number-to-electron-density conversion. 3D volume also forms the basis for the Digitally Reconstructed Radiograph (DRR), which is used as the template for verifying delivery using portal imaging or in-room or gantry mounted 2D kV imaging.

MR utilises the changing mobility of water molecules to obtain a wide range of manipulable image contrasts, complemented by an increasing range of contrast agents. Spatial resolution tipically ranges from 0.5-2.0 mm, with a slice thickness from 0.5-7.0 mm. MR often plays an important role in defining the location and local extent of disease. It provides a primary role in CNS disease, and in some diseases (e.g., prostate cancer [16] or head and neck cancer [17]). Recent developments in contrast agents and techniques may inform on tissue oxygenation [18]. MR simulation has been developed at a number of centres [19, 20], but still raises issues concearning geometric distorsions (e.g. those induced by inhomogenities in the magnetic field) and scarce informations about electron density in comparison with CT. Diffusion provides a sensitive method of identifying disease and following changes in cellularity, and it is showing promise in whole body surveys of metastatic disease, involved lymph nodes and brain tracts identification [21], with potential to better identify areas for treatment and sparing of critical functions, as well as for early prediction of treatment response and toxicity. Both for functional and anatomical MRIs, the most common approach is use image registration to combine these modalities with a more conventional CT scan. In fact, image fusion algorithms allow to combine together different imaging modalities with the aim of more accurate planning and margin personalization, merging the advantages and different insight provided by the diverse modalities.

PET is playing an increasing role in the radiation therapy treatment planning process since the emergence in year 2000 of combination PET-CT scanners followed by dedicated PET-CT simulators. There are three main areas of application of PET in radiation therapy:

1. Standardization of tumor delineation in comparison to CT with the use of 18F-FDG-PET/CT for a number of cancers that are well-imaged on PET.

2. Diffentiation of specific tumour biologic and microenvironmental features relevant to the radiobiology of the cancer by means of appropriate ad-hoc tracers. PET can provide prognostic information on the aggressiveness of the cancer as well as information for a "dose painting" treatment [22], by using such tracers (e.g., of hypoxia, angiogenesis, proliferation). Such application, however, is not yet in routine clinical practice.

3. Evaluation of tumour response based on imaging signals associated with tumour/tissue viability and metabolic changes, rather than anatomic changes acquired from the CT component of the PET-CT exam. Prospective monitoring of tumour response during a course of radiation therapy is probably neither practical nor useful given the confounds of radiation induced tumour cell death and inflammatory response. However, FDG PET scans performed at a consistent interval post-treatment may yield a reproducible measure of tumour response [23].

One of the first in-room imaging were portal images to verify patient set-up with respect to the radiation beam. Generally the position of the isocentre of the beams relative to the patient's anatomy, obtained from DRRs, is verified using either the actual treatment fields or two additional orthogonal fields. If the treatment fields are used, the portal image also provides information about the correct beam aperture or positioning of the blocks. Portal films are applied to verify patient set-up during treatment. A disadvantage of the use of the film technique is its off-line character, which requires a certain amount of time before the result can be applied clinically. For this reason on-line electronic portal imaging devices (EPIDs) have been developed for acquiring megavoltage images during patient treatment. In order to apply portal imaging in the clinic, local protocols have to be established stating the frequency of portal imaging, the criteria for acceptability of observed set-up deviations, and the responsibility for making decisions for changing the patient position. A detailed review on the clinical use of EPIDs for portal imaging purposes has been given by Langmack [24] and Herman et al [25] .

There are five major imaging methods employed in the systems that are currently available on the market:

- Ultrasound (US), in which conventional systems are employed in conjunction with a tracking system (optical or robotic) to allow US images of internal anatomy to be related to the isocentre of the treatment unit. These systems have the advantages of low-cost, easy integration within the RT process, and freedom from toxicity. Controversial points include the dependence of precision and accuracy on the skill and training of the operator, and the potential for errors arising from displacement of the relevant anatomy during the placement of the US probe [26, 27].

- Megavoltage CT, which is mainly used in conjunction with tomotherapy treatments and results in a loss in contrast-to-noise as compared to kilovoltage systems when equivalent doses are applied (~3 cGy to isocenter) [28, 29]. The geometric accuracy and quality of these images is high and should permit precise and accurate positioning of the patient. The use of the megavoltage beam provides accurate electron density estimation and reduces the magnitude of

artifacts associated with metal implants. The issues facing this technology include lower spatial resolution in the longitudinal direction (typically ~3 mm) and limitations on the ability to monitor motion during treatment.

- kilovoltage 2D radiography, which offer high geometric targeting precision and accuracy for high contrast surrogates of the target and normal tissues. The low imaging dose and high level of integration make it possible to perform multiple localizations during the course of a single fraction to verify correct and stable targeting of the treatment beam, despite its low contrast and 2D nature. The potential for very frequent monitoring (~15 fps) is made possible with the development of high-performance fluoroscopic modes, that can be used to monitor also intra-fraction changes [30].

- Cone-Beam CT (both kilovoltage and megavoltage), in which the accelerator itself is equipped with a source-detector couple. Provided that appropriate mapping of the system geometry has been performed, it is possible to employ filtered back-projection methods [31] to reconstruct an x-ray cross-section map (effectively a 3-D, high resolution CT image) of the patient's internal anatomy while positioned on the couch of the treatment unit. This allows the operator to detect, localize, and adjust the location of the internal anatomy with respect to the treatment beam just before the start of irradiation. These approaches have the advantage of soft-tissue detection and imaging of the patient in the treatment position. The greatest issues facing these systems arise from intra-acquisition motion, x-ray scatter on the detectors, detector lag, and limited field of view of the x-ray detector. These lead to variable image quality, inaccuracy of CT numbers, presence of shading, and truncation artefacts.

- CT-on-rails, a conventional CT scanner which is located near the accelerator used for EBRT, allowing to move the patient back and forth between accelerator and CT scanner.

In addition, it is worth mentioning the development of in-room verification systems, such as in-room PET dosimetry [32, 33] or proton radiography [34, 35, 36]. The additional dose due to the imaging is a growing concern in the medical community and is considered in the selection of the imaging technique.

A general scheme for IGRT is shown in Figure 2.4. The first step relies upon the acquisition of a CT scan in treatment position for planning. This is a joint work of the attending physician and physicist, who define both geometrical constraints (i.e. avoidance of OARs and target volumes) as well as dose and beams constraints (i.e. max/min dose delivered to each structure, beam angles and intensities and fractionation scheme). These two first phases can take up from one week to one month, mainly depending on clinical workload and complexity of the region to be irradiated. The next five phases constitute the real treatment. At each fraction, the patient is positioned on the couch, using both external tracking systems (e.g. lasers) and personalized immobilization devices. At all fractions or only at protocol specified ones, in-room images of the patient as on treatment couch are acquired. Agreement with planning CT is checked based on bony anatomy, soft tissues such as OARs or tumor volume directly, and finally rigid (i.e. rotations and translations)

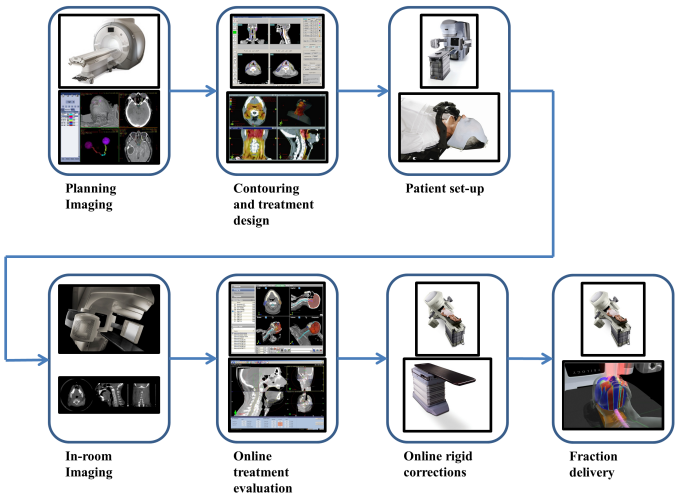


Figure 2.4: Workflow of a classical RT treatment. After the acquisition of volumetric images, the attending physician and physicist individuate structures and plan beams directions and parameters, in order to maximize the therapeutic benefit and minimize side effects to organ at risk. The patient is than treated according to institution and site specific fractionation schemes. At each session, the patient is accurately repositioned exploting both customized immobilization devices and in-room imaging. Rigid corrections are applied acting on the treatment couch degrees of freedom, after manual evaluation of the discrepancies between current situation and planned treatment.

corrections are computed. The couch position is subsequently updated and than the treatment is executed. The process from patient immobilization on is repeated at each fraction, until the prescribed total amount of dose is delivered.

2.3 IMAGE GUIDED ADAPTATION OF THE THERAPY (OF HEAD AND NECK)

The rapid development of IGRT and IMRT has provided tools for improving accuracy of patient positioning, target localization and conformality of dose distribution. Achieving a higher conformality does not necessarily imply better outcome, but may decrease toxicity in the OARs, especially for Head and Neck (HN) patients, as demonstrated by clinical experience [37, 38, 39]. Standard clinical protocols are mainly *one-plan-for-all-fraction* approach, thus relying on a single 3D snapshot of the patient anatomy at time of planning. In this phase, an accurate target definition is crucial, both to fully exploit the IMRT potential, delivering the highest prescribed dose to the highest tumor cell density (i.e. GTV volume), and to minimize dose to normal surrounding anatomy. Correlation between irradiate volumes, delivered dose and therapy induced side effects, such as, in HN patients, salivary production modification [40, 41], dysphagia and aspiration [42, 43], has been investigated.

The motivation for breaking out of conventional strategy comes from the violation of the assumption that the daily anatomy remains the same as initial planning, as suggested from clinical evidence of both disease and normal anatomy being significantly modified during a 6-7-week course of treatment, such as that of HN

patients. These changes include positive evolution of primary and nodal disease, alterations in normal glands and mucosa, resolution of post-operative soft tissue effects, and weight loss [44, 45]. Barker et al [44] reported median GTV and parotid volume loss of 0.2cc per treatment day in 14 patients imaged with CT scans three times a week during the whole therapy, as well as a median 7.1% weight loss from first fraction. These data were confirmed also by a megavoltage CT study from Lee et al [46], whereas Wang et al [47] reported an average volume loss of 20% at parotid glands after 3-weeks treatment, correlating significantly with delivered mean dose. Parotid glands volume loss and medial shift towards high doses areas was recently demonstrated also by Ricchetti et al [48], whereas tumor volumes do not generally follow predictable trajectories. There is also clinical evidence that patients may benefit from treatment update before or around the 25th fraction [49, 50], to avoid underdosing the target and/or overdosing the healthy tissues/organs, which may translate into compromised tumor control and/or increased adverse effects.

Image Guided Adaptive RT (IGART) is an approach to correct for daily tumor and normal tissue variations through online or offline modification of original IMRT target volumes and plans. An adaptive approach breaks the conventional sequential procedure of RT simulation, treatment planning and dose delivery, thus aiming at re-designing the approach from a *one-plan-for-all-fractions* to an *anatomy-of-the-day* scenario. IGART can be divided into two distict phases: monitoring and adaptation. The monitoring phase relies on the same technology used in standard IGRT, with the three main goals:

- modelling organ motion and correction of patient setup.
- distribution of the dose delivered to changing patient geometry.
- on-line monitoring of tumor response and normal tissue acute damage.

A general trend in IGART is to image the patient at a frequency comparable with or finer than the time scale of the anatomy changes thanks to the availability of in-room equipment and to use the data for as an online/offline feedback, to better direct the subsequent treatment. Daily in-room volumetric imaging is essential to both IGRT and IGART. In fact, even though no adaptation is foreseen, Region Of Interest (ROI) alignment on bony anatomy and only in exceptional cases on treatment volume is conducted daily or at specific fractions identified from an institution specific protocol. Current standard clinical protocols do not foresee fraction suspension, but rather new CT acquisition corresponding to in-room monitoring, even though inter-fractional changes do not urge at replanning.

Among all the technologies (see §2.2.1), CBCT stands as state-of-the-art in-room imaging technology, which could be exploited also as basis for treatment plan real-time adaptation. In contrast with the CT-based planning, CBCT is at its early stage in radio– and proton-therapy application and it is not yet considered a reliable tool for planning, mainly because of its HU inaccuracies [51, 52] due to larger contribution of scattered radiation and limited field of view [53]. CBCT has the advantage of being an instant 3D image of the patient as on the treatment couch, of delivering a lower dose to the patient and of being acquired in the same frame of reference of the treatment beam. Nevertheless, due to its 3D volumetric ability, good soft tissue

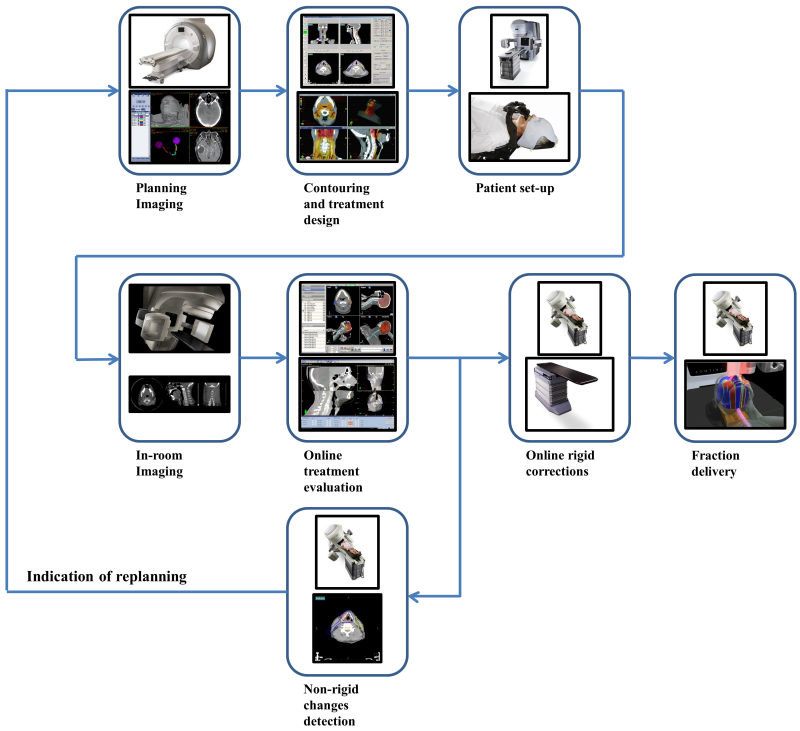


Figure 2.5: Image Guided Adaptive RT workflow. In comparison to the workflow in Figure 2.4, the adaptive procols foresee an active decision after the detection of changes in the patient that cannot be addressed acting on the treatment couch degrees of freedom, because they actually are non-rigid modifications. In the current clinical practice, if the change is bigger than an institution specific threshold, a second CT needs to be acquired and manually replanned. For HN district, the image volume acquisition does not generally involve contrast medium injection, thus limiting the physicians possibility of discriminating between different soft tissue.

contrast and low patient dose, CBCT also offers potential in evaluating the delivered dose distribution, online modification of the treatment plan and even online re-optimization. Jaffray et al [54] demonstrated that the system is capable of producing images of soft tissue with excellent spatial resolution at acceptable imaging doses, as well as the interfractional setup error reduction when using CBCT for soft tissue alignment. Barney et al [55] showed the accuracy of CBCT-based IGRT in comparison with kV fiducial imaging, despite the need of longer computational time, whereas Shi el al [56] reported suboptimal performances for gray-valued alignment against shifts calculated on fiducial markers alignment.

Aside from the full exploitation of the potential offered by the state-of-the-art technology, reliable tools enabling a real-time and physically consistent plan update, accounting also for those non-rigid shape modification which have occurred before the treatment day and/or on the treatment couch (e.g. breathing) and cannot be successfully compensated by shifts and/or rotation of the couch, are still missing.

A general IGART workflow is shown in Figure 2.5. With respect to Figure 2.4, a feedback line is added, representing the adaptation of the plan, based on a new CT acquisition, after the detection of non-rigid modifications of the patient. In theory, ART can occur at three different timescales: offline between fractions, online prior to a fraction and in real-time during treatment delivery. As such, the feedback line along which the image volumes are brought back to the treatment planning workstation, has to satisfy different time requirements. In addition, adaptation often implies dose recalculation and rarely proper re-contouring, given the restrictive time constraints imposed by the clinical workflow. Technical components of IGART strategies have been recently reviewed by Wu et al [57] and Schwartz et al [58]. First attempt of IGART clinical implementation by Nijkamp et al [59] for prostate patients demonstrate the potential effectiveness in reducing overdose to organ at risk and high-dose regions. Replanning clinical benefit was also assessed recently by Jensen et al [60] .

Given the hard constraints imposed by the clinical routine, a very popular strategy is the offline adaptation of plan and dose between fractions, thus implying either that the in-room image volume is not used until the next fraction of treatment or that a brand new CT volume has to be acquired. Among others [61, 62, 63, 64], Yan et al [65] developed a strategy to re-optimize treatment plan based on a target volume built from the first several fractions of image data. Wu et al [66] looked at possible margin customization and reduction as a conseguence of adaptation call, whereas Wu et al [67] used treatment re-optimization for compensating cold spots, i.e. underdosage area, in the volume from the next fraction. In addition, Woodford et al [68] demonstrated, using megavoltage CT, the clinical benefit of treatment plan adaptation if GTV volume decreases more than 30% during therapy course.

To overcome issues introduced from offline replanning, several researchers have been concentrating on online or hybrid adaptive paradigms, this latter meaning that the modifications occur at the same fraction in which the imaging is acquired only if major changes are detected, whereas all patients undergo online repositioning. Court et al [69, 70] developed an efficient method for dose homogeneity improvement based on the modification of MultiLeaf Collimator (MLC) positions by means of 2D registration of corresponding slices in planning and daily CT. Alternative approaches were proposed by Mohan et al [71] and Fu et al [72] by means of projecting anatomy in beams-eye-view and subsequently deforming intensity maps based on the new geometrical constraints. Fluence map and direct aperture modification on the basis of planning-to-daily non rigid deformation map was designed by Wu et al [73] and Feng et al [74] respectively. Perhaps the algorithm nearest to a clinically feasible implementation was proposed by Ahunbay et al [75] and relies on the subsequent optimization of MLC segments and monitor units.

Online replanning implies not only an efficient beam re-optimization and dose re-calculation but also a fast re-contouring method not subjected to intra- and inter-observer variability [76], thus being the uncertainity associated with single and multiple rater contouring respectively. In the head and neck area, treatment planning is especially complex and can require hours of tedious manual contouring work (2~4h). Fast and robust autosegmentation of structures of interest can be than achieved propagating original planning contours by means of the calculated transform with minimal manual input. In this kind of application, there is no *a*

priori determined relationship between a voxel's intensity value and the label that should be assigned to it, as we are seeking to label anatomical structures rather than tissue. Therefore, clustering methods, such as [77], cannot be employed, and a neighborhood relationships needs to be included in the segmentation process. A representation of these relationship is an atlas. Mathematically speaking, an atlas A is a mapping $A : \mathbb{R}^n \rightarrow \Psi$ from $\mathbb{R}^n$, n-dimensional spatial coordinates to labels from a set of classes Ψ and as such it is a special type of image, that is, a label image. A typical realization of an atlas is a manually segmented image and/or a computational phantom. Therefore, we typically have access not only to a spatial map of labels, the actual atlas, but also to a corresponding realization using at least one particular imaging modality. In order to segment a new image using an atlas A, we need to compute an anatomically correct coordinate mapping between them. When compared to other segmentation algorithms, such as level sets [78] or watersheds [79], this approach has a major advantage of having an *a priori* knowledge about the shape and the distribution of the segmented structures, that can be used to guide the segmentation.

In this context, image registration algorithms offer a huge potential to be exploited. Although details will be given in chapter 3, the main idea is to describe the modification occurred between two datasets by means of a mathematical transform with increasing number of degrees of freedom depending upon the complexity of the physiological phaenomena to be described. The idea is that, given an accurate coordinate mapping from the image to the atlas, the label for each image voxel can be determined by looking up the structure at the corresponding location in the warped atlas [80, 81].

Several strategies for atlas based segmentation have been developed, which can be classified based on templates numerousness [82, 83, 84]. The basic idea is to deform the structures manually outlined on one scan to predict the segmentation of a new patient on the basis of a vector field estimated by means of Deformable Registration (DR) [85]. This procedure suffers in first place from the fact that the chosen atlas might not be representative of test subject anatomy and in second place from DR inaccuracies. A popular choice for this algorithm is the employment of computational phantoms, whose segmentation is guaranteed to be optimal [86, 87]. A second possible strategy relies on a database of atlases, amongst which the most similar to the new patient is selected. Here, the main advantage is the ability of maximizing the probability of the chosen subject to be similar to the test patient, if the database is sufficiently varied. An approach involving the construction of an average atlas may overcome this issue, despite a degradation of the information due to the co-registration of diverse model cases, shown as diffuse blur, getting more intense at structures borders [88, 86, 89, 90, 91]. Several studies have demonstrated that label fusion outperforms the single-atlas approach, when the anatomical variability is too high to be accurately captured by mean statistics [92, 93, 94]. Therefore, the attention has been concentrating in constructing the so-called probabilistic atlases [95, 96, 97] or in developing strategies in which each subject in the database is registered to the target image independently and the resulting segmentations are combined [93, 98, 99, 100]. Groupwise registration and introduction of a prior into registration have also been recently explored, to boost algorithms further [101, 102, 103, 104, 105]. Most of these algorithms were devel-

oped for Magnetic Resonance Images segmentation, where boundaries between structures are generally evident and well defined. This image modality, although employed as support source of information, is generally not employed as in-room imaging in offline/online radiation therapy adaptation.

Besides pure segmentation problems, a fundamental issue in IGART is to obtain and guarantee geometrical alignment at re-planning. Wu et al [106] and Zhang et al [107] focused on CT-CT registration and tested a full replanning approach on a restricted set of patients. Nithiananthan et al [108] concentrated on Cone-Beam Computed Tomography (CBCT) guided intervention, demonstrating its feasibility and ability to capture the 3D patient anatomy during treatment on a cadaver phantom and patient study. Faggiano et al [109] and Hou et al [110] worked on multi-modal image volumes and were able to predict parotid glands position on mega voltage CT and soft tissue control points defined on CBCT respectively.

From this brief description of IGART, it is clear that one of the most critical point is represented by DIR itself, that not only needs to be fast and accurate, but also needs to reach convergence no matter which images and of which quality are employed. Validation of a registration algorithm also represents a critical issue in a clinical environment, in terms of possible metrics employable in a real clinical case. These questions will be addressed in chapter 4.2 and 4.3. We then will illustrate and validate a strategy for inter-patient multi-atlas segmentation of HN CT anatomical structures applicable both at planning and adaptation stage in chapter 5.1. Finally, we will present a strategy for generating a *virtual*CT volume to be exploited for fast online and/or offline adaptation of the HN treatment in chapter 5.2, with the double aim of dropping the need of a re-planning CT scan and to generate reliable virtual scan with corresponding geometry on which to re-calculate dose and re-optimize treatment.

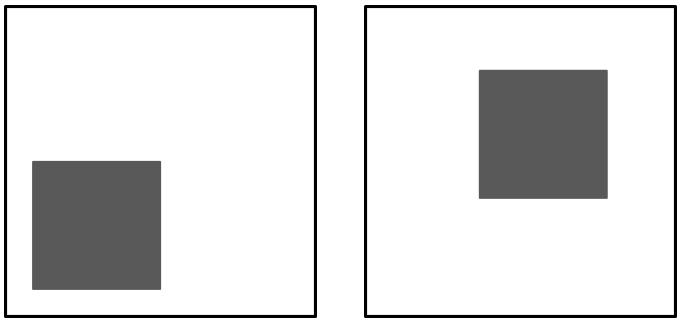
3 IMAGE REGISTRATION IN RADIATION THERAPY

Within the current clinical setting, there is a increasing request of integrating all information gathered with imaging, in order to be able to formulate a more complete profile of the patient. A key component of this process is Image Registration (IR), which is the transformation of the different sets of image data into one geometric coordinate system. IR is also often called *data fusion,* to underline its capability of bringing images not only into geometric alignment, but also literally to make them match.

3.1 AN INTRODUCTION INTO IR PROBLEM

Given a *reference* $I(x, y, z)$ and a *floating* image $J(x, y, z)$, IR goal is to find a suitable transformation ϕ such that the transformed *floating* becomes similar to the *reference*. In other words, given $I(x, y, z)$ and $J(x, y, z)$, we aim at optimizing ϕ so that $I(x, y, z) = J(\phi(x, y, z))$.

Though the IR problem is easy to state, it is generally hard to solve. As demonstrated in [111], The problem is *ill-posed* and its solution may *not be unique*. In fact, not only small changes of the input images or initialization can lead to completely different registration results, but also we may find several equally optimal solutions to fuse the input datasets. Suppose we want to register the two images in Figure 3.1, allowing for similarity transformations only. We can find several *equivalent* solutions (pure translation, a rotation followed by a translation, etc.), which cannot be sorted out without additional knowledge. As testified by the still vivid interest into algorithm developments and custom adaptation, each application has its own demands with respect to the level of accuracy required by the problem, to the definition of which kind of transformation model is allowed, and to constraints in terms of computational resources and time.



(a) *reference* image (b) *floating* image

Figure 3.1: Ambiguity of IR problems. How many ways do you have to fuse *floating* (b) and *reference* (a) image?

In general, we can divide IR applications into four groups, according to the relation between the datasets to be registered [112]:

1. Multiview analysis (images of the same scene taken from different viewpoints), e.g. reconstruction of 3D coordinates of points from 2D camera views.
2. Multitemporal analysis (images of the same scene acquired in different times), e.g. treatments follow-up, adaptive radiation therapy.
3. Multimodal analysis (images of the same scene acquired from different sensors), e.g. CT/PET, MRI/CT integrations.
4. Scene to model registration, e.g. comparison of digital images and anatomical atlas.

In addition to this classification, IR methods can be broadly clustered according to the type of inputs or of employed algorithm [113]. The available input sources include 2D and 3D images (e.g. portal and CT images) as well as 3D digitized points obtained for example by means of laser scanning.

Roughly speaking, sources could be divided into geometric objects and intensity data. Using sparse, accurately segmented, geometric data results in fast running times, and intra- and inter-modal registration are not distinct cases. This method has the advantage of being biologically valid, thus meaning that the result can be interpreted in terms of the underlying anatomy. Intensity-based approaches match intensity patterns using mathematical or statistical criteria, and thus do not use pre-segmented data, but may not be biologically consistent. Therefore, hybrid methods have been developed [114]. These algorithms mix together the geometric and model-based approaches with the intensity-based one, thus exploiting the advantages of both techniques. Dimensionality classification would further define IR as a mD-to-nD problem, where m and n are not necessarily correspondant (e.g. 2D-to-3D) [115].

The majority of commercially available registration methods include only rigid motion, thus implying that the relative distances between different points of the imaged object are not modified by the algorithm. From a strict mathematical point of view, this results in the description of the transformation by means of only translations and rotations. Rigid IR (RIR) concept is generally extended to affine transformation too, where the scaling can be modeled by a single factor. In contrast, non-rigid (Deformable) Image Registration (DIR) implies some stretching of the volume to be matched with the reference. DIR algorithms compute the deformation at each voxel using specific transformation models, e.g. diffusion, elastic or fluid. We can further divide the transformations into global and local. By definition, a transformation it is called global if it applies to the entire image, and local if subsections of the image each have their own transformations defined. Local transformation are rarely used alone, because they might compromise the local continuity and bijectiveness.

3.2 ALGORITHMS CONSTITUTIVE ELEMENTS

An IR algorithm can be decomposed into four main elements [**?**], which interact as in Figure 3.2:

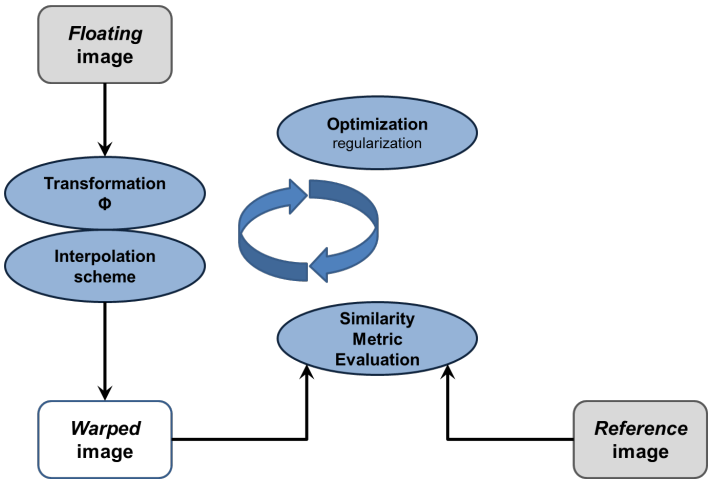


Figure 3.2: Elements of IR and their interaction in an iterative optimization scheme.

- a similarity metric (i.e. cost function to be optimized)
- a transformation model ϕ
- an optimization scheme
- an interpolation scheme

First of all, it is necessary to define a cost function (similarity metric), which provides a fitness value of the match of transformed *floating* (i.e. warped) and *reference* image. The transformation model defines how an image can be deformed to match another. They have two purposes: first, to control how image features can be moved relative to another to improve the similarity; second, to interpolate between features where there is no usable information. The model is then optimized with a specified optimization scheme. The output transformation estimated from the optimizer is used to generate the warped image, after proper regularization (if deemed necessary) and with the help interpolation scheme.

3.2.1 *Similarity measure*

The similarity measure describes how well the two images match. Ideally, this would imply the definition of a continuous monotonic function, having a single global minimum or maximum in correspondence of the optimal transformation parameters. In practice, this requirement is relaxed to just have a local minimum or maximum at some optimal parameters combination.

Geometric approaches build explicit models of conspicuous anatomical elements in the images, e.g. curves, landmarks or surfaces. The features can be classified as hard, i.e. fiducial markers which are positioned before imaging at certain spatial positions on a patient, or soft, i.e. retrospective landmarks deduced from the images themselves. The spatial location of these "anatomical" landmarks requires expert knowledge and/or sophisticated image analysis tools for automatic detection. The most distinctive element in these type of algorithms is definitely the definition of correspondences, which can be established manually, using a distance measure

Table 3.1: Common image similarity measures (from [**?**]). N is the number of voxels, $I_I(x)$ is the intensity of an image at voxel x, $I_J(\phi(x))$ is the intensity at corresponding point given by the current estimation of transformation $\phi(x)$; H_{I_I}, H_{I_J} and $H_{I_I I_J}$ are respectively the Shannon entropies of image I_I and I_J and the joint entropies of I_I and I_J.

Voxel similarity measure	Formula
Sum of Squared Differences	$SSD = \frac{1}{N}\sum_x (I_I(x) - I_J(\phi(x)))^2$
Correlation Coefficient	$CC = \frac{\sum_x (I_I(x)-\overline{I_I})\cdot(I_J(\phi(x))-\overline{I_J})}{\sqrt{\sum_x (I_I(x)-\overline{I_I})^2 \cdot \sum_x (I_J(\phi(x))-\overline{I_J})^2}}$
Mutual Information	$MI(I_I, I_J) = H_{I_I} + H_{I_J} - H_{I_I I_J}$, where $H_\cdot = -\sum_i (p_i \cdot \log(p_i))$

or evaluating some image properties and deeply affects the final result. Such algorithms optimize measures such as the average distance (L^2 norm) between each landmark and its closest counterpart, or iterated minimal landmark distances (e.g. Iterative Closest Point algorithm [116], Head and Hat algorithm [117, 118]). Curves and/or surfaces appear in literature under the form of snakes or active contours, whereas 3-D deformable models are sometimes referred to as nets.

The most frequently used similarity measures are summarized in Table 3.1. The sum of squared differences supposes that intensities of corresponding voxels of registered images differ only by Gaussian noise and therefore this measure is only appropriate for mono-modal registrations. Another possibility for single mode registration is to compute the cross-correlation coefficient, based on the assumption that registered images have linear intensity relationship and objects of interest are in the field of view of both images. Mutual information assumes, instead, only a probabilistic relationship between the intensities. It is based on the computation of the entropies of the intensities distribution and can be used also in multimodal registration [119, 120, 115] . This metric reaches its maximum when the two images are aligned, but it requires normalization to exclude dependence from initial alignment. Pluim et al [121] proposed the Normalized Mutual Information (NMI) as

$$NMI(I_I, I_J) = \frac{H(I_I) + H(I_J)}{H(I_I, I_J)}$$

where $H(I_I)$ and $H(I_J)$ are the Shannon entropies of the *reference* and *floating* images respectively and $H(I_I, I_J)$ is the joint entropy of the two input volumes. A drawback is that MI is not a least-squares criterion and the calculation of derivative information is not straightforward. To get a rough feeling of what these quantities mean, in Figure 3.3, we show a famous example by Pluim et al [122]. From left to right, the joint entropy distribution spread is increasing. The corresponding joint entropy values are (a) 3.82; (b) 6.79; (c) 6.98; and (d) 7.15. If they would have computed the corresponding value of MI, (a) would have the maximum value, while (d) the minimum.

Figure 3.3: Joint gray value histograms of a Magnetic Resonance Image with itself. (a) Histogram shows the situation when the images are registered. Because the images are identical, all gray value correspondences lie on the diagonal. (b), (c), and (d) show the resulting histograms when one MR image is rotated with respect to the other by angles of 2, 5, and 10, respectively. [121]

3.2.2 *Rigid transformation models*

An image coordinate transformation is called rigid, when only translations and rotations are allowed. Rigid transformations are a subset of the broader class of affine transformations. An affine 3-D transformation can be described using a single constant matrix equation, such as $y = Ax + b$, where x is a vector of the original coordinates and y is the output point. Therefore, the affine model is described as a linear transform followed by a translation (b). Matrix A describes the linear transform, that can include rotations, scaling, shear, reflection. If we restrict to rigid models, the only allowed transformations are rotations and translation and therefore, the matrix A is orthogonal (i.e. its columns form an orthogonal basis, with determinant equal to 1). In homogeneous coordinates, the transform matrix for a 3D registration problem can than be written as:

$$\overrightarrow{p_R} = \underline{\underline{R}} * \overrightarrow{p_I} + \overrightarrow{t}$$

where $\underline{\underline{R}}$ is a 3x3 rotation matrix, $\underline{t}$ is the translation vector and $\overrightarrow{p_I}$ and $\overrightarrow{p_R}$ are the original and rototranslated point.

Rigid registration has increasingly become a clinical standard, to calculate rotations and translations to be applied to the couch to cope for differences in patient positioning. In this case, the problem can still be seen as an optimization problem with nsix of freedom.

3.2.3 *Non-rigid transformation models*

Models used in medical image registration to describe an elastic deformation can be divided into parametric and non-parametric transformations. The parametric methods can be generally described by a quite limited number of parameters, whereas non-parametric methods typically feature a transformation function that is based on a vector per voxel describing the displacement of the point represented by this voxel. We will describe one model of parametric and non-parametric registration, whose performances will be discussed in chapter 4.

3.2.3.1 *Parametric transformation models*

The transformation between *reference* and *floating* image can be seen as a deformation field defined by a linear combination of a class of basis functions. A typical example is the use of *Thin Plate Splines* (TPS) [123], which can intuitively be understood thinking of a thin metal plate being bent by point constraints. This plate will arrange itself in a configuration in which the bending is smoothly distributed. The point contraints are landmarks individuated manually or automatically both on the *reference* image and on the *floating* image and voxel intensity are not explicitly used. The deformation field in this case is basically an interpolation function, that satisfies the condition

$$\overrightarrow{\phi}(\overrightarrow{i_k}) = \overrightarrow{j_k}$$

where $\overrightarrow{i_k}$ and $\overrightarrow{j_k}$ are corresponding sets of points defined on *reference* image $I(x, y, z)$ and *floating* image $J(x, y, z)$ respectively. The displacement field is calculated as

$$\overrightarrow{\phi}(\overrightarrow{i_k}) = \sum_{k=1}^{N} \overrightarrow{c_k} M(d) + \underline{\underline{A}}\overrightarrow{i_k} + \underline{b}$$

where $\underline{\underline{A}}\overrightarrow{i_k} + \underline{b}$ is an affine transformation, N is the number of control points, $M(d)$ and $\overrightarrow{c_k}$ are the chosen basis function depending on the Euclidean distance between points $d = \left|\overrightarrow{i_k} - \overrightarrow{j_k}\right|$ and its coefficients respectively. Intuitively, $\overrightarrow{c_k}$ is the bending force applied at each control point. For TPS, $M_{TPS}(d) = d$. This type of registration is also called landmark based registration and the most influent paramenter, is definitely the distribution of the control points. In fact, its dispersion deeply influences the vector field shape and the degree of localization of the displacements.

Another approach is to base the transformation on a number of control points arranged in a regular grid. An often used technique is *cubic B-splines*, which have the advantage of local support [124, 125, 126]. In this type of model, a change of a parameter only affects the transformation in a spatially limited neighborhood while other parts of the deformation remain unchanged. Hence, with respect to image deformation, only the relevant part of the image has to be resampled, which significantly improves the computational performance.

A *B-spline* is a piecewise polynomial with a uniform spacing between the control points $n_x n_y n_z$, which is extended to higher dimensions by a tensor product, thus in 3D:

$$\phi_x(x, y, z) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} \beta_l(u)\, \beta_m(v)\, \beta_n(w)\, P_x(i + l, j + m, k + n)$$

where P_x is the spline coefficient defining the x component of the displacement vector for one of the control points that influence the voxel. The basis function $\beta_l(u)$ is defined as follows (and similarly $\beta_m(v)$ and $\beta_n(w)$):

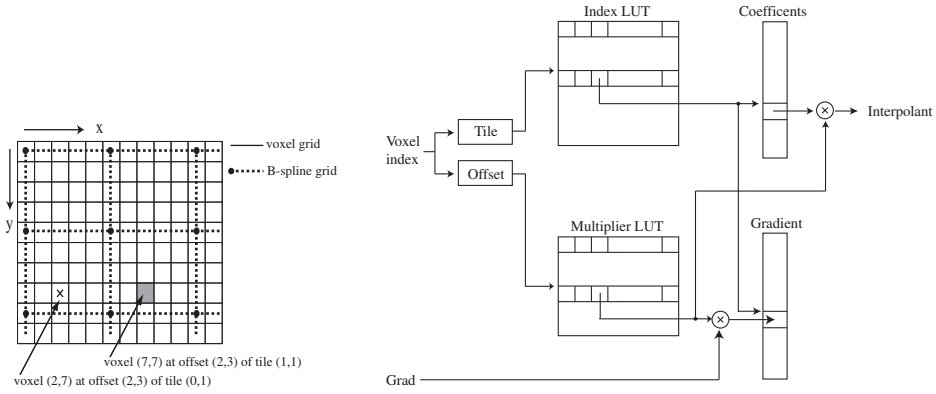


Figure 3.4: (a) Overlay between 2D image and a 5 × 4 control-point grid. The latter is aligned to the voxel grid. Marked and grayed voxel have the same relative offset within their respective tiles, and therefore they will use the same $\beta_l(u)\,\beta_m(v)$ value. (b) For aligned grids, lookup tables can accelerate the registration process by eliminating redundant computations. (from [127])

$$\beta_l(u) = \begin{cases} \beta_0 = \frac{(1-u)^3}{6} \\ \beta_1 = \frac{3u^3-6u^2+4}{6} \\ \beta_2 = \frac{-3u^3+3u^2+3u+1}{6} \\ \beta_3 = \frac{u^3}{6} \end{cases}$$

One efficient method to model the deformation field and to achieve reasonable computational cost, was developed by Wu et al [126] and succesively GPU accelerated [124, 128] and validated [127, 129, 130]. It is distributed in the form of open source multi-platform package under the name Plastimatch (www.plastimatch.org). It consists in aligning the *B-spline* control point grid to the voxel grid, thus partitioning into many equal-sized tiles of dimensions $N_x x N_y x N_z$, where N_x, N_y, and N_z denote the distance between control points, in terms of voxels, in the x, y, and z directions, respectively (Figure 3.4a). Basically, each voxel in the tile is influenced by the 64 control points in the neighbouring tiles, while the values of the basis functions depends only on the local coordinates within the tile.

Therefore, the coordinates of the tile, in which the voxel (x, y, z)falls, are:

$$i = \left\lfloor \frac{x}{N_x} \right\rfloor - 1 \quad j = \left\lfloor \frac{y}{N_y} \right\rfloor - 1 \quad k = \left\lfloor \frac{z}{N_z} \right\rfloor - 1$$

The coordinates in the tile, normalized in [0,1], instead:

$$u = \frac{x}{N_x} - \left\lfloor \frac{x}{N_x} \right\rfloor \quad v = \frac{y}{N_y} - \left\lfloor \frac{y}{N_y} \right\rfloor \quad w = \frac{y}{N_z} - \left\lfloor \frac{z}{N_z} \right\rfloor$$

This approach allows to pre-compute all relevant *B-spline* basis function products once, instead of recomputing the evaluation for each individual tile, because two voxels in two different tiles having the same offset in the tile results in the same basis functions product (Figure 3.4a).

In order to further speed up the computation, a complex data structure was developed (Figure 3.4b), further exploiting the symmetries derived from the alignment. The strategy relies on two LookUp Tables (LUT), called *Index* and *Multiplier* LUT respectively. The set of control point indices influencing each tile in the volume are stored in the *Index* LUT, while *Multiplier* LUT contain $\beta(u)\,\beta(v)\,\beta(w)$ products, which are the same for all valid normalized coordinate combinations. Therefore, for each voxel, the absolute (x,y,z) coordinates are used to calculate the tile it falls within as well as its relative coordinates in the tile. The *Index* LUT is than queried to derive the value of the 64 control points that influence that voxel, while from the *Multiplier* LUT, appropriate $\beta(u)\,\beta(v)\,\beta(w)$ products are derived. In comparison to the classical methods, this strategy reduces necessary computation to 64 queries and multiplications plus final accumulation for each voxel and vector field component, against 192 computations of the cubic polynomial B-Spline basis function and multiplications plus 63 additions for each voxel.

Once the displacement vector field is generated, the moving image is deformed and compared to the static image by means of the chosen registration metric. For example if we assume Mean Squared Error (MSE) as metric, the cost function becomes

$$S(I_I, I_J \circ \phi) = \frac{1}{N} \sum_x \sum_y \sum_z \left(I(x,y,z) - J(x+\phi_x, y+\phi_y, z+\phi_z)\right)^2$$

The optimization of this cost function, requires the computation of the derivatives at each control point coefficient value, thus

$$\frac{\partial S}{\partial P} = \sum_{(x,y,z)} \frac{\partial S}{\partial \phi(x,y,z)} \frac{\partial \phi(x,y,z)}{\partial P}$$

where the first term does depend just on the cost function S and as such is not depending on the B-Spline parametrization employed. For example, for MSE case, the equation becomes $\frac{\partial S}{\partial \phi(x,y,z)} = [I(x,y,z) - J(x+\phi_x, y+\phi_y, z+\phi_z)]\,\nabla J(x,y,z)$, which depends on intensity values in the *reference* image $I(x,y,z)$ and *floating* image $J(x,y,z)$ as well as on the current vector field estimate u. As the deformation field is updated at each iteration, $\partial S/\partial \phi(x,y,z)$ needs to be computed at every iteration. The second term instead depends just on B-Spline parameterization and thus remaining constant over the optimization process, such that

$$\frac{\partial \phi(x,y,z)}{\partial P} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} \beta_l(u)\beta_m(v)\beta_n(w)$$

Furthermore, given the built data structure, the values needed for $\partial \phi(x,y,z)/\partial P$ are already pre-calculated in the *Multiplier* LUT.

Since the displacement of the control points is not constrained during optimization, a folding of the points may occur resulting in an inconsistent topology of the deformation field. To overcome this difficulty, several techniques for regularization are proposed such as adding an energy term to the similarity measure or using multi-level B-Spline approximation techniques [131, 125]. The strategy chosen by Plastimatch developers relies on a second derivative regularization term, which can be computed either analytically or numerically. The regularization term

is added to the similarity term directly into the cost function and its amount can be directly controlled by the user via the parameter λ. Therefore the general form of the regularizated cost function c becomes:

$$c = S(I_I, I_J \circ \phi) + \lambda c_{RM}$$

where $S(I_I, I_J \circ \phi)$ is the intensity metric term, which penalizes images that do not match well, and c_{RM} is the regularization term, which penalizes non-smooth transformation. Typical values of λ range between 0.005 and 0.1. We define c_{RM} as the square of the vector field second derivative, thus:

$$c_{RM} = \int_\Omega \left(c_{RM,x} + c_{RM,y} + c_{RM,z}\right)$$

where $c_{RM,x} = \left(\frac{\partial \phi_x^2}{\partial x^2}\right)^2 + \left(\frac{\partial \phi_x^2}{\partial y^2}\right)^2 + \left(\frac{\partial \phi_x^2}{\partial z^2}\right)^2 + \left(\frac{\partial \phi_x^2}{\partial x \partial y}\right)^2 + \left(\frac{\partial \phi_x^2}{\partial x \partial z}\right)^2 + \left(\frac{\partial \phi_x^2}{\partial y \partial z}\right)^2$, $\phi = (\phi_x, \phi_y, \phi_z)$ is the vector field and Ω is the domain of the fixed image. In the analytical approach, the regularization is computed directly from the 192 control points of the B-Spline at each tile, which are formed into a vector $\nu = (\nu_x, \nu_y, \nu_z)$. The smoothness is computed using a 64x64 matrix K, as a quadratic form:

$$c_{RM} = \nu_x' K \nu_x + \nu_y' K \nu_y + \nu_z' K \nu_z$$

The matrix K is precomputed, and can be reused for each tile. The results are than numerically integrated tile-by-tile over the fixed image. In the numerical approach instead, finite differencing is used to compute the derivatives, which are than integrated over all voxels.

3.2.3.2 *Non-parametric transformation models*

Non-parametric transformations rely on physical properties and functions to guide the registration process. In this case the deformation is represented by means of a dense vector field (up to one vector per voxel) and needs to be subjected to regularization constraints. Generally, constraints are defined using an energy function computed from the deformation field. The DIR problem becomes the one of finding an equilibrium between external forces applied to the elastic continuum and internal forces that arise from elastic properties being modeled. The external forces are derived based on local similarity of voxels in the images. The internal elastic forces are based on the model itself.

Linear elastic energy is one of the most commonly used. Since the linear elasticity assumption is only valid for small deformations it is hard to recover large image differences with these techniques. Replacing the elastic model by a viscous fluid model allows large and highly localized deformations. The higher flexibility increases the opportunity for misregistration, generally implying growth of one region instead of distorting another [132]. For viscous models, the DIR solves the following Navier equation, which is a partial differential equation implicitly regularising the registration and limiting the types of deformation field allowed:

$$\mu \nabla^2 v(x,t) + (\lambda + \mu) \nabla \left(\nabla^T \cdot v(x,t)\right) + \overrightarrow{F}\left(\overrightarrow{x}, t\right) = 0$$

where v is the velocity field, F is the force on the object at x that depends on t, and μ and λ are Lames coefficients, determined from Young's Modulus (E) and Poisson's ratio (υ), through the equations

$$E = \frac{\mu(3\lambda+2\mu)}{\lambda+\mu} \qquad \upsilon = \frac{\lambda}{2(\lambda+\mu)}$$

Demons algorithm solves this PDE and was introduced by Thirion in [133]. It uses optical flow[1] to find a driving force at each point based on the intensity gradient of the image. The allowed transformations are described using a velocity field where each voxel has an associated deformation vector describing which voxel has to be pulled to the reference image. To regularise the flow a Gaussian filter is used.

An interesting and mathematically robust variation of the classical demons schema has been proposed by Vercauteren et al in 2009 [134]. The Log-Domain Diffeomorphic Demons (LDDD) combines the advantage of optimizing a diffeomorphic transformation with a computationally efficient framework, in which the optimization of the cost function happens in two different steps.

In order to cast the demons algorithm into a minimization of a well posed criterion, a hidden variable was introduced in the registration process: correspondences. The idea is to consider the regularization criterion as a prior on the smoothness of the transformation Φ. Instead of requiring that point correspondences between image pixels, a non-parametric spatial transformation s, be exact realizations of the spatial transformation Φ, one allows some error at each image point. The algorithm minimizes the cost function:

$$E(\phi, s; I_I, I_J) = \frac{1}{(\sigma_i^2)} S(I_I, I_J \circ \phi) + \frac{1}{(\sigma_c^2)} dist(\phi, s) + \frac{1}{(\sigma_T^2)} c_{RM}(s)$$

Where $S(I_I, I_J \circ \phi) = ||\, I_I, I_J \circ \phi \,||^2$ is the image disparity measure, $dist(\phi, s) = ||\, \phi - s \,||^2$ quantifies the similarity between the deformations Φ and s, $c_{RM}(s)$ is the degree of smoothness of the deformation and $\sigma_i^2, \sigma_c^2, \sigma_T^2$ balance the contribution of the three terms of the cost function. The introduced auxiliary variable allows to decouple the complex minimization into two separate steps: the first step solves for $1/\sigma_i^2 S(I_I, I_J \circ \phi) + 1/\sigma_c^2 dist(\phi, s)$ (i.e. correspondences) with respect to c and with s being given, while the second step solves $1/\sigma_c^2 dist(\phi, s) + 1/\sigma_T^2 c_{RM}(s)$, with respect to s and with c being given. The symmetric log domain is introduced by an efficient second-order minimization in which the warp Φ is represented with a smooth and stationary velocity field v. The relationship between the two is defined as $\phi(x) = exp(v)(x)$ which is a diffeomorphic deformation, with inverse $\phi^{-1}(x) = exp(-v)(x)$. At each iteration, an update velocity field u is computed minimizing:

$$E_{diff}(\phi, s; I_I, I_J) = S(I_I, I_J \circ \phi \circ exp(u)) + \|u\|^2$$

Therefore, the algorithm proceeds as follows:

1. Compute $u_{forward}$ and $u_{backward}$ that minimize E_{diff}.
2. Smooth with a Gaussian kernel K so that $u = \frac{1}{2} K \star (u_{forward} - u_{backward})$.

1 Optical flow is by definition the appearent motion from brightness patterns.

3. Update $v = v + u + \frac{1}{2}[u, v]$, where $[u, v] = |Jac^v(x)| \cdot u(x) - |Jac^v(x)| \cdot v(x)$ and $|Jac^v(x)|$ is the determinant of the Jacobian of the velocity field.

The update rule can be described in several different ways. At each voxel x, we compute [134, 135]:

$$u(x) = \frac{I_I(x) - (I_J \circ exp\,(v))\,(x)}{\|G\,(x)\|^2 + \frac{\sigma_i^2(x)}{\sigma_c^2}} G\,(x)$$

where $\sigma_i(x)$ is estimated from image noise, σ_c controls the maximum step length and $G\,(x) = -\frac{1}{2}\left(\nabla_x I_I + \nabla_x\,(I_J \circ exp\,(v))\right)$ is the symmetric gradient evaluated usign both images.

Optical flow methods (and therefore the demons too) rely onto two basic assumptions:

1. **Brightness constancy**, according to which in the local neighboorhood the intensities of the two images do not change over time, thus $f\,(x + u, t + \delta t) = f(x, t)$. For small displacements, using first order Taylor expansion, his assumption leads to an optical flow constraint $\nabla f \cdot u = 0$, with ∇f are the partial derivatives of the images.

2. **Small motion**

To satify these two constraints, histogram matching and pyramidal approach is often applied.

3.2.4 *Optimization scheme and strategies*

To find the best transformation, we seek parameters that minimize an appropriate cost function. The optimization problem in the case of non rigid registration becomes more complicated with the increase of the flexibility of the transformation. This means that there are more parameters to choose from compared to the rigid case. Moreover, in some cases, the algorithm converges to a solution, but this is not physically meaningful. Seeking the optimization of a parameter set is a problem solvable with standard methods, such as Steepest Gradient Descent, Conjugate Gradient Method, Levenberg – Marquart method or Limited-memory Broyden – Fletcher – Goldfarb – Shanno (L-BFGS) .

The Steepest Gradient Descent Method is an optimization strategy, where one takes steps proportional to the negative of the gradient of the function at the current point. The Conjugate Gradient Method is an algorithm for the numerical solution of particular systems of linear equations, which have to be symmetric and positive definite. The conjugate gradient method is an iterative method and it can be applied to sparse systems which are too large to be handled by direct methods. It is often applied for solving partial different equation but it can be used also to solve unconstrained optimization problems. Levenberg – Marquart algorithm is a numerical method used to find a solution to non linear least squares problems. It is particularly used for curve fitting and non linear programming. Compared with the Gauss-Newton Algorithm (GNA), it is more robust, thus meaning that even if starting off from very far off the minimum, it finds a solution. If instead the initial

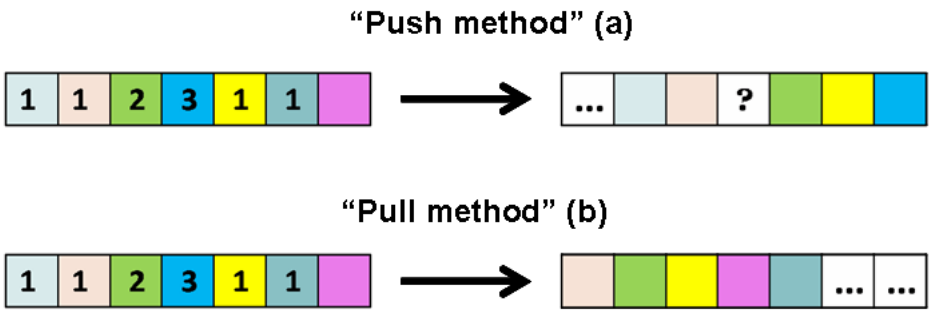


Figure 3.5: (a) The intensities are pushed to the target, located according to the deformation field. (b) The intensities are pulled to the origin of the displacement vector from the box individuated by its tip.

condition are reasonable and the function is well-behaved, it tends to be slower than GNA. L-BFGS instead is an efficient Newton optimization method applicable to bound-constrained optimizations. No matter what is the optimizer chosen, a robust and efficient stopping rule shall be defined. We address this problem in chapter 4.2.

Given the ill-posiness of the problem, generally people never start off directly with a DIR registration, but the vector field is rather modeled in first instance by means of a rigid or affine registration, thus adjusting for global location and scale. This also helps preventing folding and vortices in the final deformation, which make it non-physical. A very popular approach, for increasing accuracy of registration and at the same time reduce computational cost, is to use a *coarse-to-fine* approach. The idea is to build a pyramid of resampled images and to let the algorithm rely on new features as the resampling rate approaches 1, thus as the resolution approaches the initial voxel size. Therefore the deformation approximated at a coarser stage is than used as input for finer stages, thus reducing the risk of being stuck in local minima at the highest resolutions. Often the downsampling of images is combined with an image filter like a Gaussian filter, to enhance even more the differences in the features at the diverse scales[131].

Detecting convergence and guaranteeing that the chosen DIR stopping condition is always optimal for the considered images is a fundamental but rather complicated in translating these algorithms into the clinical practice. We analyze, compare and propose a series of stopping conditions and method to compare those objectively in chapter 4.2.

3.2.5 *Push and Pull warping method*

Once the deformation has been computed, it is applied to the image. There are two possible warping methods (Figure 3.5):

1. *Push* method
2. *Pull* method

In (1), the voxel intensities are moved to the target pixel, individuated by the coordinates in the vector field. If we look at a 1D example as in Figure 3.5(a), representing

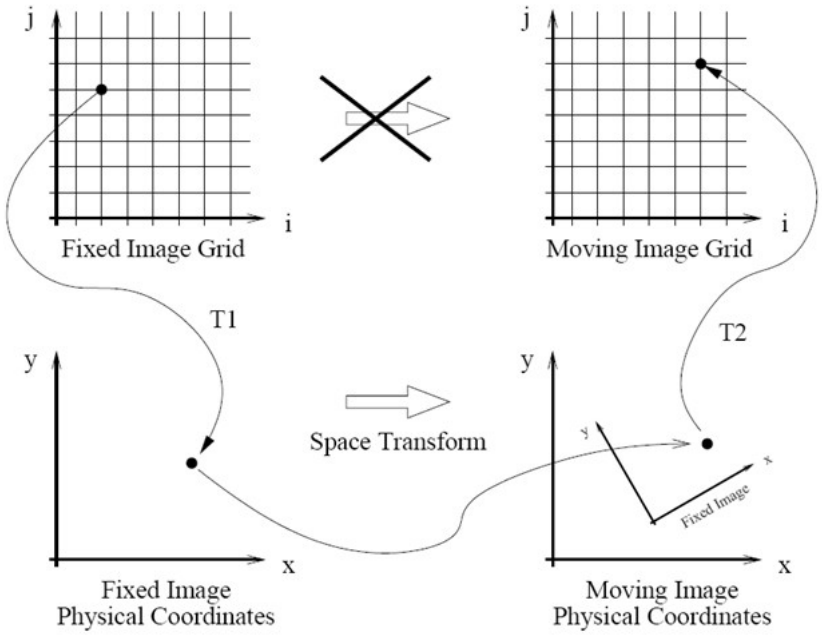


Figure 3.6: The registration problem in Plastimatch and in ITK (from [136]) This computation involves taking the coordinates of the pixel in the image grid, mapping them into the physical space of the fixed image (transform T1), mapping those physical coordinates into the physical space of the moving image (transform to be optimized), then mapping the physical coordinates of the moving image in to the coordinates of the discrete grid of the moving image (transform T2), where the value of the pixel intensity will be computed by interpolation.

a non rigid deformation, the numbers in the box stand for the displacement to be applied and the different colors for the intensities. Let us consider the case of light blue, green and blue boxes. The light blue one will be pushed one, the green two and the blue three to the right. As in the original image the green box is the third and in the warped needs to be the fifth, we have no information to fill the fourth box, which is therefore called *hole*. On the contrary, the blue box has to become the seventh in the final image, exactly as the greyish one. In this case, we have an *overlap* and again no mean to decide whether the last box has to be blue or greyish. The pull method (ii) solves the issues generated by the holes and the overlap, because the intensities are pulled to the voxel on which the vector is applied. In other words, the application of (ii) implies that if we are standing on the first box on the right in (b), the first box on its right will be pulled onto it, while the yellow will be attracted on the third one. In this way, there will be neither holes nor overlaps in the output image. After the warping procedure, the voxel left empty on one side of the image, will be filled with a default value.

Obviously, the whole estimation of the transform and warping happen in the world coordinates, i.e. physical coordinates, not in the voxel space. In summary, the computation of the transformed intensity involves (Figure 3.6) [136]:

1. Taking the coordinates of the pixel in the fixed image grid, usually called the "(i,j)" coordinates, and mapping them into the physical space of the fixed image (transform T1).

2. Mapping the physical coordinates from step 1 into the physical space of the moving image (transform to be optimized).

3. Mapping the physical coordinates of the moving image into the coordinates of the discrete grid of the moving image (transform T2 in the figure), where the value of the pixel intensity will be computed by interpolation (generally linear or nearest neighboor).

3.3 VALIDATION AND/OR PERFORMANCE ASSESSMENT OF A REGISTRATION METHOD

Several flavours of registrations, cost function, optimization algorithm and validation metrics have been tested, according to the different input datasets [137, 138**?** , 139, 115, 140, 141, 142, 143]. The open question is how to make this approach sufficiently robust and fast to be plunged into the clinical routine.

Besides ensuring optimal convergence in any run (see chapter 4.2), DIR results need to be quantified and the algorithms itself need to be validated. Finding a comprehensive metric that could classify registration results is of primary importance.

Validation does not only mean measuring the accuracy or error in a specific portion of the image, but should also refer to the analysis of the following features [115]:

- Robustness or stability, thus the request that small variations in the inputs should result in small output variations.
- Repeatability, thus having the same behaviour for the different clinical cases.
- Complexity, thus the need of charachterizing its computational cost, which has to comply clinical requirements in terms of time and resources.
- Clinical usability and relevance.
- Precision, thus the error associated with an ideal input.

The validation of a DIR algorithm tipically occurs at multiple levels. First of all, a series of synthetic experiments is conducted to study the properties in a controlled environment, in which a gold standard is available. Phantoms applications and pre-clinical studies on cadavers are key elements to the second step towards real applications in which no gold standard is generally provided, but landmarks points are tipically provided. Finally, the algorithms are tested on real clinical patient datasets in retrospective and prospective studies.

The validation and performance quantification metrics can be classified into point-, surface-, intensity- and volume-based methods [144]. Point-based methods mainly imply the computation of 3D residual error on chosen landmarks [145, 146, 147]. In clinical practice, most applications are based on the manually identification of anatomical landmarks, a time-consuming and operator-dependent procedure. Therefore, several algorithms have been developed to perform automatic or semi-automatic landmarks extraction and matching, with the goal of increasing the accuracy of detection and decreasing the cost in terms of time [148, 149, 150, 151, 152, 153]. These methods will be described in chapter 4.3.

Each metric presents its own advantages and disadvantages and can be more or less suitable depending upon the assigned task [154, 155, 156, 157]. As all the metrics analyze different aspects of the registration problem, generally a pool of metric is employed to assess quality of the results. Recently, Rohlfing [158] demonstrated on a paradigmatic case how classical validation indices are not reliable for validation and accuracy estimation. In chapter 4.3, we will illustrate a method for extraction of salient features from medical images relying on gaussian filtering and compare it to classical performance indices.

Part II

WE CAN SEE A LITTLE FURTHER

4

DEFORMABLE REGISTRATION: FROM THEORY TO CLINICAL PRACTICE

Adaptive Radiotherapy (ART) is a possible strategy to compensate for changes and improve delivery, where patients are re-imaged and re-planned several times during the entire treatment time (see §2.3). It mainly relies on DIR, which has become a fundamental tool to analyze target motion and measure physiological changes by fusing imaging diverse modalitites or time series volumetric imaging. Several algorithms have been developed for DIR [159, 8, 132, 114, 111, 112, 160, 113, 115, 119, 120, 142, 161, 147], but the lack of a gold standard in clinical practice still compromises their validation. We previously discussed possible strategies for DIR (see chapter 3), describing in particular one parametric and one non-parametric model of DIR. Here we will try to evaluate the performance of two popular implementation on the example of HN CT-to-CT DIR, at the same time comparing both approaches to a commercial DIR (see §4.1).

Despite the rapidly growing interest in DIR and the increasing number of works illustrating its clinical feasibility and relevance, yet DIR is not a standard clinical practice. In fact, clinically feasible DIR algorithm must ensure accuracy (i.e. the resulting deformation should exactly reproduce the real patient one), flexibility (i.e. sufficiently robust for all the patients), ease of use (i.e. tool learning curve should be fairly steep) and a reasonable computational cost, both in terms of time and memory.

One of the main limitation to the clinical use of DIR is the need for reasonable computational time and a robust parameter set. Several efforts have been made towards the first objective [162, 163], thanks in part to the introduction of powerful and affordable GPUs [164, 128]. Besides the most advanced technologies and the most complex algorithm, a good stopping condition that prevents extra optimizer iterations when convergence is reached would decrease the computational effort, at the same time not introducing complex pre-calculations and not requiring extra dedicated hardware. The choice of what should be employed as escape condition and how it should be computed with respect to previous values is, however, not obvious given the complexity of the problem. We worked on development of a robust stopping condition on the example of a non-parametric method developed by Vercauteren et al [134, 135], i.e. LDDD Registration. In this chapter we will review the results and try to provide operative guidelines for clinical usage (see §4.2).

The second limitation to systematic deployment is the lack of precise guidelines for performance evaluation in the clinical practice. Given the absence of a ground truth deformation field describing the transformation between two input dataset, the validation or performance assesment is confined to evaluation of the quality of reproduction of image charachteristics, features and/or anatomical structures. In the third part of this chapter, we will therefore concentrate on reviewing classical indices used in literature and propose an alternative approach based on automatic

salient features extraction. Finally, we will trace guidelines for choosing a metric appropriated to specific DIR tasks.

4.1 COMPARISON OF A PARAMETRIC AND A NON-PARAMETRIC DIR ALGORITHM

From modeling and compensation of setup uncertainties, the focus of research has gradually moved to deformations induced both from anatomic and physiological motion (e.g. bladder deformation, respiratory motion) and from therapy response (e.g. tumor growth or regression) for the implementation of adaptive radiotherapy. In addition, inter-subject probabilistic segmentation of anatomical volumes has become more important, with the end goal of supporting physician contouring. Each of these applications depends on Deformable Image Registration (DIR). Amongst these, two of the most popular algorithm are B-Spline based methods (also called free-form methods) and the demons algorithm, introduced by Thirion (see §3.2.3.2 and [133]). We analyzed an efficient implementation of B-Spline registration by Sharp et al [127, 129, 124, 126, 128] contained in Plastimatch (www.plastimatch.org), an open-source cross-platform suite for IR in radiation therapy, and one of the several flavours of demons algorithm, called LDDD by Vercauteren et al [135].

The testing dataset is composed by twenty patients, treated at European Institute of Oncology (Milan, Italy) for local HN malignancy (mostly rhino- and oropharynx). The patients were imaged at planning (CTsim) and after 40 Gy (CTrepl_a) and 50 Gy (CTrepl_b) according to the institution specific IGART protocol. CTsim was acquired after contrast medium injection and mostly during the concomitant chemotherapy treatment. Contrast medium was not used for CTrepl_a and CTrepl_b. The total number of cases analyzed is 28. We performed DIR between CTsim and each of the CTrepl for each patient using Plastimatch, LDDD and a commercial DIR algorithm MIMvista (MIMsoftware, Cleveland, Ohio, USA) implemented in a clinical workstation . Manual segmentation of mandible, parotid glands and GTV from attending physician were associated with the corresponding CT scan and were used as main mean of performance assessment. Therefore, we apply the computed deformation to binary rendered CTsim contours, obtaining sets of warped structures, which we compare with the corresponding manual segmentation on CTrepl, in terms of Dice Similarity Coefficient (DSC) [165], Mean Surface Distance (MSD) and 3D residual vector length between Center Of Masses (3DCOM) of warped and reference image volumes. Dice Similarity Coefficient (DSC) is defined in particular as:

$$\mathrm{DSC} = \frac{2\,|A \cap B|}{|A| + |B|}$$

where A and B are the two structure volumes to be compared. Its value goes from 0 to 1, the latter indicating the perfect overlap between the considered labels. MSD is instead calculated as median of the distance calculated via Danielsson Distance Map at each voxel, while 3DCOM is simply the modulus of the vector defined by means of coordinate-wise subtraction between the COM positions in the warped and reference image. Statistical comparison between the three methods was carried

out for each structure according to each index by means of Bonferroni corrected Friedman test at 5% confidence level.

Plastimatch B-Spline were regularized with $\lambda = 0.03$, but a single set of parameters were used for all the patients. These parameters were experimentally determined on the basis of residual Mutual Information between output warped and reference image. LDDD required the implementation of a compensation algorithm as the length of patient scans might not be the same between planning and replanning phase, but no patient specific tuning is foreseen. LDDD stopping condition was based on harmonic energy variation analysis throughout the iterations.

The results are presented in Figure 4.1, 4.2 and 4.3, in form of median and percentile of the overall distribution for the 20 patients. The mean volume overlap described from DSC provides a global indication of performance, but is not able to accurately describe finer local errors. In fact, if we look at Figure 4.1, we note that all algorithms were able to recover the deformation between CTsim and CTrepl, and few or small differences are present. Statistical difference between the methods was reported in the case of GTV and parotid glands, whereas the small variations in mandible DSC value were not significative. In particular, LDDD performance was significantly different in all soft tissues structures, whereas Plastimatch was significantly different from commercial software only for the case of GTV. We also report a larger variability, expressed in terms of difference between the 75^{th} and 25^{th} percentile, for LDDD on most structures. If we look at Figure 4.2 and 4.3, we notice that here diversity between the three methods are more evident than for DSC. Friedman test confirmed that LDDD might be slightly less favourable than Plastimatch B-Splines implementation and its performances are further from the commercial algorithm. Generally better results are obtained for bony structures, where open-source algorithms are fully comparable with commercial ones and both demons- and B-Splines-based registrations can recover mispositioning. The performances of the other structures mainly depend on relative tissue contrast, which for istance definitively influences GTV propagation. Contrast enhancement as well as histogram matching would prevent LDDD failures and serve as anchor features for vector field determination in all implementations. COM distances are very influenced from singlThe clinical protocol foresees the acquisition of a CT_repl only if modification bigger than 5 mm are detected on in-room imaging by means of visual inspection. Therefore, te voxel difference and therefore their quantitative evaluation should be compared to other indices, such as MSD. All algorithms however, reached a reasonable level of accuracy for all patients (i.e. less or around 3 mm of error and DSC comparable with inter-rater variabilities reported in literature [94, 109, 166, 167]).

This study served mainly as validation of both approaches, and did underline the feasibility of both approaches. One main disadvantage of LDDD is that, although regularized, it provides no stopping condition criteria at convergence and has few parameters to modify its behaviour. B-Splines instead are more flexible, but may also benefit of a more robust stopping criteria than just looking at absence of improvement in the metric value in the last ten iterations. We therefore moved on in studying behaviour and robustness of diverse escape conditions.

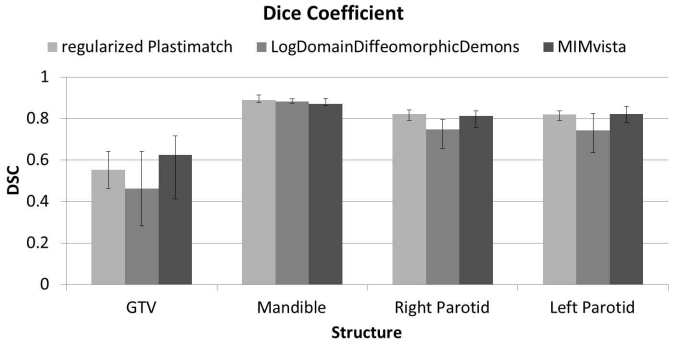


Figure 4.1: DSC at GTV, mandible and parotid glands for the three analyzed implementations (B-Spline Plastimatch, LDDD and MIMvista). Regularized B-Spline DSCs are statistically equivalent to MIMvista results, whereas LDDD performed slightly worse than the other two implementation. Only for GTV, also Plastimatch demonstrated to be less accurate than MIMvista.

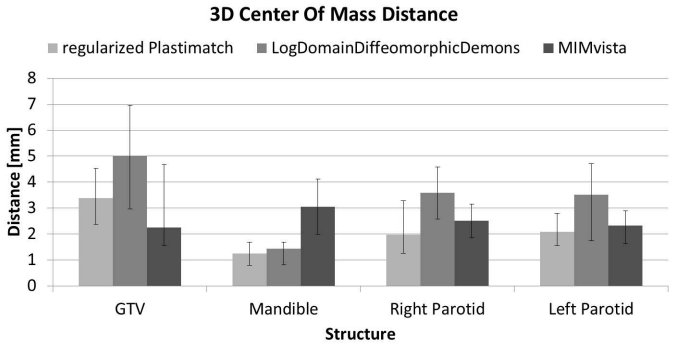


Figure 4.2: COM difference at GTV, mandible and parotid glands for the three analyzed implementations (B-Spline Plastimatch, LDDD and MIMvista). Regularized B-Spline outperformed both MIMvista and LDDD in all structures but GTV. The results achieved on GTV are anyhow around resolution level and as such acceptable. LDDD fails to map GTV and parotids correctly, whereas mandible contour differences in MIMvista translate in a higher COM mispositioning.

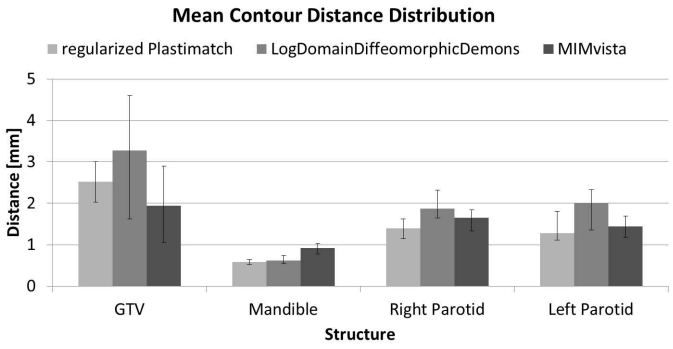


Figure 4.3: MSD difference at GTV, mandible and parotid glands for the three analyzed implementations (B-Spline Plastimatch, LDDD and MIMvista). The most of the differences can be found on the GTV, which definitely is the most difficult structure to discriminate because of its poor constrast with respect to surrounding soft tissues.

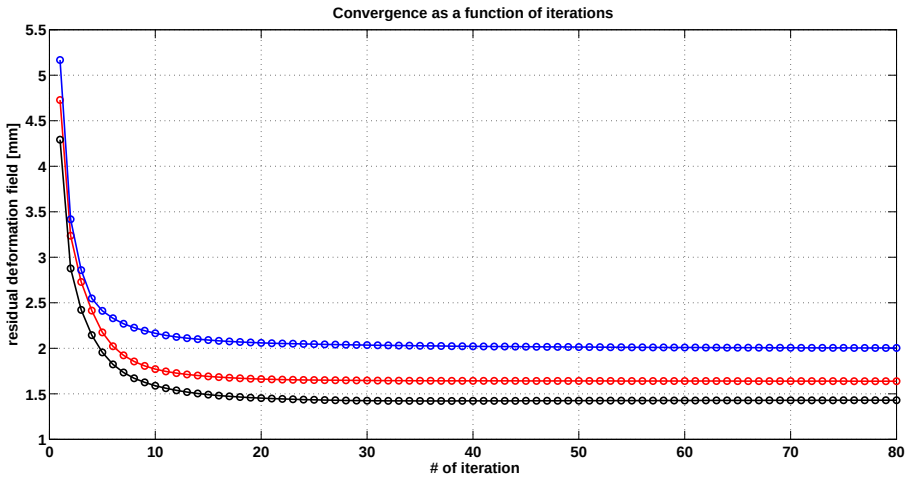


Figure 4.4: Mean of the modulus of the residual deformation field difference as a function of iteration number in a coarse stage of DIR. The initial mean deformation can be seen in the first iteration. We note that for all the three deformations, the trend of convergence is similar, but at each iteration the residual value is different depending upon the initial value of deformaiton.

4.2 RANKING OF STOPPING CRITERIA FOR CLINICAL DIR

The most basic stopping condition is to terminate the algorithm after a predetermined number of iterations. Although very simple, this criteria is only weakly related to the actual convergence and relies rather on user level of expertise, as well as on task and image quality. What can actually happen is that the number of iterations that seem to be optimal for one patient scan, it is actually beyond or much before optimal number of iterations needed to reach the convergence. For example, in Figure 4.4, we show a typical convergence graph for recovery of three different synthetic deformation applied to the same image. We note that the behaviour of the graph are similar, but at any given number of iteration the mean of the modulus of the residual deformation field is different, thus confirming that the solely number of iterations is not robustly ensuring convergence.

To relate the stopping criteria with registration convergence, we compare four different stopping rules for DIR, each focusing on a different aspect of the registration algorithm. One of the possible metric is using similarity measure as escape condition [168], without taking into account how this is related to the vector field. In [169], a metric based on updates to the deformation field is used, which we now extend to the velocity field. Also in [134], the authors used the harmonic energy and the number of vector field voxels with negative Jacobian elements to assess algorithm performance. Here we propose to use these as an escape condition. Experimental evaluation of these stopping criteria were performed with five synthetic non-rigid deformations applied to clinical quality CT image volumes of a dosimetry phantom and three HN patients.

Criteria based on image intensities are generally aimed at observing changes in similarity between *reference* image $I(x, y, z)$ and *floating* image $J(x, y, z)$ at current iteration t with respect to the previous iteration. Using Mean Square Error (MSE)

of image intensity as the image similarity measure, we define a threshold ε such that

$$\frac{MSE_{current} - MSE_{previous}}{MSE_{previous}} < \varepsilon_{MSE}$$

Here, ε_{MSE} is an user-controlled parameter, which quantifies the percentual error reduction between iterations. If the registration is converging (i.e. the warped image at t is more similar to $I(x, y, z)$ than the one at some previous iteration), the ratio would be negative, while it goes to zero if the algorithm has plateaued or converged. The ratio will be positive if the chosen step direction is suboptimal, but the registration can be allowed to continue if the degree of non-convergence stays below the user acceptance threshold ε_{MSE}.

We next consider convergence based on the Harmonic Energy (HE) of the deformation field. HE is defined as the average over all voxels of the squared Frobenius norm of the Jacobian of the vector field. As with MSE, we compute:

$$\frac{HE_{current} - HE_{previous}}{HE_{previous}} > \varepsilon_{HE}$$

HE is expected to increase with convergence, and therefore the ratio should be always positive unless the optimization is diverging. $\varepsilon_{HE}$is again the measure of relaxation allowed by the user, quantified as percentual increment.

The third quantity we consider is the use of the Jacobian of the update field as stopping criteria. Because the Jacobian matrix of an unstable deformation vector field is negative, we compute the ratio between the number of voxels with Jacobian below a slightly positive threshold to the total number of patient voxels. This quantity is than compared with a user defined threshold, thus

$$\frac{\#\left(Jac\left(x\right) < \tau\right)}{N} < \varepsilon_{JAC}$$

where $\tau = 10^{-3}$for our purposes and ε_{JAC} is set by the user.

Finally, we analyze the Quantity of Update (QU) between iterations. QU is defined as:

$$QU_t = \frac{\sum |dv_t|}{\sum |v_{t-1}|}$$

where t is the iteration identifier, v_{t-1} is the velocity field at a previous iteration and dv_t is the update field at current iteration. QU_t decreases with convergence and therefore we can compute

$$\frac{QU_{current} - QU_{previous}}{QU_{previous}} < \varepsilon_{INCR}$$

At convergence, $QU_t \to 0$ and as a consequence, the ratio decreases below the allowed percentage, thus stopping the iteration process.

Each of the metrics, besides the Jacobian, requires the comparison with a previous Stopping Condition Value (SCV), i.e. of what we called $\bullet_{previous}$. Its choice is of critical importance, as $SCV_{previous}$can be considered the actual descriptor of the shape of the convergence curve. Table 4.1 summarizes the methods we analyzed in terms of convergence speed.

Table 4.1: Current and previous Stopping Condition Value (SCV) tested in terms of convergence speed. These formulations are used in the final computation of the chosen stopping criteria. For conditions from D to F and stopping criteria HE, the minimum shall be substituted with a maximum, given HE definition.

	Current Value	Previous Value
A	SCV_t	SCV_{t-1}
B	SCV_t	SCV_{t-3}
C	SCV_t	SCV_{t-5}
D	SCV_t	$\min(SCV_{t-\alpha})$ with $\alpha = [1;6]$
E	$\min(SCV_{t-\alpha})$ with $\alpha = 0,1,2$	$\min(SCV_{t-\alpha})$ with $\alpha = [1;6]$
F	$\min(SCV_{t-\alpha})$ with $\alpha = 0,1,2$	$\min(SCV_{t-\alpha})$ with $\alpha = [3;6]$

4.2.1 *Dataset*

We first tested our approaches on an image of a RANDO® dosimetric phantom acquired on a clinical CT scanner, using supine setup and clinical acquisition protocols. The volume acquired is 512*512*123 voxels and [0.94, 0.94, 3] mm element spacing. After the phantom study, we analyzed three head and neck patients acquired with the same CT scanner and image resolution of the phantom. An example slice per patient and of RANDO® is presented in Figure 4.5. We chose two patients with few image artifacts (patient 1), one with a moderate degree of artifacts in the dental region (patient 3) and one where the dental capsules severely compromised the Hounsfield Units accuracy (patient 2). On patient 1 scan in addition, the chemotherapy tube is still in place (red circle in Figure 4.5).

Five artificial non-rigid vector fields were used as ground truth deformations. Each of the deformation field was obtained simulating the distribution of left-right, anterior-posterior and inferior-superior components with a smooth, localized Gaussian. The chosen Gaussians are centered at different voxels in a matrix of size 128x128x128 and have diverse standard deviation and amplitude, as reported in Table 4.2.

After re-sampling the Gaussian deformations on the image grid, they are superimposed to define a continuous and smooth vector field, that we aim to recover by means of LDDD registration (see §3.2.3.2). We applied the same deformations both to original phantoms (J_{ph}) and patients (J_{pa}) image volumes, obtaining their warped correspondent volume (I_{ph} and I_{pa}respectively). The five warped phantoms (I_{ph_i}) are shown in Figure 4.6, to get a rough idea of the applied deformations. The apparent swapping of *floating* and *reference* images is motivated by the warping strategy used to generate I_{ph} and I_{pa}. In fact, this happens according the *Pull* method (see §3.2.5) and as such, applying the synthetic vector field to the original volumes, it is as if these were the *floating* images of a previous registration, while the warped scans become the *reference* images.

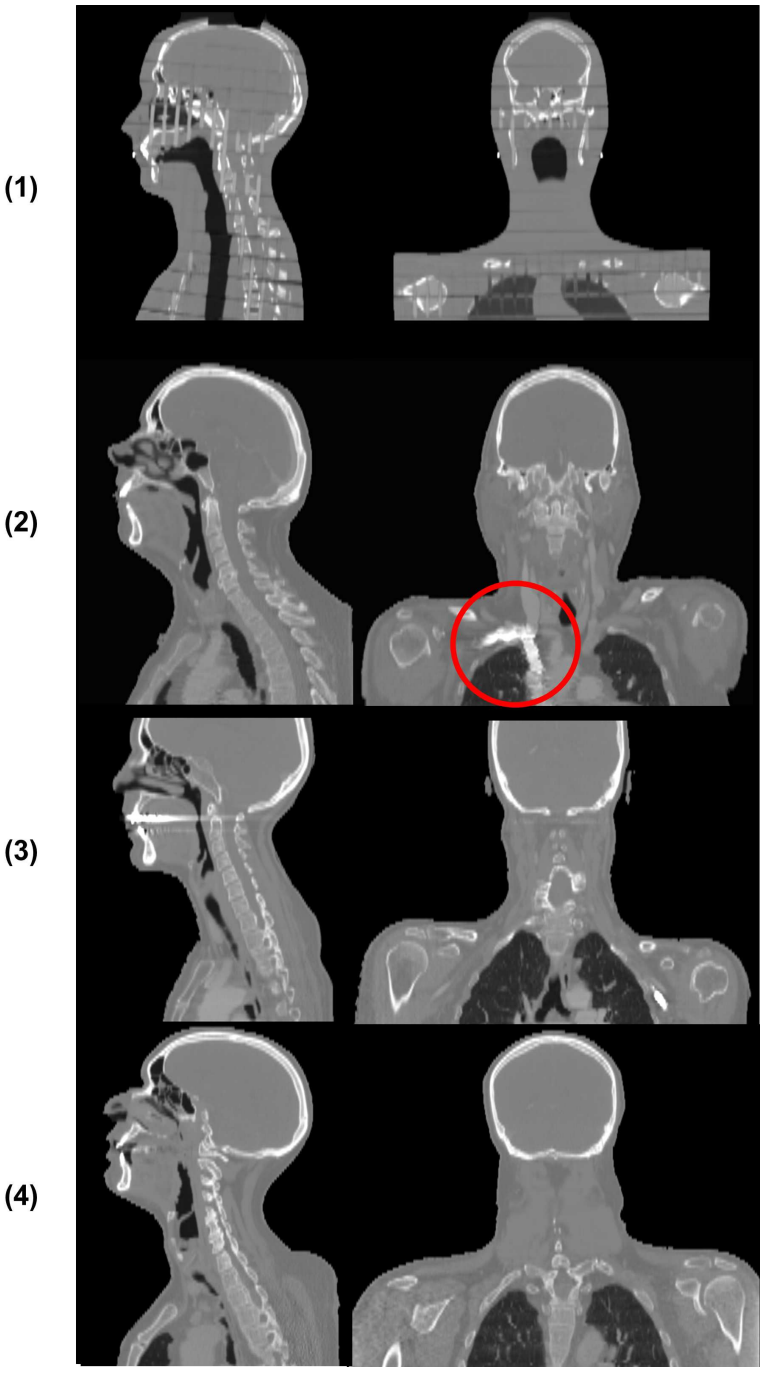


Figure 4.5: An example slice per patient. Panel (1) shows one slice of RANDO®, in which we notice distinctly the several layers it is composed of. In the red circle, chemotherapy tube for patient 1. Note the presence of metal artifacts in patient 2 and 3.

Table 4.2: Parameters of the Gaussian distribution used to generate five artificial non-rigid 128*128*128 vector fields. Besides dislocating the centers and changing the standard deviation, we further weighted the distributions by a scaling factor to enhance either one of the components (Relative Weight).

Test number		Application voxel	Radius [# voxels]	Relative Weight	Standard Deviation [# voxels]
1	x	[30 30 30]	15	5	12
	y	[50 50 50]	30	5	18
	z	[80 80 80]	10	100	36
2	x	[60 60 60]	25	1	12
	y	[50 50 60]	30	1	18
	z	[60 60 55]	45	100	36
3	x	[60 60 60]	25	10	12
	y	[50 50 60]	30	10	18
	z	[60 60 55]	45	1000	36
4	x	[60 60 30]	40	0.15	10
	y	[30 30 60]	20	0.15	15
	z	[60 60 90]	30	0.15	22
5	x	[60 60 90]	30	0.1	22
	y	[60 60 90]	30	0.1	22
	z	[60 60 90]	30	0.1	22

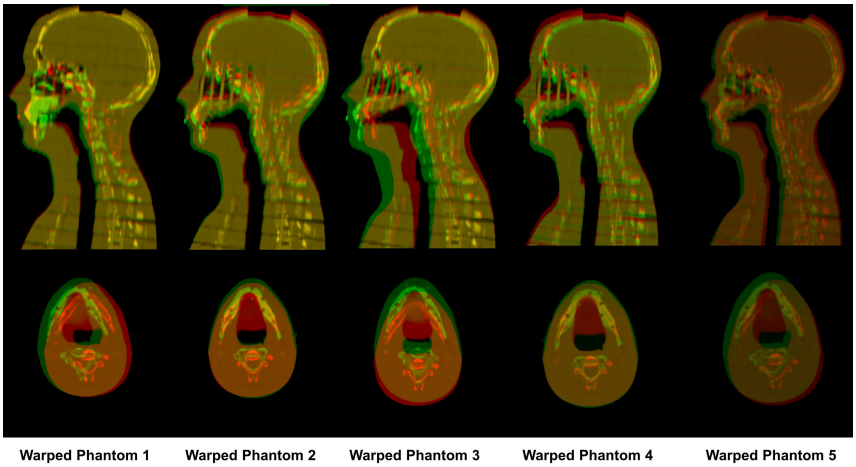


Figure 4.6: Overlay of warped (green) and original (red) phantom. The same deformation were used to warp the patient images.

4.2.2 *Experiments*

We register J_{ph} and J_{pa} to each of their five warped volumes (I_{ph_i} and I_{pa_i}) aiming at reproducing the same deformation vector fields we simulated. We run 500 iterations at one-eighth of the image resolution (i.e. a coarse registration stage).

We looked at the Normalized Mutual Information (NMI) before and after registration between the images to detect any possible failure in the registration procedure, based on the fact that the NMI between two registered images should be bigger than the NMI between two misaligned datasets.

To compute the convergence values, we exclude background (air) voxels from the computation, because, on one side, we are only interested in patient volumes and, on the other side, nothing can be said from deformation calculated in homogeneous areas. At each iteration, the algorithm computes the deformation vector field and we compute the stopping condition values described in Table 4.1.

To rank the stopping condition performance, we compute the vector length of the residual deformation at each iteration, by means of component-wise subtraction of the ground-truth displacements from the demons vector field. Given the image resolution, we set the threshold on the vector length residual error to 1.875 mm (i.e. 2 voxels at full resolution, or half a voxel at one-eigth resolution), and assess the number of iteration required to obtain the desired level of accuracy ($t_{required}$). We combine this information with the SCV at threshold iteration and we call it $SCV_{critical_i}$ with i=[1;N] (N=5 in this case). Ideally, if a given escape condition is optimal for the problem, $SCV_{critical_i}$ would match $t_{required}$ for all experiments. We then rank the performance of the stopping criteria in terms of number of extra iterations needed for all experiments to reach the best $SCV_{critical_i}$. In addition, we rank the SCV calculation strategies in terms of convergence speed and stability.

4.2.3 *Results*

4.2.3.1 *Convergence properties*

Figure 4.7 illustrates the median of MSE, HE and QU values over the five warped phantoms at each iteration and for each of the computational strategies described in Table 4.1. To compare with phantom study, Figure 4.8 shows the median of the median value of the stopping condition over the five different deformations for the three patients included in this study. As the values are medians of medians, the slight instabilities are smoothed out. We notice that all of the strategies are potentially good for the assigned registration problem, but some of them are not as computational efficient as the others. For example, strategies B, C and F are suboptimal, as they require more iteration than e.g. E. Moreover, B and D require a longer initialization (respectively 3 and 5 iterations) to achieve a reasonable metric value, thus possibly leading to unnecessary iterations. Strategy A might be undesirable, because it monitors just the immediately previous iteration. Strategy D is a good candidate for MSE (Figure 4.7 and 4.8, Panels 1) and QU (Figure 4.7 and 4.8, Panels 3) but it is unstable for HE (4.7 and 4.8, Panels 2). Strategy E seems to capture the best tradeoff between speed of convergence and stability, in particular if we look at the HE case and we analyze the patient median values separately.

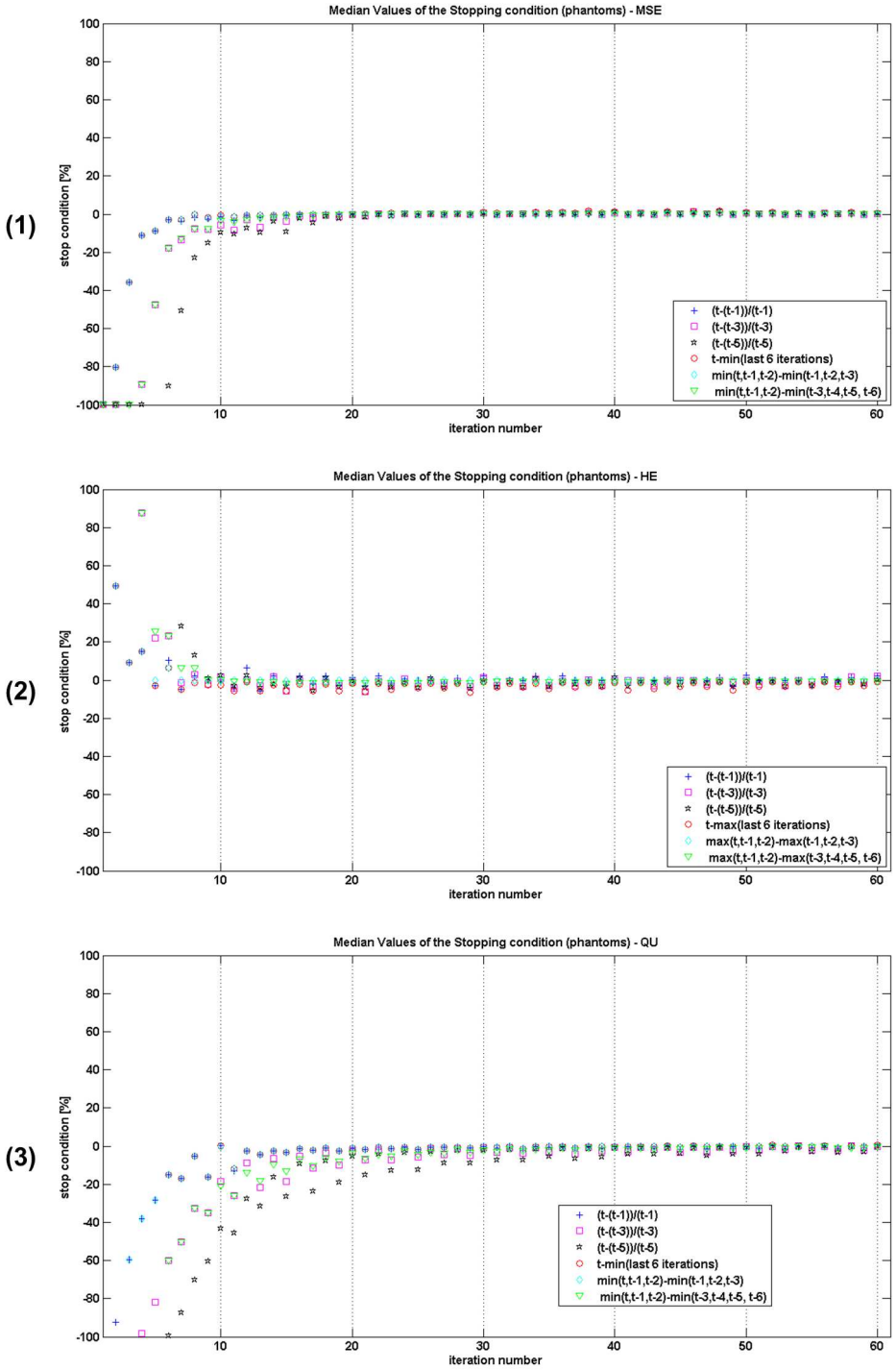
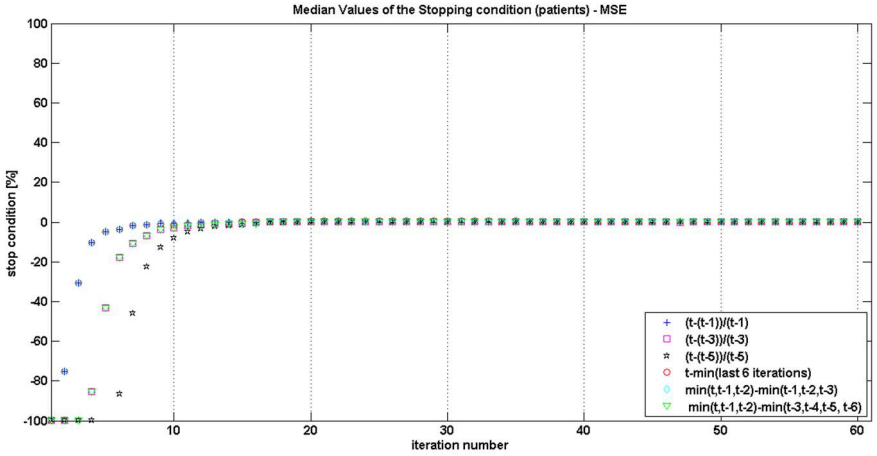
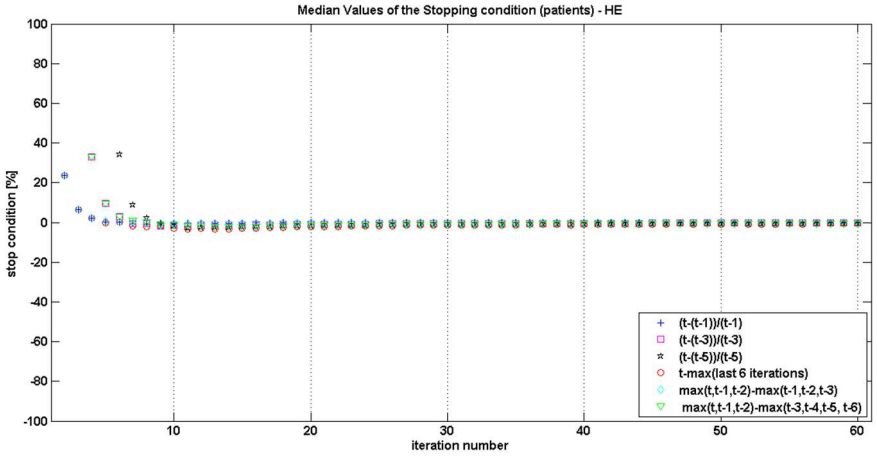


Figure 4.7: Median SCV values at each iteration for the phantom study over the five deformation for MSE, HE and QU. What should be noted here is the steepnees of the initial part of the curve. Better performances of a SCV are identified by fewer iterations used to reach convergence.

(1)



(2)



(3)

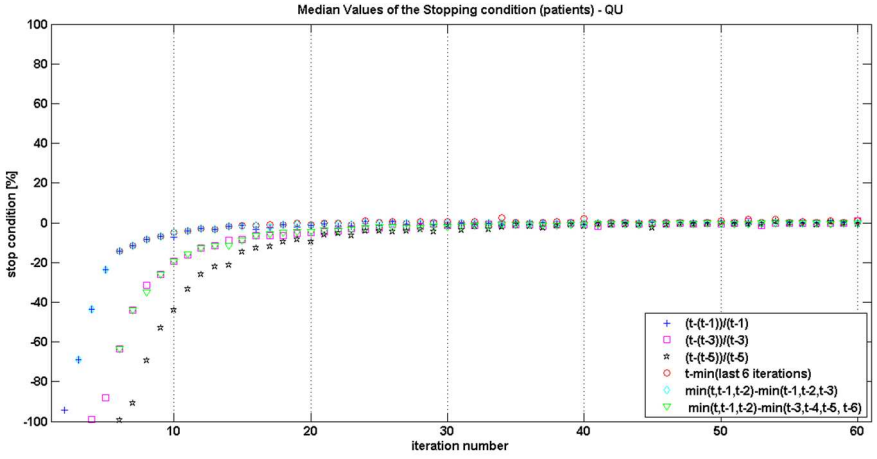


Figure 4.8: Median SCV values at each iteration for the patient study over the five deformation for MSE, HE and QU. With respect to phantom study, we notice that values are less smoothed out than in the previous case, likely because of the presence of artifacts. Better performances of a SCV are identified by fewer iterations used to reach convergence.

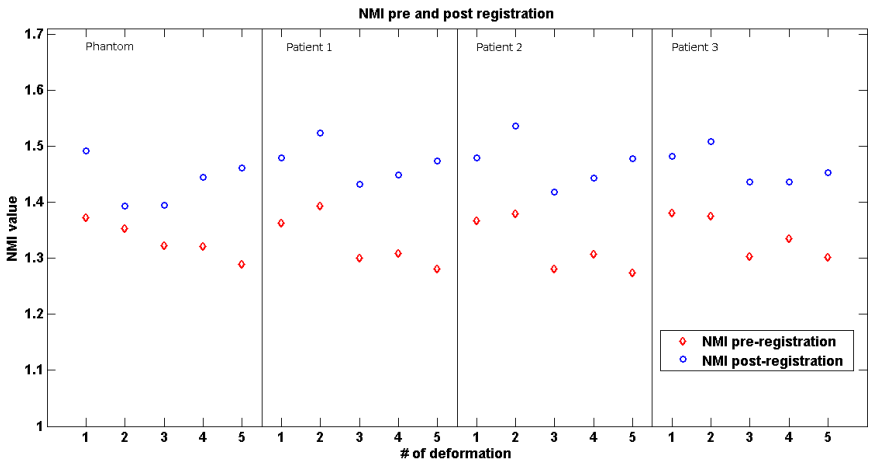


Figure 4.9: NMI values pre- (red) and post- (blue) DIR registration. Note that for all the registrations, NMI increased after DIR alignment.

4.2.3.2 *Deformation recovery capability*

We report the results about NMI pre and post deformable registration in Figure 4.9. As expected, the NMI increased, thus testifying that the two datasets have been correctly aligned. Figure 4.10 and 4.11 shows the plot of the median, 25th and 75th percentile for the residual of the displacement field of the five phantom and three patient cases respectively. It is evident that more iterations do not necessarily imply better convergence, which reinforces the need for a stopping condition. The chosen threshold should stop the algorithm near the minimum residual. Figure 4.10 and 4.11 also show how the slope of the initial part of the recovery changes slightly between phantom and patient cases. To illustrate this better, we show a direct comparison between phantom and patients deformed by the fifth displacement field in Figure 4.12: residual error reaches 2 mm in 35 iterations and 15 iterations respectively.

For the phantom study, we report the results on the metric performances. Table 4.3 illustrates the $SCV_{critical}$ at $t_{required}$ (1.875 mm). The chosen $SCV_{critical}$ is shown in bold. We then compute the number of extra iterations required to reach the $SCV_{critical}$ at $t_{required}$ (Table 4.4). For example, considering MSE, we look at the $SCV_{critical}$ for last two phantoms in the other simulations. The best performance algorithms are HE and MSE, while QU tends to need more iterations at a given threshold value. We than repeated the same procedure for the three patients. Out of the fifteen tested registrations, two did not reach the desired level of accuracy (i.e. median residual than 1.875 mm). In addition, out of the remaining registrations, in five cases the Jacobian was not evaluable. The final results are presented in Table 4.5. Again the HE outperformed MSE in terms of capability to detect the stopping iteration to be used, while QU demonstrated to be not sufficiently robust to changed image quality. In fact, it was greatly outperformed by MSE and HE in patient 2, which presented the most severe image artifacts.

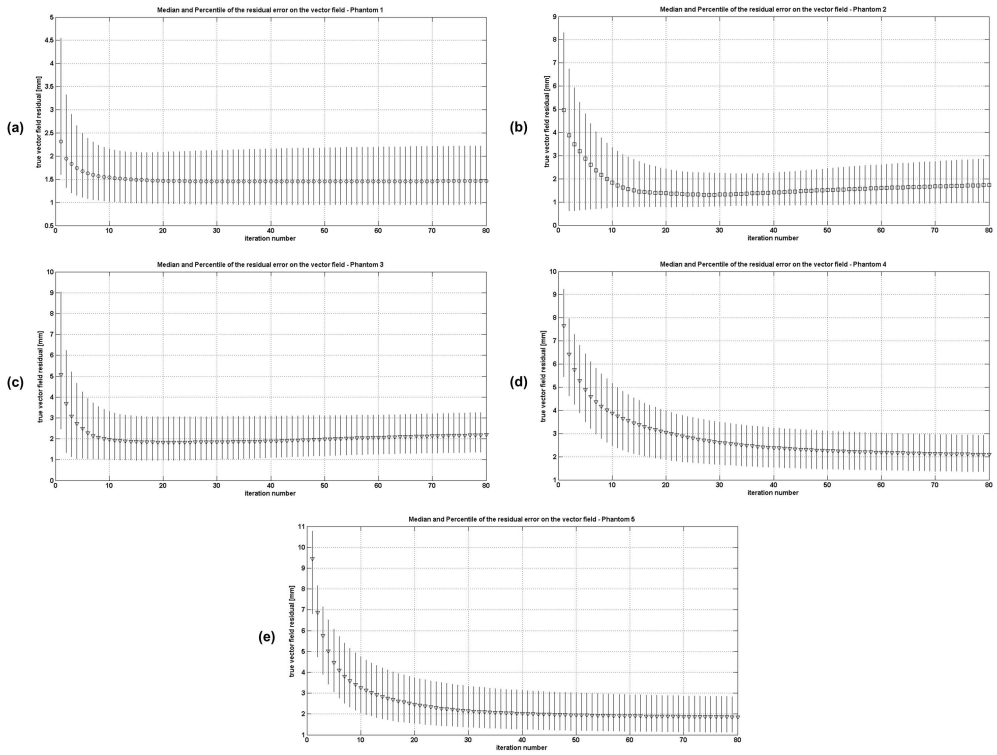


Figure 4.10: Residual errors on the vector field for phantom study. Graph shows median, 25^{th} and 75^{th} percentile for each analyzed deformations (from a to e). From the images, it is evident that more iterations do not always imply better convergency level, evidenciating the importance of a robust stopping condition stopping DIR at the minimum residual.

Table 4.3: $SCV_{critical}$ at the iteration corresponding to the chosen threshold (1.875 mm)

Test number	$t_{required}$	MSE	HE	QU	Jac
1	3	-0.373	0.027	-0.243	$6.53e^{-04}$
2	10	-0.033	0	-0.022	0.002
3	14	-0.019	0	-0.014	$3.64e^{-04}$
4	295	0	-0.002	$-2.38e^{-05}$	0.004
5	71	0	-0.003	-0.001	0.002

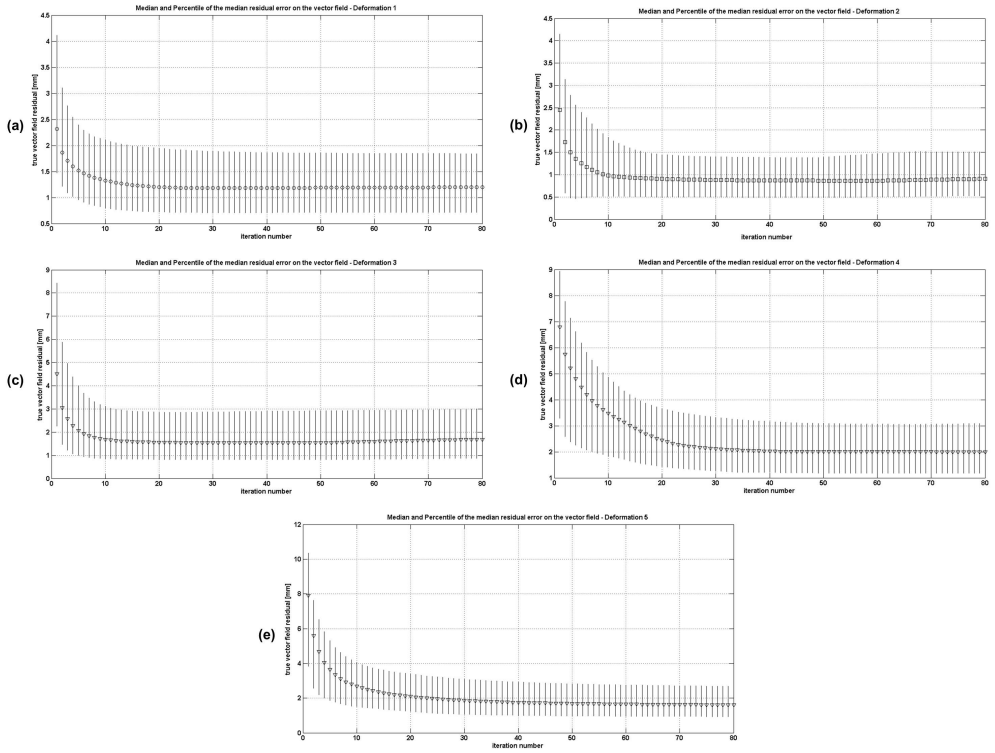


Figure 4.11: Median residual errors on the vector field for patient study. Graph shows median, 25^{th} and 75^{th} percentile for each analyzed deformations (from a to e). Besides the definition of an appropriate threshold, it should be noted that residual trend is equivalent to the patient's one.

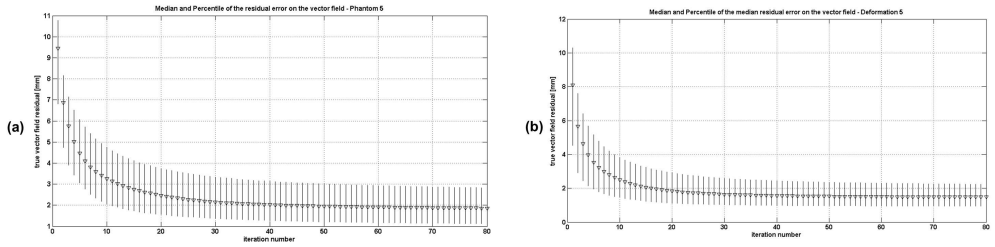


Figure 4.12: Mean residual errors on the vector field for selected patient study after 15 DIR iterations. Speed of convergence is different between the two graphs: in fact the level of 2 mm is found only after 35 iteration for patients, while that's reached after 15 iterations in phantom study.

Table 4.4: Number of extra iteration the algorithm needed to reach the minimum of the $SCV_{critical}$. *In this case we were not able to find the desired $SCV_{critical}$ in three cases out of five

Test number	MSE	HE	QU	Jac
1	19	2	197	4
2	11	0	99	—
3	8	0	216	0
4	0	2	0	—
5	0	1	96	—
Total	38	5	608	4*

Table 4.5: Number of extra iteration the algorithm needed to reach the minimum of the $SCV_{critical}$. The Jacobian was not evaluable in two cases for patient 1 (£), two for patient 2 ($) and one for patient 3 (&).

Patient number	# of registration converged	MSE	HE	QU	Jac
1	5	171	135	288	$10^{£}$
2	4	25	9	93	$0^{\$}$
3	4	65	30	68	$7^{\&}$
Total	13	261	174	449	17

4.2.4 *Discussion and conclusion*

We implemented and compared four different stopping conditions strategies for a deformable registration algorithm and tested them on LDDD. Our aim is to provide non expert users with an easy and immediate tool that would require the setting of an allowed degree of error for the algorithm and at the same time guarantee the optimal convergence of the registration. So far, most algorithms employ maximum number of iterations as escape condition, which, not only is non-intuitive, but also is not robust with respect to image quality and patient specific issues. Our testing dataset is composed by five synthetic deformations obtained by simulating the entity of displacement at each coordinate with a Gaussian distribution, centered at a different voxel of the image grid (see Table 4.2). One anthropomorphic dosimetric phantom and three patient scans were warped with the generated synthetic deformations.

The problem of finding a robust stopping condition is first of all influenced by the way in which the condition itself is calculated. Aiming at requiring just a percentage error from the user and having no target value in mind, it is hard to define with which value the current stopping condition value should be compared to. We tested six different strategies (Table 4.1), in terms of speed of convergency and stability. For our dataset, the winning strategy is the one that combines the minimum of the last three iterations including the current one and the minimum of the three iterations before last. The minimum shall be substituted with the maximum for the case of HE, which by definition is increasing during registration. The worst performing strategies in terms of convergence speed are those relying on the subtraction between the current and some previous value, which suffer from a long initialization time. The longer initialization time would definitely have greater impact in the finer stages of a multi-resolution registration strategy, thus leading either to registration errors and/or to longer computational times. On the other end, although fast, just looking at the previous iteration might induce the algorithm to stop in a local minimum or into a plateau, thus subsequently leading to registration errors.

We studied MSE, HE, QU and Jacobian in terms of speed of convergence and of deformation recovery capability. Each of the strategy used captures different aspects of the DIR problem. While MSE monitors just the evolution of the difference between intensities, HE can indeed be related to the smoothness of the field, which is correlated both to the physical meaningfulness of the transform and to the degree of convergence. Together with HE, QU quantifies how much forces are pulling the transformation and is computed on the velocity field, which is actually driving the deformation in a demons based strategy. If the relative update in the pulling force goes to zero, the algorithm shall move to a finer registration stage or stop. By computing the Jacobian of the update field, we intended to detect any possible discontinuity leading to a non-physical result. The ratio between the number of elements below a positive ε and the total number of voxels just requires just to be compared with a user threshold, making this metric a favorable candidate as easy stopping criteria. However, analyzing the metric evolution, we notice its instability and irreproducibility between patients, as showed by the high number

of cases, in which either the convergence or the desired critical value were never reached.

Comparing phantom and patients results, we notice that the number of extra iterations needed increases for patients and in particular when the image quality was severely compromised by artifacts in the image. In addition, we had some cases in which the deformation was not recovered to the desired accuracy (residual < 1.875mm). This seems to be correlated with image quality. In both datasets, though, the winning metric seems to be HE, which outperformed both QU and MSE. QU is also sensitive to image artifacts, which might induce abrupt changes in the update field. These changes have a limited impact on HE, because of the Gaussian smoothing filter applied from the algorithm prior to the calculation of the iteration deformation field. MSE does not monitor the deformation field throughout the computation and takes advantage of the sharp differences due to artifacts.

The concept used here can be easily extended to any transformation type, but according to the employed model, it might need tuning and features extension. Furthermore, this initial work needs to be extended to real deformation cases, in which the ground truth deformation is either unknown or derived from a non-image based DIR strategy, such as for example thin plate spline registration.

4.3 VALIDATION OF A DIR ALGORITHM

4.3.1 *Classical validation indices*

Popular intensity based methods include voxel-by-voxel difference and correlation indices, which can be global or restricted to a portion of the image around a landmark [147, 146, 145]. Typical example is the Correlation Coefficient R^2, where R is defined as:

$$R = \frac{cov(A, B)}{\sigma_A \sigma_B}$$

with $cov(A, B)$ and σ_A, σ_B are the covariance and the standard deviation of the two images A and B respectively. The value of R^2 is such that $0 \leqslant R^2 \leqslant 1$ and gives the proportion of the variance of A that is predictable from the B, denoting the strength of their linear association. For example, if $R = 0.922$, then $R^2 = 0.850$, which means that 85% of the total variation in B can be explained by the linear relationship between A and B (as described by the regression equation). The other 15% of the total variation in B remains unexplained. A famous example from Anscombe [170] demonstrated the ambiguity of linear correlation and therefore the limit of the correlation coefficient. Figure 4.13 illustrate four different sets of points all having the same statistical properties, but very different graphical aspect. The case in which correlation coefficient returns meaningful results is the one illustrated on the top left graph.

To overcome this issue, often MI or NMI (see Table 3.1 for their definition) are employed as descriptors, looking in particular to the spread of the joint entropy graph, which is increasing with images disalignment (see Figure 3.3). It shall be noted that, in contrast to other indices, it is not possible to quantify a maximum and minum value for MI or NMI. Therefore, people prefer to compute relative

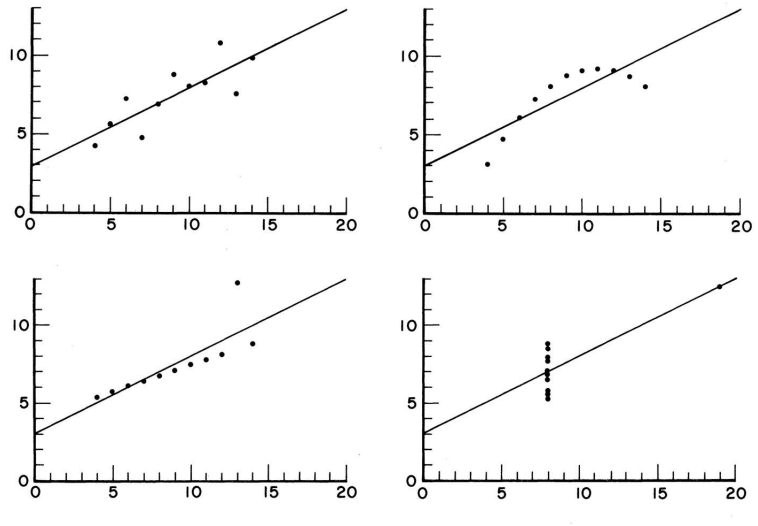


Figure 4.13: Anscombe's quartet. We plot four different datasets that have identical statistical properties, whose graphs appear very different. [170]

increase or decrease and/or they refer everything to the theoretical maximum, thus being the MI or NMI of an image with itself.

Volume based metrics require the accurate segmentation of structures of interest on all images to be compared and/or the propagation of previously outlined structures on all the scans involved in the analysis. Manual delineation is often preferred for this purposes, despite being affected from intra- and inter-rater variability [144, 171, 172, 173, 174, 175]. Amongst volume-based metrics, we count the Jaccard and Dice Similarity Coefficient [165], which express the mean overlap between two structures, besides the classical volume estimation. Perhaphs the most popular is the DSC, whose value varies between 0 and 1, i.e. from no overlap to complete overlap, but depends on the total structure volume, i.e. the largest the structure the nearest to 1 the value of DSC is, even though the overlap is the same.

Different authors have also come up with several solutions to compute Surface Distances (SD) [176, 177, 178, 179, 180, 181]. To evaluate the local surface distances, after triangulation of the contour points, one idea is to compute the Root Mean Square Error (RMSE) as

$$\mathrm{RMSE} = \frac{\|d\|}{N}$$

where $\|d\|$ is the norm of the vector of the shortest Euclidean distance between all the N correspondent closest points in the datasets without any further refinement. The correspondence between points is defined as in [116].

Other possible measures of SD include both Mean SD (MSD) and Hausdorff SD (HSD), thus both an average and a maximum measure of distance between the analyzed surfaces. Given that no a priori knowledge of the reference contour can be assumed, a bidirectional calculation of the shortest distance between an arbitrary voxel and the contours by means of a Distance Map is often performed. A distance map indicates, for each pixel in the objects (or the background) of a binary picture, the shortest distance to the nearest pixel in the background (or the

objects) [182]. In this way, the problem of finding correspondences between points is directly overcome. MSD and HSD are defined as follows:

$$\mathrm{MSD}(A,B) = \frac{1}{|\Psi(A)| + |\Psi(B)|}\left(\sum_{\psi_A \in \Psi(A)} d(\psi_A, \Psi(B)) + \sum_{\psi_B \in \Psi(B)} d(\psi_B, \Psi(A))\right)$$

$$\mathrm{HSD}(A,B) = \max\left\{\max_{\psi_A \in \Psi(A)} d(\psi_A, \Psi(B)), \max_{\psi_B \in \Psi(B)} d(\psi_B, \Psi(A))\right\}$$

where $\Psi(A)$ and $\Psi(B)$ indicate the subset of ψ voxel in a structure belonging to the contour and $\sum_{\psi_A \in \Psi(A)} d(\psi_A, \Psi(B))$ and $\sum_{\psi_B \in \Psi(B)} d(\psi_B, \Psi(A))$ are the sum of the shortest distances of the voxels $\psi_A \in \Psi(A)$ from the contour $\Psi(B)$ and vice-versa.

In addition, from clinician perspective, volume differences and distance between Center Of Masses (COM) are still the most immediate and clear. In all cases, the desired level of accuracy (i.e. metric value) is determined by image resolution. For example, if the voxel of one image is 1x1x3 mm, the acceptable RMSE value is given by the $\sqrt{1^2 + 1^2 + 3^2} = \sqrt{11} \simeq 3.32$ mm (i.e. it is the diagonal of the voxel). For simplicity, often this upper boundary for accuracy is approximated with the largest value of the three voxel dimensions, i.e. in the previous example 3 mm.

4.3.2 *Landmark based validation*

Another approach is to compare the position of anatomic or external landmarks after registration has been performed. The manual individuation of landmarks is affected by inter- and intra-observer variability. Therefore, in order to purge the variability, the centroid of a series of repeated localization from different raters could be assumed as stable landmark. This is obviously not possible in a clinical application, both because of the time consuming procedure and of the need of a sufficiently numerous group of independent observers. The availability of a robust and automatic feature extraction method would introduce an independent verification tool both during validation of new algorithms and in clinical use, not relying on visual inspection but rather directly derived from image features, and the deployment of point-based registration methods such as the ones described in §3.2.3.1 and in [123]. The landmarks located on medical images can be classified as anatomical, if they describe relevant anatomical features (e.g. pulmonary bifurcations), or as geometrical, such as angles or extrema. They can be further divided into extrinsic, if they located on external solid objects located in the image volume, and intrinsic, thus based only on image information [115].

According to Forstner et al [183, 184], requirements for extracted features include:

- Distinctiveness, i.e. distinguishable from their neighbors.
- Invariance, i.e. not influenced by geometrical distorsions as these are directly influencing accuracy and repeatability of the extraction.
- Stability, i.e. noise robustness.

- Rarity, i.e. guaranteeing global separability, whereas distinctiveness ensures the local one.
- Interpretability, i.e. extracted points should be meaningful for the operators.

Commercially available features extraction methods individuate points by means of cross-correlation between a template and the image volume. An example is provided by Castillo et al [185] for lung bifurcation extraction. The limit of this kind of approaches is definitely the need of an apriori knowledge on the shape of the neighborhood of the feature, as well as the rotation, scale, illumination and 3D position invariance of the object.

One of the first differential operators for automatic features extraction was developed by Kitchen and Rosenfeld [148] and is based on the calculation of first and second order derivatives of the 2D image $g(x,y)$. Corners are identified as local extrema of the operator

$$KR(x,y) = \frac{(\nabla g)^T H_g \nabla g}{|\nabla g|^2}$$

with $\nabla g = \binom{g_x}{g_y}$ and $H_g = \begin{pmatrix} g_{xx} & g_{xy} \\ g_{yx} & g_{yy} \end{pmatrix}$ are the gradient (i.e. first derivatives matrix) and Hessian (i.e. second order derivatives matrix) of $g(x,y)$. The operator is calculated at pixel level and only those lower than a certain threshold are considered salient.

Forstner et al [183, 184], instead, proposed to detect features maximizing the following equation:

$$F(x,y) = \frac{detC_g}{trC_g}$$

where C_g is the covariance matrix in an elliptical neighborhood of the voxel. Similarly Rohr et al [186] defined the features thresholding the determinant of the covariance matrix. All of these operators were subsequently extended to 3D volumes and used for medical imaging applications [187].

The Harris corner detector [188, 150] is defined from the auto-correlation matrix centered on the point. If the image intensities are approximatevely constrant in a region, the matrix eigenvalues are small, whereas corners are represented as large eigenvalues (i.e. the auto-correlation function has significant variations in all directions). An edge is individuated when the autocorrelation has a spike and eigenvalues are of different order of magnitude. The main limit of this algorithm is the scale invariance, as the method works on a single scale. Crowly and Praker [189] developed a graph representation to locate peaks at different scales, so that the features can be matched also at different scales.

A method that can extract and match stable and trustworthy points at different scales between two images, is Scale Invariant Features Transform (SIFT). D. Lowe [151, 152] initially designed it for 2D images, and Cheung & Hamarneh extended it to nD images [153]. The aforementioned method has already been used for DR and contour propagation based on Thin Plate Spline [190, 191, 192].

We present an application of SIFT to volumetric medical images with anisotropic voxels to validate DR. The invariance properties of a feature extraction method are

especially important, if we want to find salient points in images acquired with different modalities, or scans of different patients, or, as in our case, temporal series of the same patient. For this purpose, we test SIFT invariance properties on a phantom, by applying different rigid transformations and, for the first time, non-rigid transformations. Finally, we apply SIFT to pairs of registered images of twenty head and neck patients, with the aim of providing a new index for assessing DR performance.

4.3.2.1 *Scale Invariant Feature Transform Feature Detector*

SCALE-SPACE EXTREMA DETECTION To prevent aliasing due to noise, we firstly pre-smoothed the input image $I(x, y, z)$ with a Gaussian kernel with $\sigma = 0.5$. Then the image is represented as a family of smoothed images, i.e. a scale space, parametrized by the smoothing kernel dimension. In practice this results in convolving $I(x, y, z)$ with a Gaussian kernel $G(x, y, z, \sigma)$:

$$L(x, y, z, k\sigma) = G(x, y, z, \sigma) * I(x, y, z) = \frac{1}{\left(\sqrt{2\pi\sigma}\right)^3} e^{\frac{\left(x^2+y^2+z^2\right)}{2\sigma^2}} * I(x, y, z)$$

The output is a stack (or octave) of blurred Gaussian Images $L(x, y, z, k\sigma)$ separated by a constant multiplicative factor $k = 2^{\frac{1}{s}}$, with $s = 3$ number of intervals. The value of the Gaussian kernel sigma is updated moving from a scale to another, exploiting the fact that sequential filtering of independent variables corresponds to a single filter with variance equal to the sum of variances.

Once a complete octave has been processed, the Gaussian image that has twice the initial value of σ is subsampled by a factor of 2, serving as source image for the next octave. Finally, Difference of Gaussian Images (DoGs) are obtained by subtracting each Gaussian Image from the adjacent image in each octave, thus

$$DoG\left(x, y, z, k^i\sigma\right) = \left[G\left(x, y, z, k^{i+1}\sigma\right) - G\left(x, y, z, k^i\sigma\right)\right] * I(x, y, z)$$

and therefore

$$DoG\left(x, y, z, k^i\sigma\right) = L\left(x, y, z, k^{i+1}\sigma\right) - L\left(x, y, z, k^i\sigma\right)$$

The choice of DoG images is motivated both from the possibility of calculating them efficiently and from the fact that they are an approimation of the Laplacian of Gaussians $\sigma^2\nabla^2 G$, which allow the detection of more stable features than for example gradient or Harris corner detector [193, 194]. A scheme for this strategy is shown in Figure 4.14.

The candidate keypoints are detected as local extrema by comparing a voxel to its 80 neighbors in 3x3x3 regions at the current DoG and adjacent DoGs. This comparison implies that $s + 2$ DoGs are required, i.e. $s + 3$ Gaussian Images. In this work, we generated 5 DoGs per 3 octaves.

KEYPOINT LOCALIZATION We defined a point as stable, if it is well contrasted and if it is not located on plate-like or tubular structures. Therefore, a simple thresholding on contrast is implemented, such that, assuming the image values to be in the range [0, 1], each candidate landmark $\hat{x}$ have $|DoG(\hat{x})| > t_{con}$.

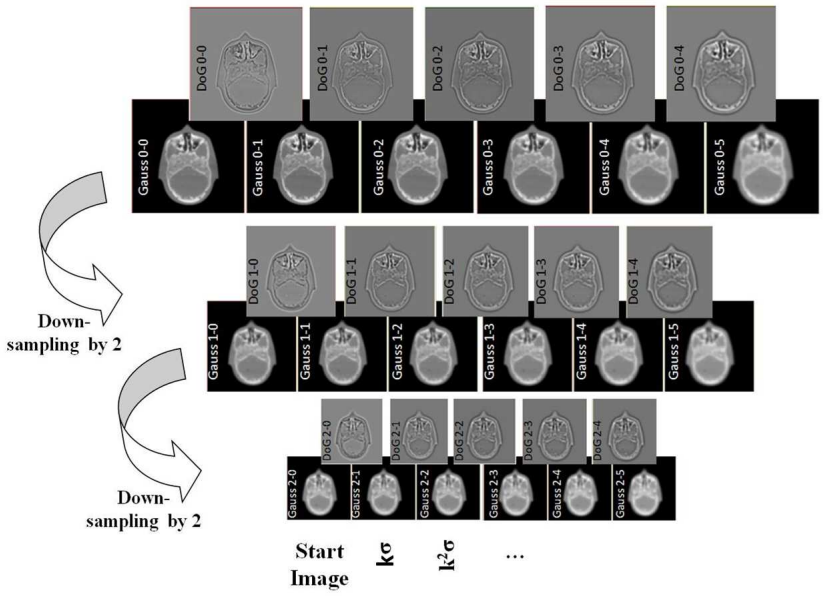


Figure 4.14: Schematic illustration of the SIFT algorithm. From top to bottom, generated gaussian images (Gauss-*m*-*n*) and corresponding Difference of Gaussian (DoG-*m*-*n*), being used by SIFT for features detection. In the notation, m indicates the octave and n the number of the considered Gaussian Image, thus between Gauss-*o*-*n* and Gauss-*1*-*n* there exist a resampling factor of 2, while Gauss-*o*-*o* and Gauss-*o*-*1* are filtered with σ which differ by a factor $k = 2^{\frac{1}{3}}$ in our case.

Table 4.6: Tested rigid transform for validation

Transform	Parameters
Translation (x,y,z)	[0,1,2,3,4,5,6] mm
Translation (x,y)	[0,1,2,3,4,5,6] mm
Rotation (z)	[0,1,2,3,4,5]°
Scaling (x,y,z)	[0.8,0.9,1.1,1.2]

Among well contrasted features, we selected those in which principal curvatures have the same sign and same order of magnitude. The condition on the sign is guaranteed if $\mathrm{tr}(H)\det(H) > 0$ and $\sum \det_2^P(H) > 0$, whereas $\frac{\mathrm{tr}(H)^3}{\det(H)} < \frac{(2t_{curv}+1)^3}{(t_{curv})^2}$ controls the order of magnitude. In the equations, H is the Hessian of the image, $\mathrm{tr}(H)$ its trace, $\det(H)$ its determinant and $\sum \det_2^P(H)$ the sum of the second order minors of H. t_{curv} instead is a threshold quantifying the maximum ratio between the biggest and the smallest eigenvalue of H.

ORIENTATION ASSIGNMENT AND MATCHING After detecting candidate features, we computed gradient magnitude and orientations of the first Gaussian Images L(x,y,z,σ) in each octave, thus describing each voxel of the gradient image with three spherical coordinates. The descriptor is defined as a vector containing the values of all orientation of a region around the keypoint weighted by a Gaussian filter. We calculated the Euclidean distance between the descriptor of feature point n of one image and that of a potential association n' of another image as

$$S_{n,n'} = \sqrt{\sum_{\alpha=1}^{k} |(\nabla I_n)_\alpha - (\nabla I_{n'})_\alpha|^2}$$

where $(\nabla I_n)_\alpha$ and $(\nabla I_{n'})_\alpha$ are orientation histograms of the two input images and α is the bin index.

After $S_{n,n'}$ is calculated for all points n, the two closest points n'_1 and n'_2 are identified. If the ratio $\frac{S_{n,n'_1}}{S_{n,n'_2}}$ is less than a previously established threshold t_{ass}, the point having the lowest $S_{n,n'}$ value is chosen as the correspondence of landmark n. Otherwise, no association is made for the point. To further increase the accuracy of feature point association, the bidirectionality of each landmark association is verified.

4.3.2.2 *Validation: results and discussion*

We tested our approaches on an image of RANDO® phantom acquired on a clinical CT scanner, using supine setup and clinical acquisition protocols. The volume acquired is 512*512*123 voxels and [0.94, 0.94, 3] mm element spacing. We applied rigid and non-rigid transformations on RANDO® creating a synthetic image of the phantom , thus simulating the stages of a possible ART registration process, while testing SIFT invariance properties.

Table 4.7: Median non-rigid transform for SIFT validation

	Right-Left [mm]	Anterior-Posterior [mm]	Superior-Inferior [mm]
vf_1	0.06	0.51	0.01
vf_2	0.06	0.51	1.51
vf_3	0.62	0.52	1.41

METHOD To study algorithm properties, we simulated and applied to RANDO® rigid and non-rigid transforms, which we sum up in Table 4.6 and 4.7 respectively. Non-rigid deformations are generated as in chapter 4.2. From application of these deformation fields, we obtained RANDO® warped. In both cases, we applied the transformation to the points individuated on the original RANDO® scan, and compared these with those obtained from SIFT in the synthetic image. Because of the definition of the warping method and consequently of the deformation field (see §3.2.5), in the non-rigid case, the warping same way as in the rigid case. The comparison was performed by computing the accuracy in terms of points distance along the three dimensions.

RESULTS AND DISCUSSION Before proceeding with proper validation, we experimentally tuned the parameters σ, t_{con}, t_{curv} and t_{ass}. We determined $\sigma = 1.5$, $t_{con} = 0.03$ and $t_{curv} = 20$, confirming Allaire et al [192] with respect to contrast, but being more permissive on the curvature, thus allowing the identification of points on the skull. Lowering t_{curv}, results in purging more points on the edges, while t_{ass} is a measure of the degree of separation between the points. In fact, the stronger the association between the features, the lower $\frac{S_{n,n_1'}}{S_{n,n_2'}}$ is, because the second-closest neighbor has greater $S_{n,n_2'}$, i.e. n and n_2' are further apart. For example, bones are well defined structures and therefore t_{ass} can be lowered even to 50%, thus avoiding false matches. For other structures, like soft tissues, whose features are not clearly defined on the image, a higher t_{ass}allows to increase the number of identified matches to an acceptable value. t_{ass} needs to be higher also for the case of low quality images, where feature disambiguation is compromised by noise. Here we used t_{ass} of 80%, but ideally one should proceed with a organ specific threshold [191].

To compute the descriptor, a 16x16x16 region is defined around the keypoint location and divided into 4x4x4 sub-regions. A circular Gaussian with σ equal to one half the width of the region is used to assign a weight to the magnitude of the voxels in the region. This avoids sudden changes in the descriptor and gives less emphasis to gradients that are far from the keypoint. For each of the 4x4x4 sub-region, the orientation histogram is built, describing 8x8 orientations of the gradient.

In Figure 4.15, we show three examples of successful individuation of corresponding features in the original and warped image (respectively left and right in Panel). In Panel (a), note also that warped points for translation are found at exactly two slices from the original ones (i.e. exactly 6 mm), thus reinforcing the hypothesis of robustness of the algorithm.

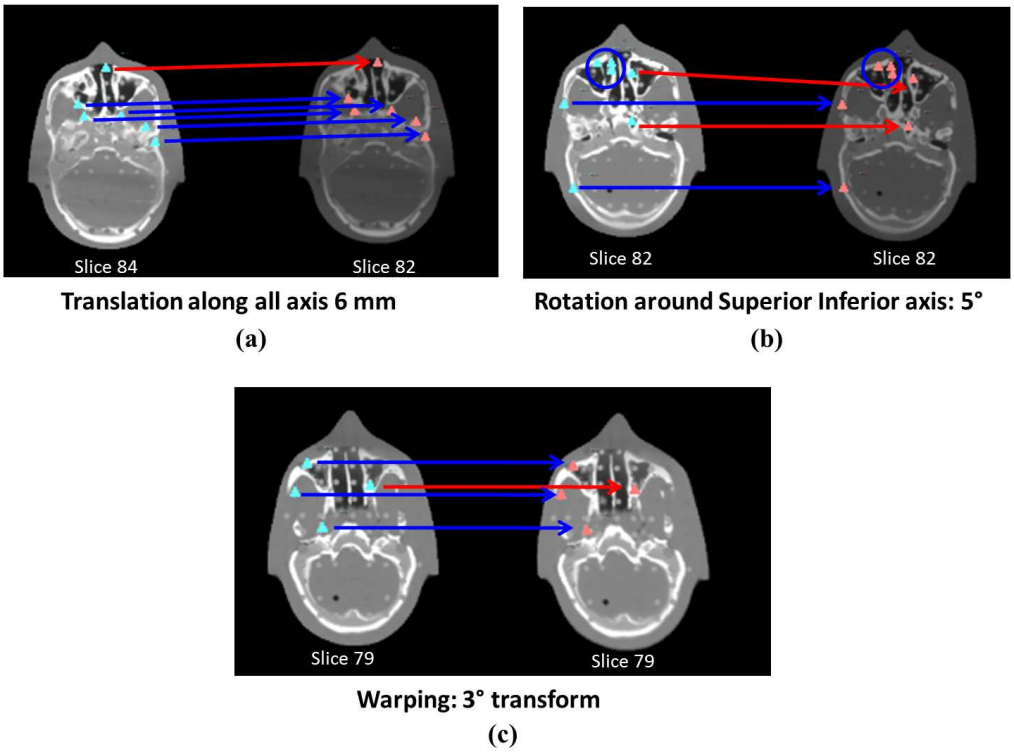


Figure 4.15: Three examples of successful corresponding features detection in the original (right) and synthetically warped (left) image. Reds codes for a feature detected as minimum of the DoG, while blue for a maximum.

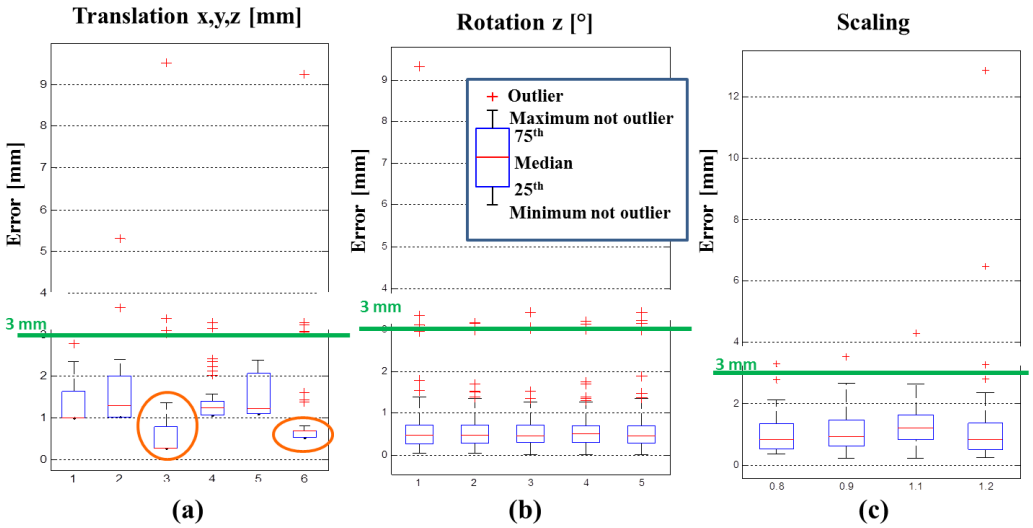


Figure 4.16: Median and percentile landmark errors for rigid transformations. In Panel (a), we report errors associated with translations, while Panel (b) and (c) show the results for rotations and scaling. We note that in all case, the error stay below 3 mm, with a negligible number of outliers. Note that for translation multiple of the prevalent image resolution (i.e. SI slice thickness, 3 mm), the error is largerly reduced. This is due to the fact that the resampling after non-multiple translations equates in voxel information integration between different slices.

Figure 4.16, 4.17 and4.18 shows boxplots of the points accuracy distribution at different transformations. We notice that, for rotation (Figure 4.16, Panel b), the median for all five cases is below the voxel dimension, confirming the invariance of the method to this transformation. There are few outliers around 3mm, due to a wrong association of the keypoint in axial direction. Scaling (Figure 4.16, Panel c) and rototranslation (Figure 4.17) demonstrated the same invariance properties. Also for growing deformations (Figure 4.18), we notice that the median accuracy is below the voxel dimension, although the variability it is bigger than for rigid transforms. Higher outliers are due to wrong matching between two features, which is likely if the images have symmetrical features (such as for example the mastoids in the head and neck).

For translations along the three dimensions (Figure 4.16, Panel a), we see a worsening of accuracy for values that are not multiple of the axial voxel dimension (i.e. 3mm). If the image content is translated along anterior-posterior and left-right directions, the accuracy stays below 1mm, with a little increase in accuracy value depending on having tested not exact multiples of the voxel dimension in these directions (i.e. 0.94mm). As shown in Figure 4.19, also the number of matches decreases in case of translations non multiple of the image resolution, while it is constant for any degree of rotation and scaling, besides for case of no scaling (i.e. scaling factor equal to one). For scaling, the reduced number of features is due to the differences in the DoGs at full scale or after transformation. We further analyzed the number ouf outliers against the number of corresponding stable features, which was neglectable in comparison to the total number of points. Results are reported in Table 4.8, in the form of total number of outliers for a given transform

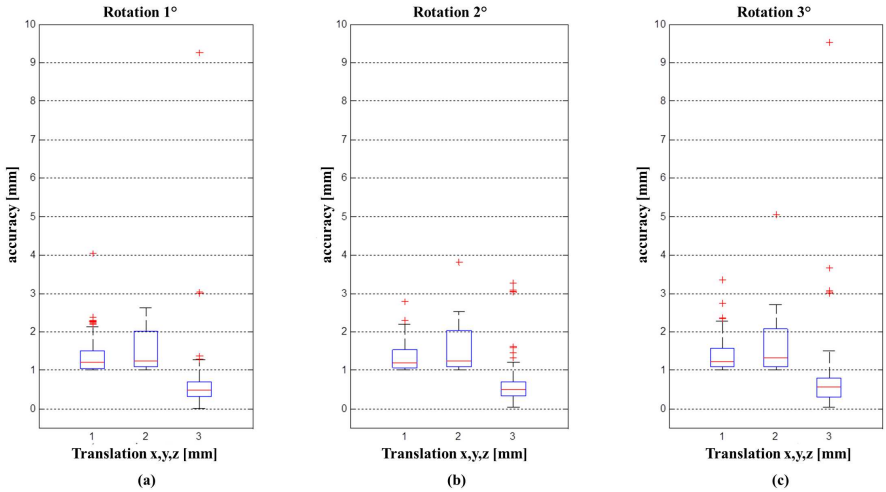


Figure 4.17: Residual errors after rototranslations. From Panel (a) to (c), we study SIFT performances at varying SI-rotation degrees and translations. We note that translation component is the most influencing one in all cases.

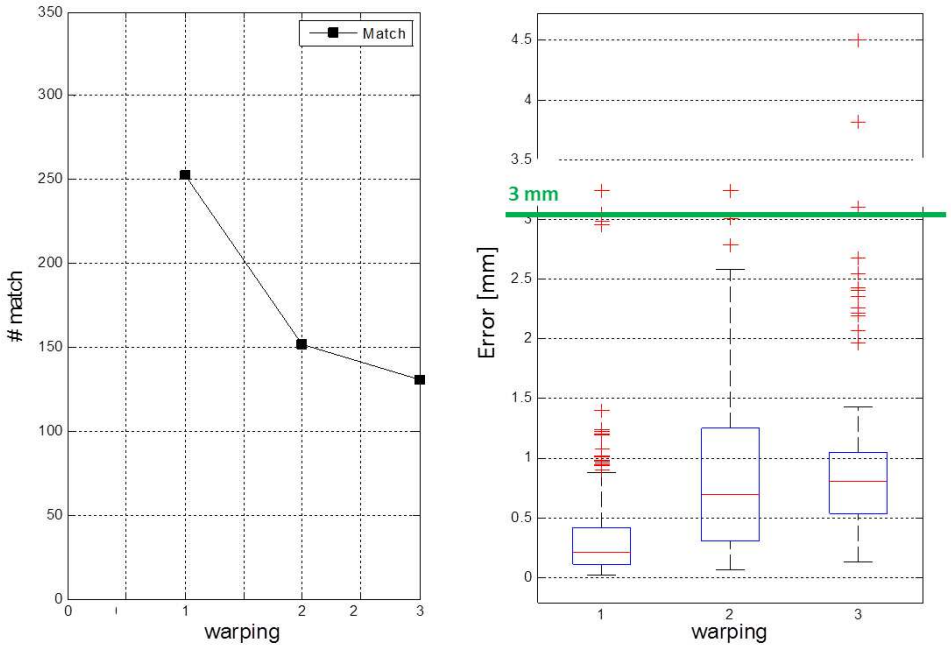


Figure 4.18: SIFT invariance to non-rigid transforms. Although few more outliers were detected, the most of them stay below the accuracy level of 3mm and therefore, the method can be considered invariant to deformation type.

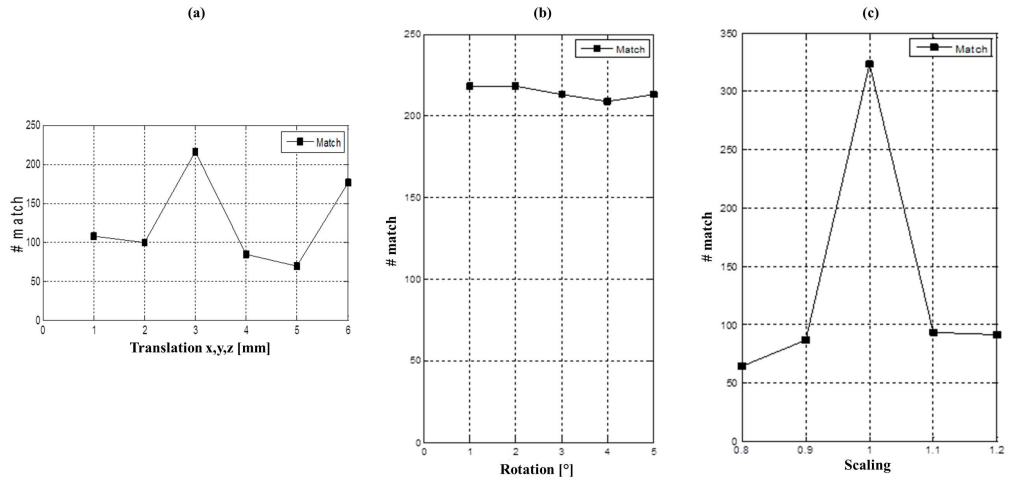


Figure 4.19: We report the number of matching features individuated on the two images (reference and floating). We note that while rotation is constant, the presence of a non-multiple translation and of scaling results in a sensible reduction in the number of mached features.

versus total number of keypoints. We note that SIFT is rather sensitive to scaling as testified by the reduced number of features identified. We clustered the rototranslation cases on the basis of the translation being multiple or non-multiple of the voxel greater dimension.

Looking at the results we can distinguish two main sources of error at keypoint detection:

1. in-voxel association: the finite difference derivative approach used in this algorithm implies that features are always associated with voxel center and as such an error of up to half of the diagonal of the voxel might be introduced (in our case around 1.64 mm)

2. between-voxel association.

The second type of error requires a more precise explanation and an example. Let us consider two different translations applied to an image whose pixelation is represented in blue in Figure 4.20. After the application of the transform, we obtain the pixelation in red. Let us further suppose to have detected keypoint K_1 in voxel i, that would be associated to the center of voxel C_1, as explained at point 1. As translation is a rigid transform, the relative position of K_1 and C_1 would be the same in the warped image, thus K_2 and C_2 are corresponding to K_1 and C_1. Given the entity of the translation considered in Panel A, C_2 is still in voxel i and therefore K_2 would always be associated to the same voxel. The error would still be the difference between the real keypoint location and the center of the voxel and be due to the resampling of the transformed image in the initial pixel grid. Let us now consider the transformation in Panel B. In this case, K_2 still falls in voxel i, but, as C_2 would be associated to $i+1$ because of resampling, the keypoint is instead associated with C_3. This intrinsic limit of the association raises the error to up to the maximum between the voxels dimensions (i.e. in our case 3 mm) plus

Table 4.8: Total and median number of outliers for all analyzed cases versus total and median number of keypoints

	# cases	#outliers / #keypoints	Median #outliers / median # keypoints
Translation	6	22 / 752	3 / 103
Rotations	5	31 / 1071	6 / 213
Rototranslations (non multiple translations)	6	16 / 506	2 / 84
Rototranslations (multiple translations)	3	20 / 622	6 / 205
Scaling	4	8 / 335	2 / 89
non-rigid deformations	3	31 / 536	13 / 152

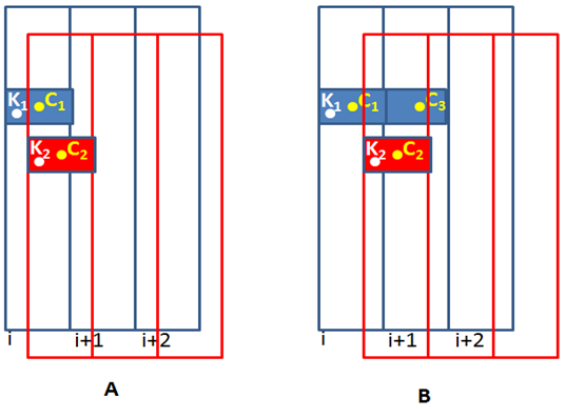


Figure 4.20: Example of inaccuracy introduced by SIFT algorithm. Let us consider Panel (a), where in blue we represent the original image, while in red there is the sinthetically translated volume. Detected feaure K_1 is identified with the voxel center C_1 and becomes K_2 in the translated image. This time it will be associated with the wrong slice because of the resampling after warping. In Panel (b) instead what should happen is shown. In this case K_1 is correctly associated with K_3, as this falls in the next slice.

the distance between C_2 and K_2. Obviously, this issue is solved if the transform is a multiple of the voxel dimension, where no integration of information occurs.

4.3.2.3 *Conclusions*

We studied an algorithm for automatic feature extraction and keypoint matching. The algorithm relies on progressive Gaussian filtering and calculation of image derivatives for the individuation of stable landmarks on medical images.

We classify the performances and limits illustrated in §4.3.2.2 according to the scheme of Förstner et al [184, 183]. The threshold on contrast and curvature allows to purge not only the less salient features (i.e. less *stable*), but also to disambiguate keypoints located on the edges which would violate the first requirement of *distinctiveness*. The main limit here are the experimental thresholds, which, not only are defined on a normalized version of the volume, but also are rather dataset and image quality dependent. In fact, the normalization favours the individuation of bony keypoints with respect to soft tissue, which correspont to a smaller sets of normalized units towards zero.

In addition, SIFT establishes matching points just looking at the descriptor, which is based on the histogram of a portion of image around the located features. Although this is an advantage in certain conditions, it could introduce false positive matches especially if the keypoints are defined on symmetrical objects, as for example the mastoid in head and neck patients. This limit could be simply addressed by introducing a regional search at the matching stage and/or using an Iterative-Closest-Point like approach, in a similar fashion as what was recently done by Lemuz-López and Arias-Estrada [195] in 2D images. We studied the *invariance* features of SIFT applying a set of known transformation to a phantom image. From the study, we verified the invarance to rotation and scaling, and looked at the decrease in the individuated features. We demonstrate a median level of accuracy in mm lower than the image resolution, with few outliers due to either wrong voxel association or to matching errors. The wrong association can be ascribed mainly either to the concentration of the information in the center of voxel or to resampling issues. Finally, while distinctiveness guarantees local separability of the points, *rarity* ensures that features are globally separable, which is of most importance in medical images. In fact a feaure could be repeated in more than one slice, compromising the validity of the approach. This is prevented at the detection step when considering not only the current but also the neighboring slices and DoGs. The *interpretability* of the extracted feaures is summing up all the previously described requirements. In fact, rare and distinct points provide a easier scenario for the medical user, which would be compromised if the algorithm would not be invariant to transforms. Saliency requirement further guarantees that located points are the most representative in the image. Therefore, SIFT algorithm, once refined, would be an accurate and operator-independent metric for deformable registration validation and preformance assessment, in comparison to more classical indices.

4.3.3 *Guidelines for performance indices choice*

Despite its immediacy and very low computational cost, the usage of a pure volume difference estimation brings along the loss of information on spatial misalignment. Therefore it might be preferable to combine this index with one describing the position. The combination with the center of mass is so far the most widespread strategy, although it should be done carefully as this metric suffers more than others under the inter-rater variability.

The main advantage of DSC instead is that it accounts for global misalignment, although it shall not be forgotten its mean nature, which implies that, no matter which structure is considered as reference, its value will not change. DSC will therefore detect misalignment, but might not be sufficiently sensitive to regional variation and not provide an indication of the location of the discrepancy (e.g. slice number). In addition, also in our results, we saw its dependence from structure volume. A smaller structure will result in a lower DSC while a bigger volume in a higher, because the inclusion or exclusion of one voxel has a different outcome depending on the total number of voxels in the structure (compare for example DSC of parotids and of GTV).

The definition of the reference volume and the individuation of intersection are fundamental for mesh distance calculation, which has the desirable property of being sensitive to both global and regional misalignment. On the down side, the construction of mesh from manually drawn points might result into non-regular and/or highly smoothed surface. Both irregularity and smoothing, together with absence of correspondence between two meshes may severely compromise the distance calculus. The probability that any of the three occurs, increases inversely with the resolution of the images, especially when the cranio-caudal resolution lowers in comparison to the in-plane one, and with the structure volume. Even though most software allow to exclude smoothing, the interpolation of the highly resolute points between two consecutive slices might be severe and affect the final mesh. In addition, the computational cost increases considerably when considering bigger structures. It is also arguable whether mesh distance is a good metrics choice for verification of registration in radiation therapy, where the main concern is the conservation of volumes and position. In addition, whether the volume differences and DSC coefficient reflect the combination of rasterization and contouring errors, meshes would need to build a triangularized surface on top of a rasterized volume, which is already affected by inter- and intra-operator variability. Moreover the whole distance distribution of the distances should be considered when analyzing the results. Popular quantity as the Hausdorff distance (i.e. the maximum of the calculated distances) may provide just a limited view on the actual discrepancies between the surfaces, and/or not be able to correctly account for the presence of concavities as well as irregularities in the structures.

RMSE between triangulated points and modified SIFT for automatic landmark extraction are affected from interpolation and extraction or matching error respectively, are the most descriptive metrics and demostrate the ability of quantifying both global and localized errors. RMSE is anyhow affected by the intra- and inter-rater variability, depends on the definition of correspondent points and of contours on the image. SIFT shows larger improvements margins, especially with respect to

the sensitivity to transformation non multiple of the voxel resolution and to the definition of a descriptor including location information, but has the advantage of being operator-independent.

5 CLINICAL APPLICATIONS

Intensity modulated radiation therapy (IMRT) allows greater control over dose distribution, which leads to a decrease in radiation related toxicity. However, it requires precise and accurate delineation of the Organs At Risk (OARs) and target volumes on a Computed Tomography (CT). In the head and neck area, treatment planning is especially complex and can require hours of tedious manual contouring work (2~4h). The final segmentation suffers from greater inter- and intra-observer variability, thus motivating the investigation of automatic segmentation algorithm, with the double goal of increasing robustness and reducing computational costs. Having a powerful instrument like DIR at our disposal, the problem of generating a robust segmentation of structures of interest of a planning CT was first addressed (§5.1).

As we underlined previously (see chapter 2), it has been proved that target and OARs are subject to significant changes during the course of treatment and there is evidence that HN patients may benefit from replanning [60, 49, 59]. Specifically, patient weight loss and reduction of parotid volumes were observed in HN patients, along with shrinkage of tumor volume [48, 44, 38, 47, 58]. The effect on parotid glands is particularly important for patients with oro- and rhino-pharynx tumors, in which the medial shift of the parotid center of mass corresponds to a shift towards the high dose coverage region [58, 48]. Here we propose a clinically feasible implementation of IGART for HN patients, relying on DIR for generation of a *virtual*CT (§5.2).

5.1 MULTI-ATLAS BASED SEGMENTATION FOR HEAD AND NECK PLANNING

Automatic segmentation can be coarsely divided into deformable or model-based approaches, and atlas-based methods. The first ones mainly rely on rely on local image features, such as edges and make use of the assumption that object boundaries are distinct. In order to cope for failures in regions where this assumption does not hold (e.g. soft tissues), prior information about the structures aspect has been introduced, but needs to be balanced in order not to limit the accuracy.

In this work, a multi-atlas strategy is developed (Figure 5.1). Generally speaking, we define atlas a CT image with its correspondent manual segmentation. After rigid pre-alignment on a common reference, we extract the most similar subjects to the current test patient. With the help of DIR, the selected atlases are registered onto the CT to be segmented and the resulting transformation is used to warp the associated manual contours. Since the warped structures are all equally probable realizations of the anatomical structure of the new subject, we develop an efficient and robust strategy for label recombination. Finally we validated our approach by comparing the automatic segmentation to the manual one in terms of Dice coefficient.

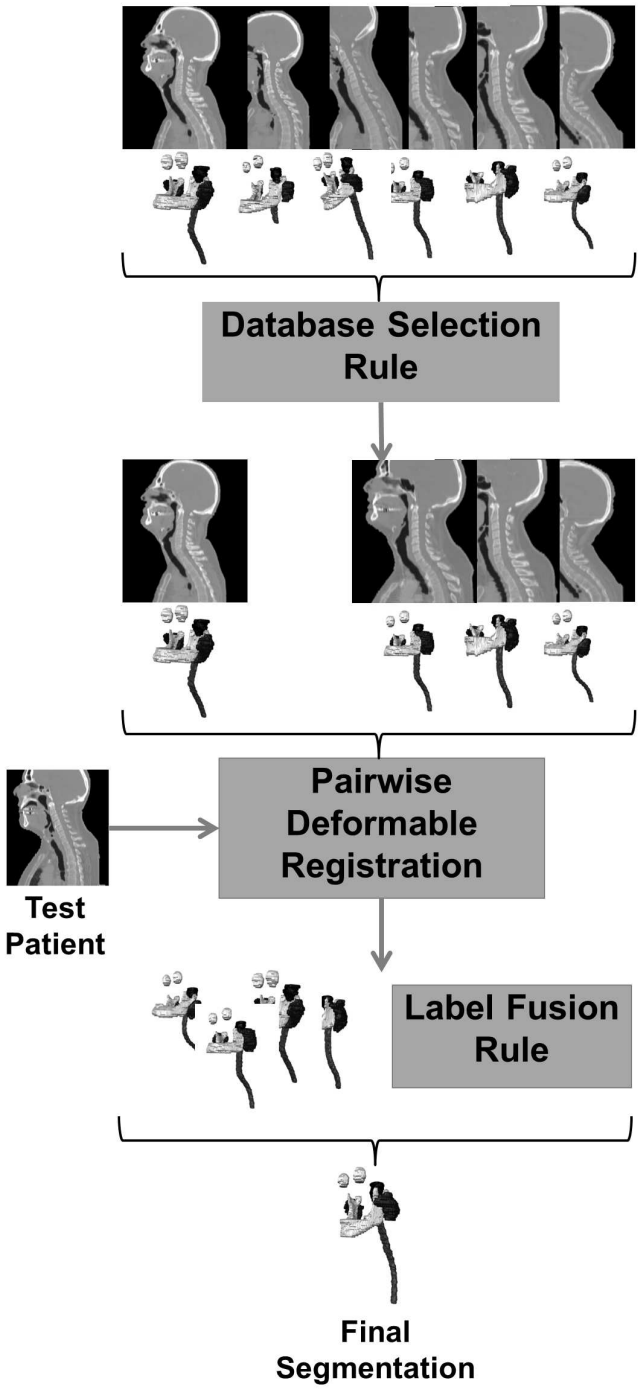


Figure 5.1: Algorithm workflow. After an initial alignment of all the atlases in a common reference frame, the chosen selection strategy extracts a subset of suitable atlases, which are registered onto the test patient awaiting segmentations. The last part is dedicated to generate the final segmentation from a series of equally probable propagated labels, by means of a chosen fusion rule.

Table 5.1: Parameters of the selection strategies studied in this work

Strategy	Tested Values
TH	NMI > [0.9:0.01:0.99]
FN	0, [50:5:100] %

5.1.1 *Materials and methods*

5.1.1.1 *Algorithm outline*

Figure 5.1 illustrates the proposed algorithm workflow. We first construct a database, including a total of thirty-one patients CTs, acquired at European Institute of Oncology (Milan, Italy) on GE Medical System Light Speed (Fairfield, CT). No demographic or pathology pre-selection is foreseen. A contrast enhanced CT was performed for the treatment planning (CTsim). Brainstem, optical nerves, eyes, mandible, parotids and spinal cord were contoured by expert radiation oncologists on CTsim, serving as Ground Truth (GT) for algorithm performance evaluation. The algorithm is divided into three separate stages: selection of the most similar subset of the database, pairwise deformable registration and finally label fusion (Figure 5.1).

5.1.1.2 *Selection strategy*

We developed a method for complexity reduction and templates sorting depending on the similarity to test patient. The database volumes were pre-aligned by means of affine registration to a common reference. In our case, we chose RANDO® (The phantom laboratory, Salem, NY) , an anthropomorphic phantom commonly used in the clinic for dosimetry quality assurance. The affine alignment is done offline each time a new atlas is added to the database.

Once the test patient is also aligned to RANDO®, similarities in terms of Normalized Mutual Information (NMI) shared by each atlas and the CT to be segmented are computed. We further normalized the NMI dividing it by the maximum value and rank the atlases in the database. Based on this ranking, the group of subjects to be actually used for image segmentation is than selected in a Fixed Number (FN) or Thresholding (TH) fashion. FN implies the selection of a percentile of the database and has the main advantage of providing the same anatomical variability for each test patient. TH looks directly at NMI values and guarantees the inclusion of just the most similar atlases. Nonetheless, if the CT scan to be segmented is not well represented in the database, the first strategy could select sub-optimal atlases, while the second method could choose just one atlas. Table 5.1 sums up the values tested in this work.

5.1.1.3 *Pairwise registrations*

Once the most similar subject have been selected, we registered each atlas to the test patient using LDDD[135, 134], which has been described in §3.2.3.2. We also previously illustrated a modification of the algorithm relying on robust stopping condition (§ 4.2 and [196]).

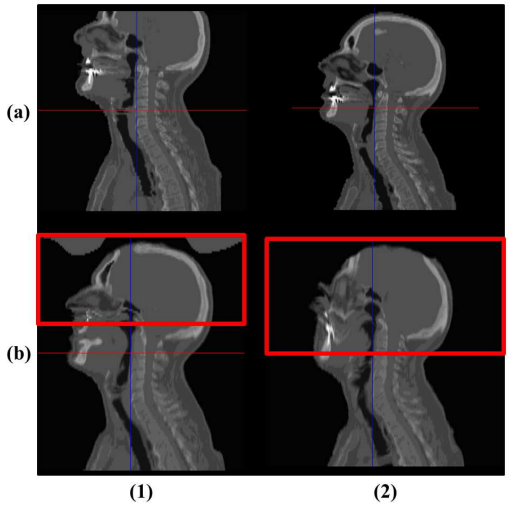


Figure 5.2: Compensation of the differences in the field of view of *reference* and *floating* images. What happens if we try to register images in Panel (a), being (a.1) the *floating* and (a.2) the *reference* volume? Possible results obtained with B-Splines and LDDD standard implementation are shown in Panels (b.1) and (b.2) respectively.

Here we also introduce an automatic compensation for different field of view of the patient scans, in order to prevent the transform being non-physical and deforming Hounsfield Unit, while trying to reproduce non existing structures. For example, let us consider the images in Figure 5.2 Panel (a), and let us assume we would register Panel (a.1) on (a.2), the latter being the *reference* image. The main issue to be prevented is the algorithm reproducing the upper part of the skull, which is not present in the *floating* image. Examples of possible resulting warped images obtained with B-Spline based registration and LDDD are shown in Panel (b.1) and (b.2), respectively. As B-splines are more flexible, the resulting volume seems to be more anatomically consistent than LDDD, which need a good initial alignment for obtaining a reasonable registration. To address this issue, in the computational step only, we canceled out those slices in the test patient that were not covered by any voxel of the atlas after an affine registration. This pre-processing step does not have an impact on the total computational time, as an affine registration is anyhow foreseen before pairwise DIR.

5.1.1.4 *Label fusion*

We propagate GT structures of the selected atlases by means of the deformation field obtained from pairwise registration. One approach for tentative segmentations recombination is averaging (AVG), thus, defining L_s as the label on atlas s amongst the S selected subjects

$$\hat{L}(x) = \text{round}\left(\frac{\sum_{s=1,..,S} L_s(x)}{S}\right)$$

A second possible approach is majority voting (MV) that is ruled by the following equation

$$\hat{L}(x) = \begin{cases} 1 & \text{if } \sum_s L_s(x) > \frac{S}{2} \\ 0 & \text{in all other cases} \end{cases}$$

The main difference between these two comparison strategies is to be found on the structures borders, where a clear majority may not be defined.

A more refined strategy uses a probabilistic framework called Gaussian Weighted (GW) fusion. It assigns to each voxel a probability of being actually part of the labeled structure, combining a Gaussian weighted image difference between patient to be segmented and registered atlas images with an exponential distance transform map of the propagated labels. The method was originally proposed by Sabuncu et al [197], for Magnetic Resonance Images, and we now extend it for CT application, where borders between structures are less well defined. We assumed the voxels to be independent realizations, thus reducing Sabuncu's model to

$$\hat{L}(x) = \arg \max_{l \in 1,\dots,L} \sum_{s=1}^{S} p(L(x) = l, I(x) | L_s, \Phi_s) \cdot p(I(x) | I_s, \Phi_s)$$

where $p(I(x)|I_s, \Phi_s)$ is the Gaussian image likelihood and $p(L(x) = l, I(x)|L_s, \Phi_s)$ is the label likelihood. The first is defined as:

$$p(I(x)|I_s, \Phi_s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(I(x) - I_s(\Phi_s(x)))^2}$$

which is a gaussian weighting of the difference between warped *floating* and *reference* image. The label likelihood is assumed proportional to the signed Maurer distance map D_s^l, thus

$$p(L(x) = l, I(x)|L_s, \Phi_s) \propto e^{\rho D_s^l(\Phi_s(x))}$$

In our case, only one label can be assigned to a single voxel, thus identifying belonging ($likelihood_1$) or not belonging to the voxel ($likelihood_0$). As soft tissues are particularly not well contrasted, we relax the hypothesis that $likelihood_1$ has to be strictly greater than $likelihood_0$ for the voxel to be included in the contour. The relationship between the two probabilities becomes $likelihood_1 > k \cdot likelihood_0$ with k being a constant between $[0; 1]$. The relationship between the two likelihoods is described graphically as in Figure 5.3, for varying k values.

We experimentally tuned k together with σ (variance of the Gaussian weight on image similarity) and ρ (scaling on the distance map) on five patients out of thirty-one, evaluating the different performances in terms of DSC variations. Given $k = 1$, we first of all evaluated possible contours for each of the five subject recombining the propagated labels according to different values of σ and ρ (Table 5.2). Once σ and ρ have been set, we tune k, testing values in the interval $[0.1; 1]$. For this purposes, we did not select the database, but rather used all the atlas as source images.

The agreement between GT and the Final Segmentation (FS) was quantified in terms of DSC, whose properties were previously described (§4.3), both for tuning parameters and for final segmentation evaluation. The statistical differences between the selection strategies as well as between the label fusion methods were evaluated by means of Bonferroni corrected Friedman test at 5% confidence level.

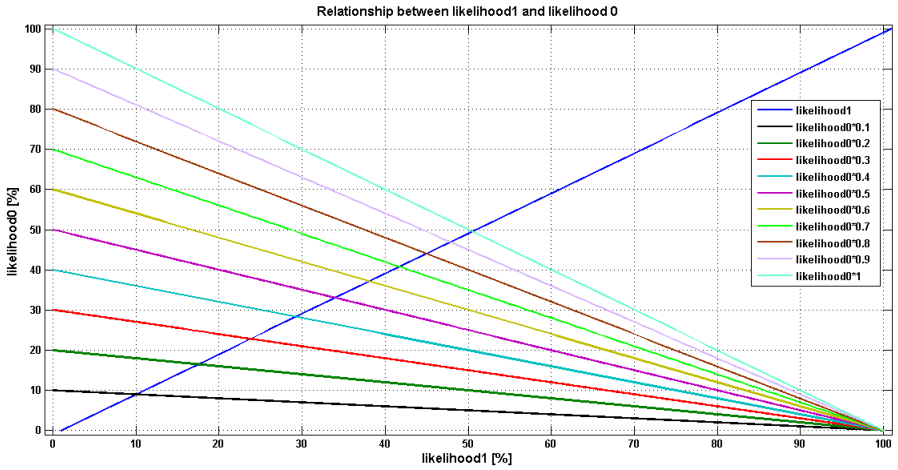


Figure 5.3: k scaling influences on the relationship between $likelihood_1$ and $likelihood_0$varying the crossing value, i.e. lowering the probability required for assigning the voxel to the target structures.

Table 5.2: Tested σ and ρ values for each of the analyzed structures

Parameter	
σ	$[1; 5; 10; 30; 50; 100; 200; 500; 1000; 2000]$
ρ	$[0.1; 0.5; 1; 1.5; 3; 10]$

5.1.2 *Results and Discussion*

5.1.2.1 *Gaussian Weighted voting parameters tuning*

Obtained results show that it is necessary to use a different combination of σ and ρ for each of the analyzed structures. Analyzed values are summed up in Table 5.3, while the corresponding results in terms of DSC are shown in Figure 5.4, 5.5 and 5.6.

The differences are connected to the relationship of σ and ρ with the organ shape and location. Change in σ discriminate between soft and boney tissue, while larger ρ values would sharpen the transition between inside and outside voxels. For example, the mandible is a well contrasted structure, generally easy to contour. In this case, we can impose a large Gaussian variance but no extra weight on the distance map is needed, as contouring is a-priori a well posed problem. For the parotid glands, instead, the border between inside and outside needs to be disambiguated from background (i.e. lower σ, but larger ρ). Although there were some differences in chosen values for symmetrical structures (e.g. left and right eye in Figure 5.5), we decided to chose same values for corresponding structures. In fact, these might be due to different image quality in the different hemispheres and therefore are non significant. We report the final chosen values in Table 5.3. Furthermore, we analyzed how k influenced the final DSC value (Figure 5.7-5.9), comparing GW trends with the Average and Majority Voting for corresponding number of patients. We notice better performance for k between 0.5 and 0.7, that

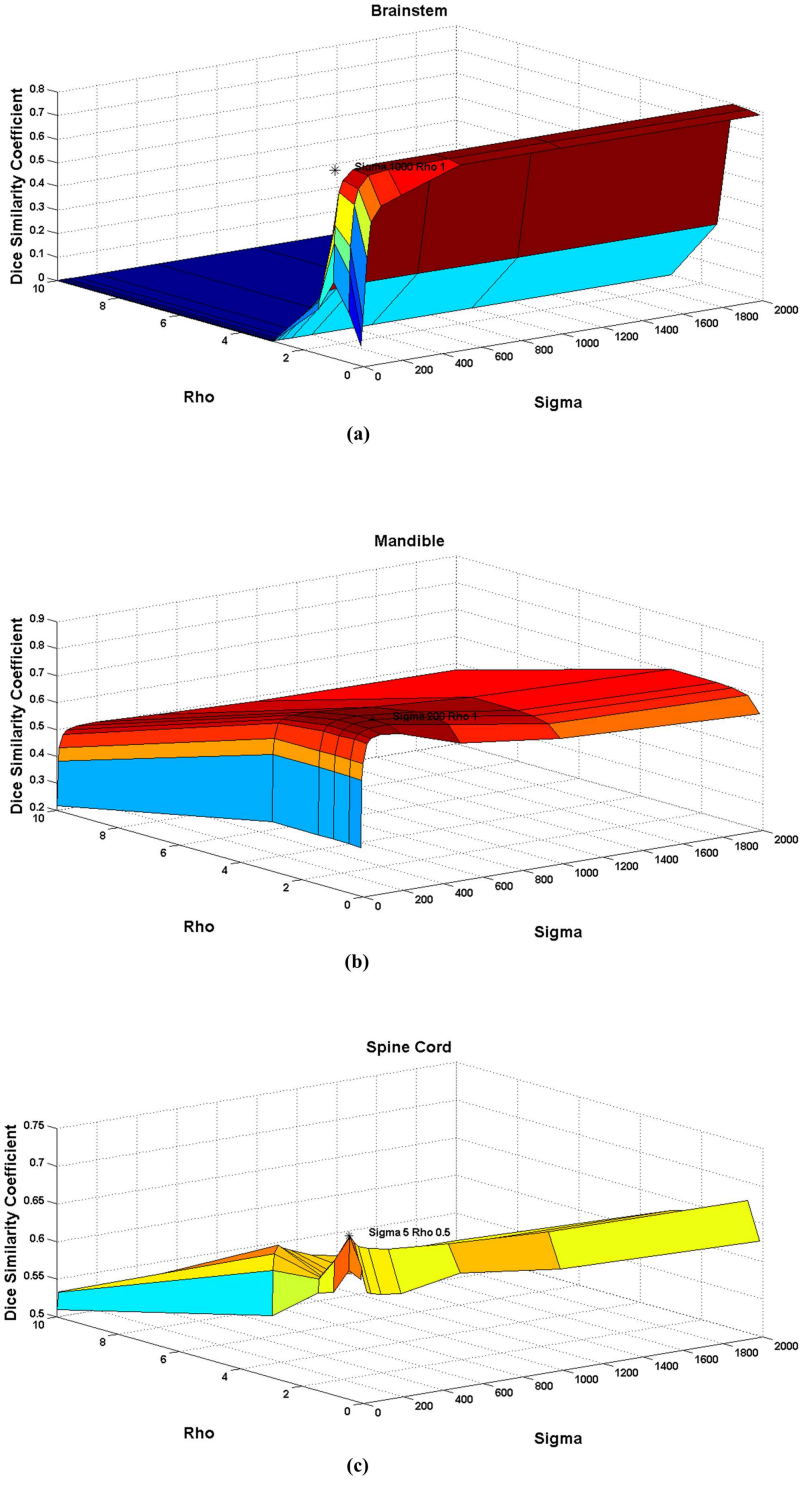


Figure 5.4: Impact of ρ and σ on segmentation in terms of variation of DSC for brainstem, mandible and spinal cord. Well contrasted structures (e.g. mandible) will need large σ, but no extra ρ.

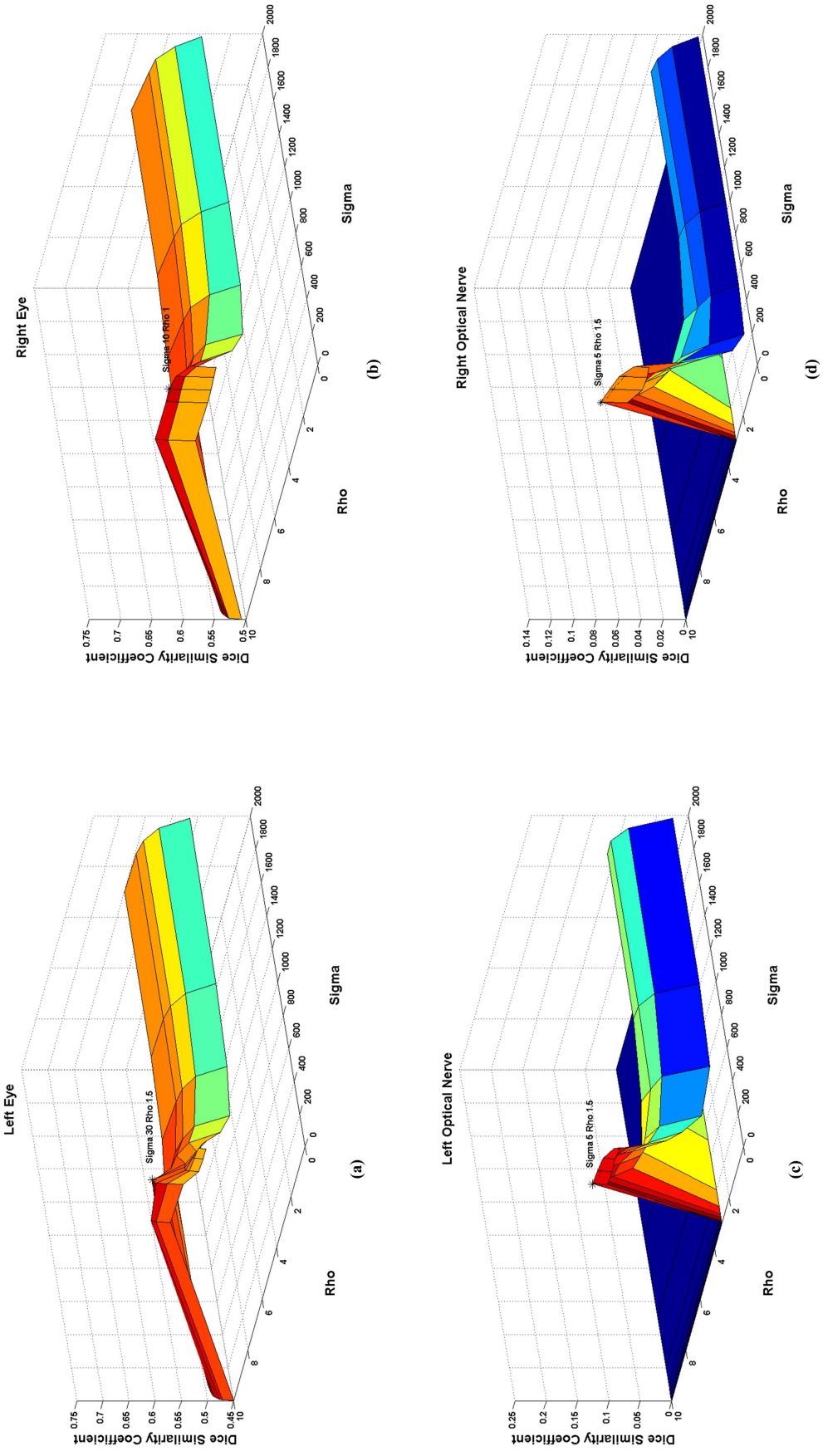


Figure 5.5: Impact of ρ and σ on segmentation in terms of variation of DSC for eyes and optical nerves. We note that there are some differences between left and right structures, likely due to differences in image quality.

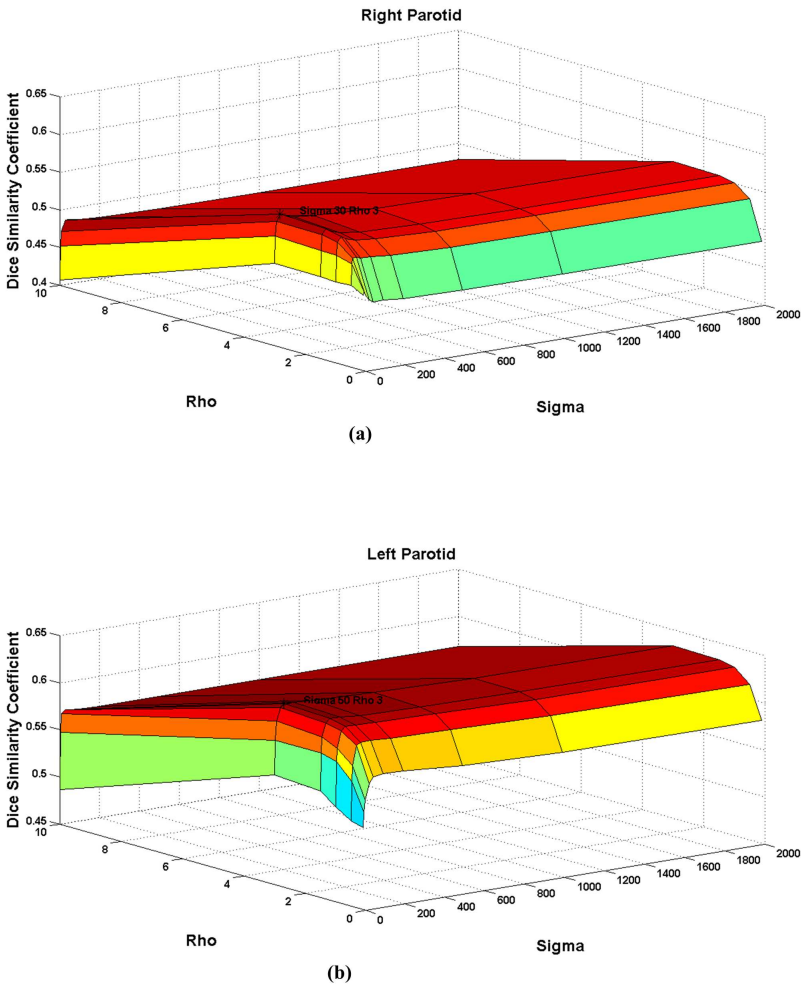


Figure 5.6: Impact of ρ and σ on segmentation in terms of variation of DSC for parotid glands. Since the border of these structures is not well defined, larger ρ will be needed.

Table 5.3: Optimal values for σ and ρ values for each of the analyzed structures

Structure	σ	ρ
Mandible	200	1
Spinal Cord	5	0.5
Brainstem	1000	1
Optical Nerves	5	1.5
Eyes	30	1.5
Parotids	50	3

we further considered in the analysis. The only structure not complying with this rule is the brainstem, for which DSC increases as a function of k (Figure 5.7).

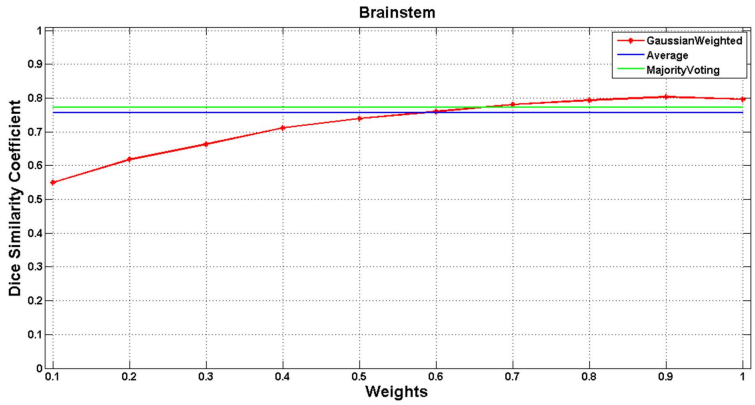
5.1.2.2 *Pairwise registrations Quality*

Before evaluating contouring results, we looked at the pairwise registrations and at the effectiveness of the introduced field of view compensation. An example is reported in Figure 5.10. Panel (a) shows overlay of patient 3 (red) and patient 26 (green) just centered in the same space, but before any registration is done. It should be noted the difference in the acquisition field of view between the two patients. Panel (b) and (c) illustrate the results of affine and deformable registration of the same two patients. This case was chosen because patient 26 has a very reduced field of view, that is particularly critical for multi-atlas based segmentation and in general for any non-rigid registration algorithm. In fact, the modelling of the deformation might be so pushed that it tries to reproduce also body districts that were not scanned in the *floating* image. If our compensation wouldn't have been effective, we would have had a final warped image similar to those illustrated in the example in Figure 5.2.

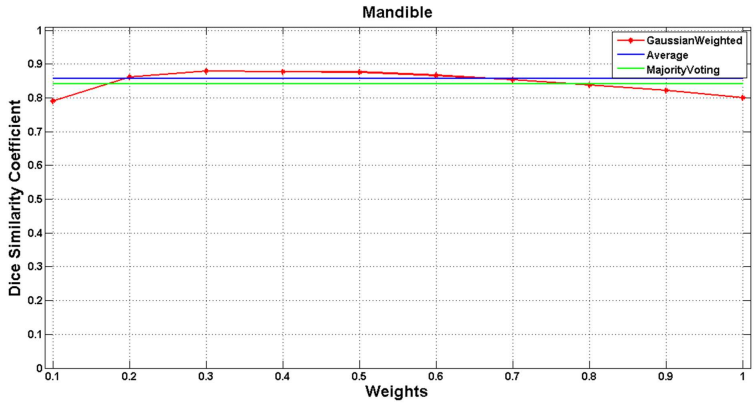
5.1.2.3 *Selection strategies and label fusion performances*

Figure 5.11 illustrate an obtained GW segmentation (solid blue) for one representative patient versus the manual outline (green) from the attending physician. It can be seen how the developed method is able to predict structure position for the test patient using a set of atlases, but refinement should be introduced both for smoothing contour irregularities and prevent holes (see for example optical nerves in Panel (d) and spinal cord in Panel (b)). Another approach to address this issue would be to add continuity constraints to the fusion rules to avoid local discontinuities. Note the very good prediction of the shape of the brainstem and of the parotid glands, which are two of the most critical structures in the HN treatments, and very influenced by the inter-rater variability.

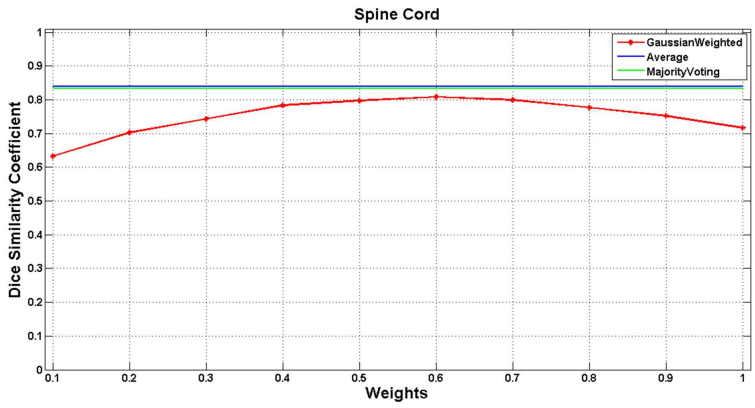
We also report the matrix of selected patients for TH and FN in Figure 5.12 and Table 5.4 respectively. Note that for TH the number of selected atlases varies with the analyzed test patient, whereas in FN different atlases are selected based on NMI ranking, but their number is constant. For some patients (e.g. 1, 10, 11, 29)



(a)



(b)



(c)

Figure 5.7: We studied the impact of parameter k on GW multi-atlas automatic segmentation in terms of DSC for brainstem, mandible and spinal cord in five patients. We compare GW DSC with AVG and MV approaches, and notice that $k = (0.5; 0.7)$ returned the best results, besides for the brainstem. It might be necessary to switch to an ogran specific tuning of k.

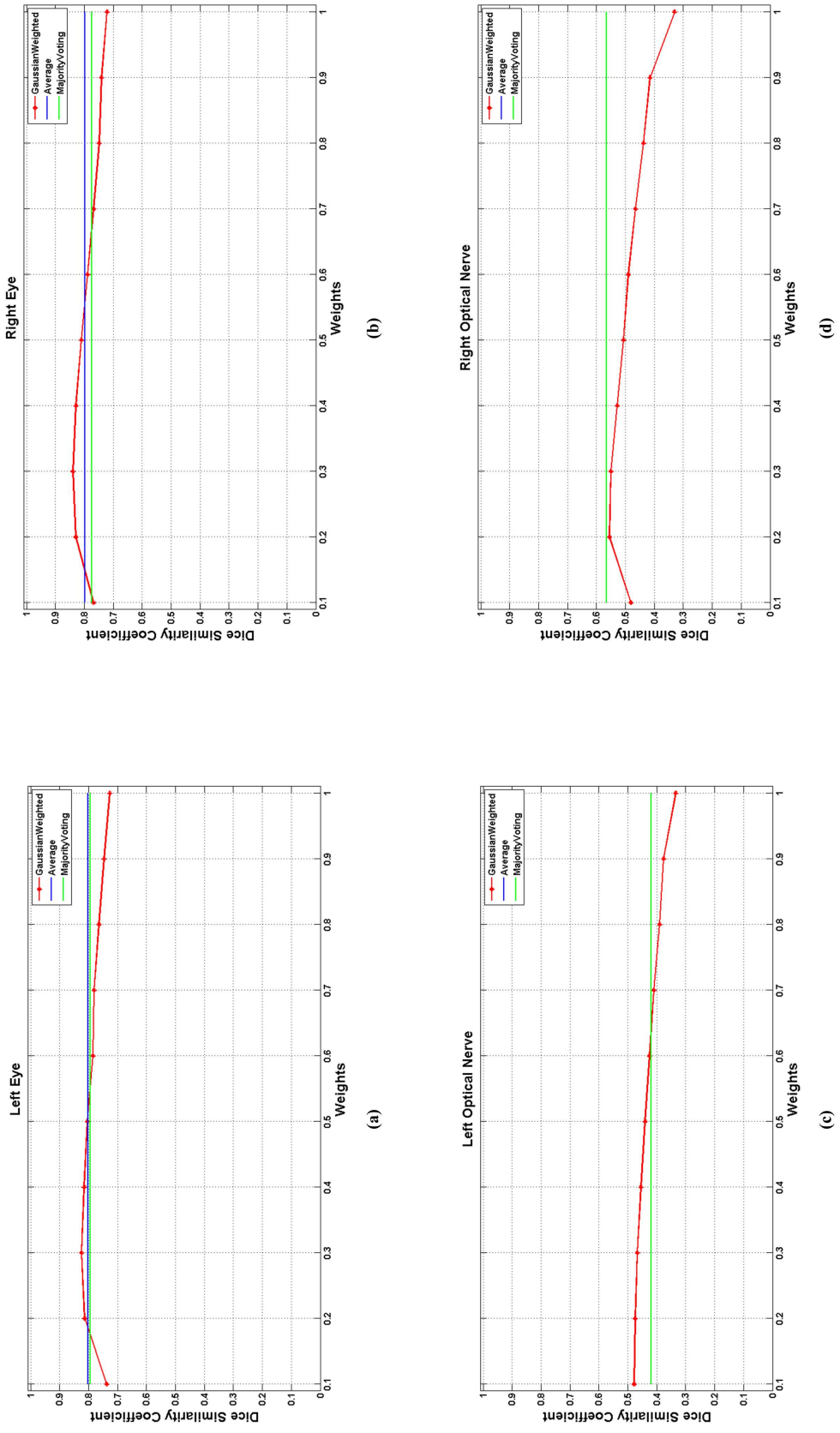
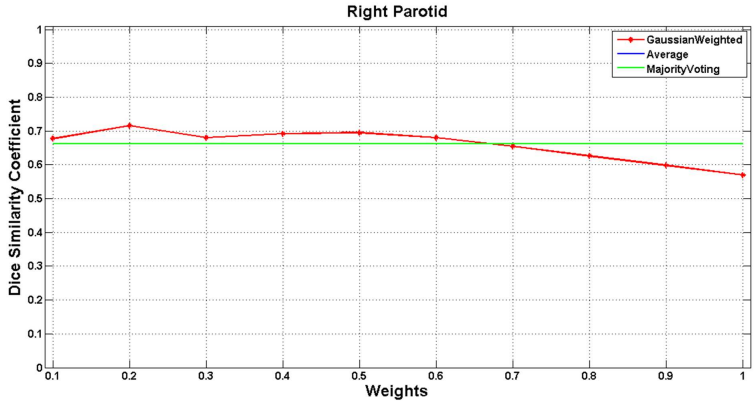
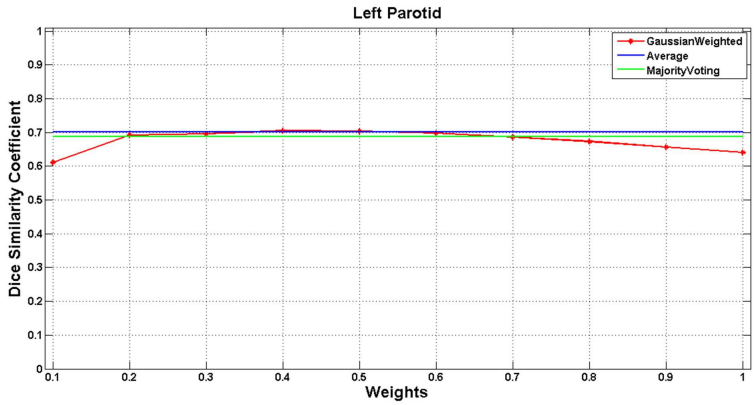


Figure 5.8: We studied the impact of parameter k on GW multi-atlas automatic segmentation in terms of DSC for eyes and optical nerves in five patients. Also in this case, $k = (0.5; 0.7)$ returned the best results.



(a)



(b)

Figure 5.9: $k = (0.5; 0.7)$ returned the best results for GW multi-atlas automatic segmentation in terms of DSC for parotid glands.

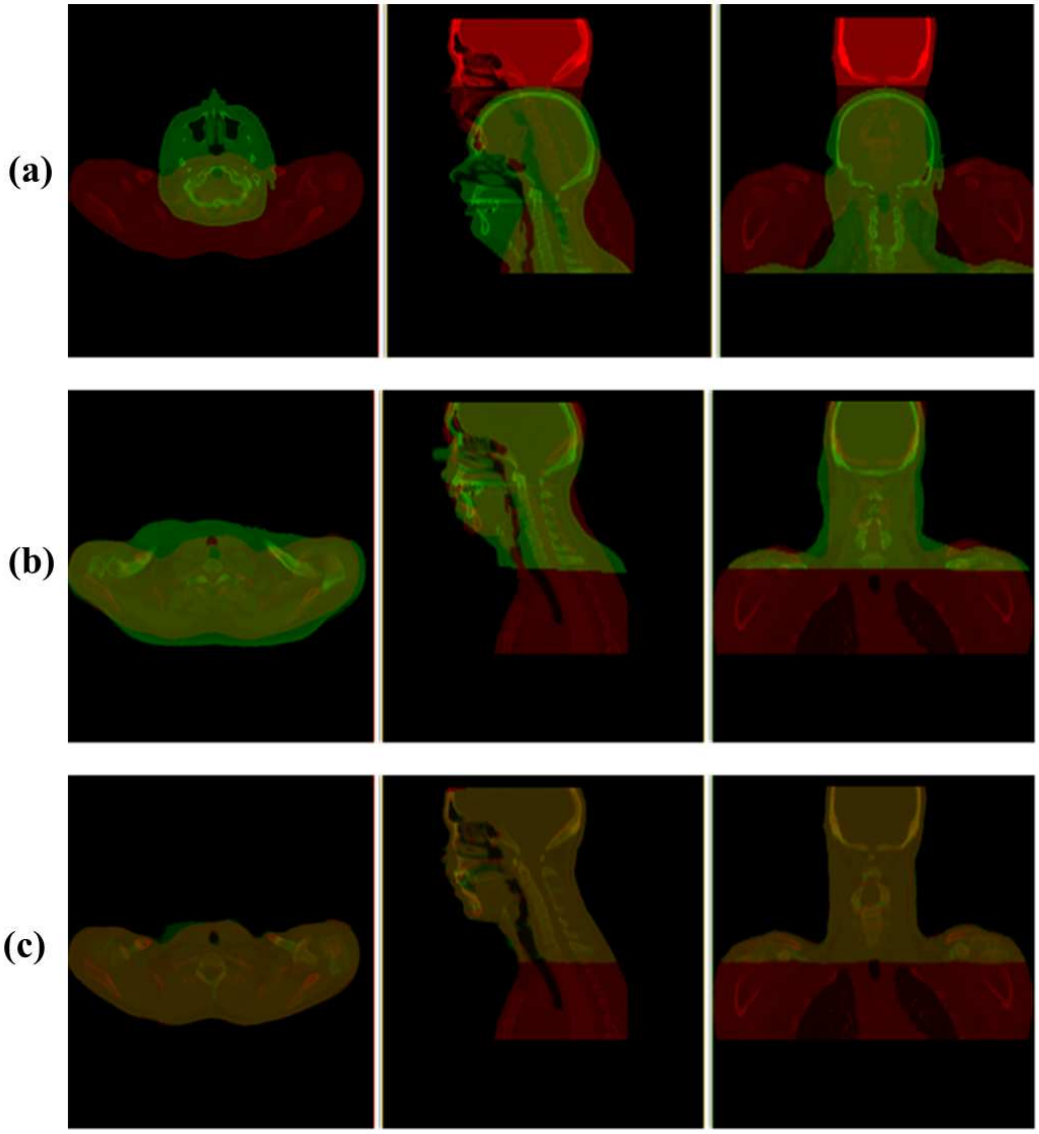


Figure 5.10: An example of pairwise registration between two patients for multi-atlas based segmentation. In red we show the test patient in this example (Patient 3) and in green the current atlas (Patient 26), from left to right axial, sagittal and coronal views. Panels (a) illustrates the initial centering of the two patients, while in Panel (b) and (c) fusion after affine and deformable registration respectively are presented. For this case, the algorithm automatically compensate for the field of view difference in the two patients, preventing unrealistic results.

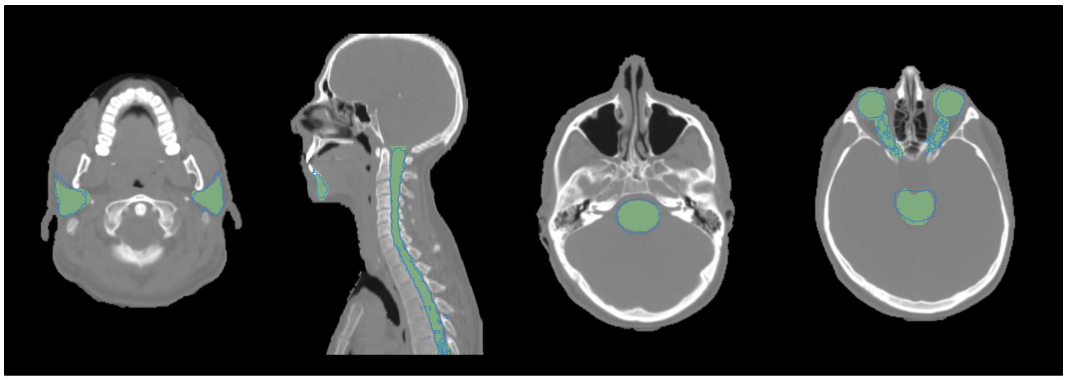


Figure 5.11: Overlay of an automatic segmentation obtained with GW multi-atlas segmentation method (solid blue) and manual outline (green) from the attending physician. Panel (a) shows the results for parotid glands, while Panel (b) the mandible and spinal cord. Panels (c) and (d) illustrate brainstem and eyes with optical nerves respectively. We note that the most difficult structures to segment are the tubular ones (i.e. spinal cord and optical nerves), which may benefit from a final stage of refinement to purge holes and border discontinuities.

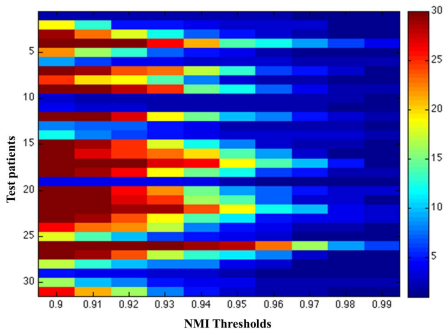


Figure 5.12: Number of patients at varying NMI thresholds. Color codes for the number of patient selected at a given patient (ordinate) for a given threshold value (abscissa).

Table 5.4: Number of patients selected for any given FN value. Note that this method does not depend on the test patient.

TH	Number of patients
0.04	30
0.50	16
0.55	14
0.60	12
0.65	11
0.70	9
0.75	8
0.80	6
0.85	5
0.90	3
0.95	2
1.00	1

just one atlas can be selected even at lower thresholds, because of the different neck positioning increasing the natural anatomical variability.

Figure 5.13 illustrates quantitative results for four selected cases. Each of the panels shows the median and percentiles of DSC values for FN and TH cases, while each column of the histograms represents one different fusion strategy. The cases in Panel (a) and (c) correspond to full database selection, while Panel (b) and (d) correspond to half of the atlas selection. Obviously, for (c) and (d), this affirmation is just approximate, as the actual subset is chosen based directly on NMI values. It shall not be forgotten that DSC values for tubular and small structures generally tends to be lower, because starting and ending voxel may differ and the metric depends on absolute volume (i.e. larger structures tends to have larger DSC).

DSC values for k equal to 0.5 and 0.7 were significantly different, both using FN and TH, with $k = 0.5$ performing better than $k = 0.7$. This is directly correlated to the relative ratio of $likelihood_0$ and $likelihood_1$, after normalization (Figure 5.14). We note a large region in which $likelihood_1$ is well below 0.5, partially due to registration errors, but also to GT contours variability.

We further looked at significant differences between tested TH and FN values. On the example of FN strategy, we report the median of the DSC distribution over the selected patients in Figure 5.15 and 5.16. Panels (a) to (b) illustrate the behaviour of all the structures at the different FN threshold according to GW in 5.15 or mean and majority voting in 5.16 . The significant degradation of DSC with decreasing number of chosen subjects discourages single-atlas approach. No significant change in segmentation performance was instead seen when selecting more than 10 atlases. If less atlases are selected, the reduction in anatomical variability and the impact of a sub-performing registration is significant. Nonetheless, the choice of a smaller subset results in a significant decrease in algorithm running time, thus making the strategy applicable in the clinic. For some structures, like

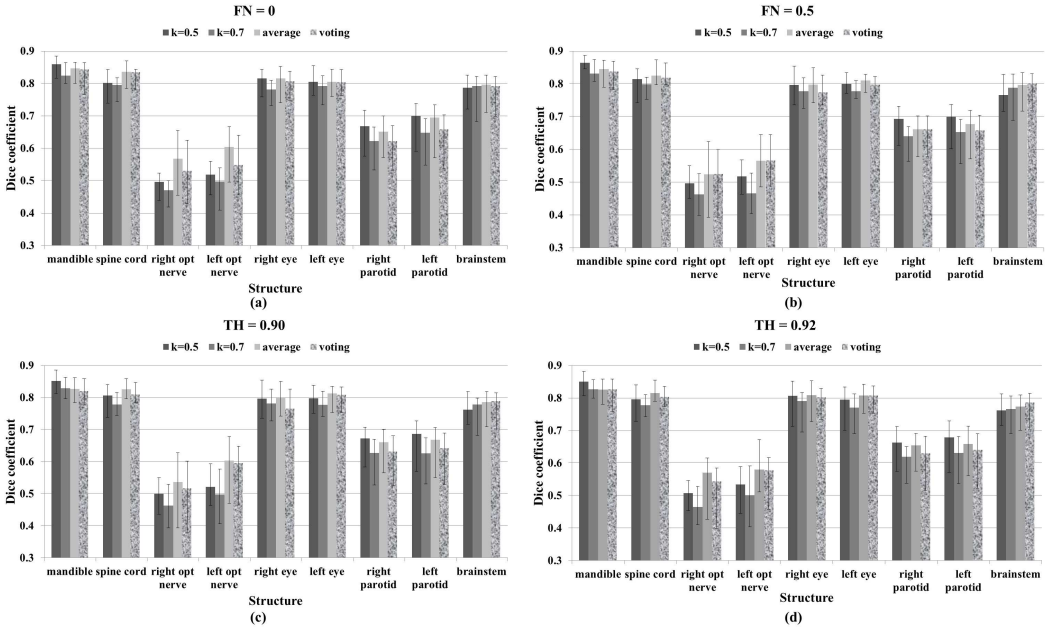


Figure 5.13: DSC values for four different selection strategies and all three fusion methods. Panels (a) and (b) show FN results for full and half database selection, whereas Panels (c) and (d) illustrates DSC results for TH value selecting approximately full and half atlases in the database. Given DSC dependence on total volume, note that it is lower for smaller structures (e.g. optical nerves). Columns indicate the different fusion strategy, i.e. GW with $k = 0.7$ and $k = 0.5$, AVG and MV.

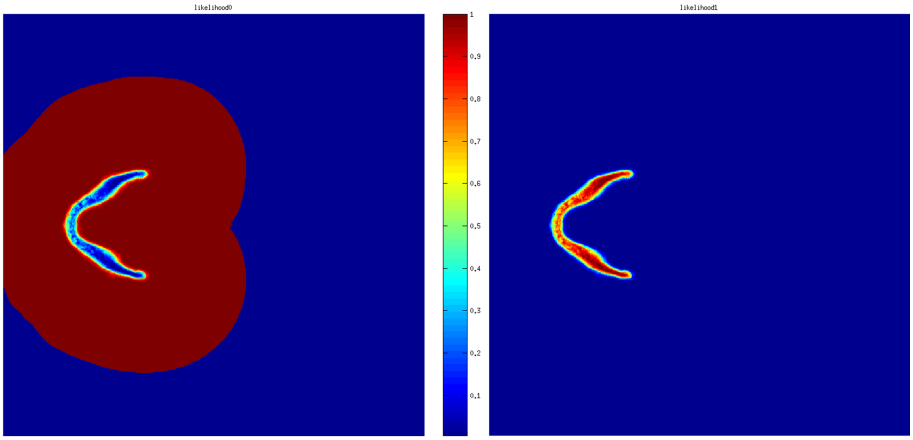


Figure 5.14: $\mathrm{likelihood}_1$(right) and $\mathrm{likelihood}_0$(left) for the case of a mandible. We note that a good amount of voxels probability do not exceed 0.4 after normalization.

eyes and optical nerves, these general rules do not hold and just the selection of the most similar patient might result in a reasonable DSC (see last entry versus all other entries in the histograms in Figure 5.15 and 5.16), because of the reduced anatomical variability at that site. The influence of the number of atlases onto the fusion rule behaviour is the same for GW with $k = 0.5$ and $k = 0.7$, the difference between the two being just the inclusion or exclusion of some voxels from the final segmented structure. The same kind of reasoning can be applied to distinguish between AVG and MV, but for a different reason. If we look at their equation, we notice that, for our database of 30 patients, AVG and MV differ only if 15 rater included the considered voxel in the structures and 15 did not. In this case, for AVG the voxel will be part of the contour, but not for MV. Also, if we look at the last entry of the histograms of all the recombination strategies, for which just the most similar atlas has been selected, we notice that GW DSC are lower than AVG and MV ones. This is due to the weighting connected to image similarity after registration multiplied to the label similarity in the GW strategy and is not present in AVG and MV.

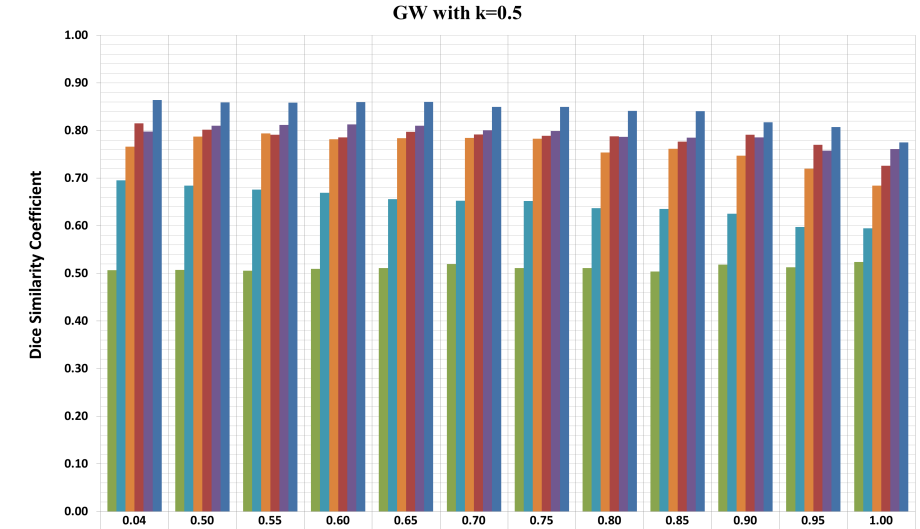
Friedman test indicates that the difference in DSC value between corresponding FN and TH results for brainstem, mandible and parotids is significant no matter which fusion rule is used. Further analysis of the ranks suggests that FN might be the slightly favourable strategy. Despite both strategies being potentially good, FN maintains an higher variability for each test patient, whilst failed or suboptimal DIR have larger impact on TH.

We further compared the recombination rules when all subjects in the database serve as atlases (i.e. FN =0 and TH=0.90). Results show that spinal cord and optical nerve may benefit from computing an average contour rather than looking at the singular voxel contribution. Besides these two structures, GW equates or beats performance of AVG and MV, suggesting its applicability.

We report a detailed comparison between GW with $k = 0.5$, AVG and MV with inter-rater variability reported in literature [94, 167, 166, 109] in Tables 5.5, 5.6 and 5.7 respectively. We can see how the multi-atlas approach is successfully reaching the inter-rater variability no matter which is the strategy used. Instead considering Table 5.8 for GW, the single atlas approach performance is not as good as the multi-atlas one. We did not report numerical results for AVG and MV, because there was no statistical difference for single atlas choice between the three approaches. The results in Tables 5.5, 5.6 and 5.7 were comparable in terms computational efficiency.

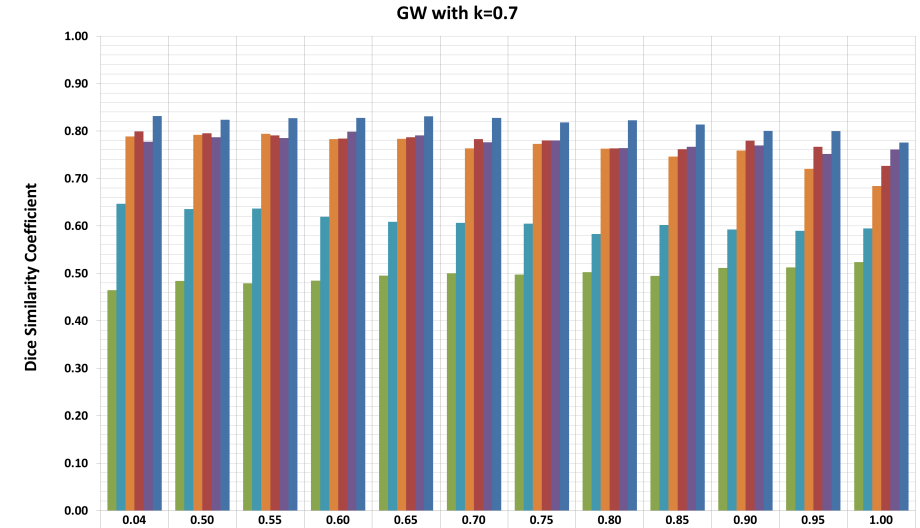
5.1.3 *Conclusion*

We developed an efficient algorithm for atlas based segmentation of RT planning CT volumes, based on demons DIR and Gaussian Weighting. The absence of preselection on atlases introduces an extra challenge besides different positioning, but also makes the results closer to real clinical application. FN selection of database atlases is foreseen to increase computational efficiency. Obtained results testify the applicability in a clinical environment, decreasing considerably the manual workload. Further work will include DR speed up, GW fusion improvement by means of a vector field discontinuity weighting term and post-processing refinement.



	0.04	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
optical_nerves	0.51	0.51	0.51	0.51	0.51	0.52	0.51	0.51	0.50	0.52	0.51	0.52
parotidis	0.70	0.68	0.68	0.67	0.66	0.65	0.65	0.64	0.64	0.63	0.60	0.59
brainstem	0.77	0.79	0.79	0.78	0.78	0.78	0.78	0.75	0.76	0.75	0.72	0.68
spine cord	0.81	0.80	0.79	0.79	0.80	0.79	0.79	0.79	0.78	0.79	0.77	0.73
eyes	0.80	0.81	0.81	0.81	0.81	0.80	0.80	0.79	0.79	0.79	0.76	0.76
mandible	0.86	0.86	0.86	0.86	0.86	0.85	0.85	0.84	0.84	0.82	0.81	0.78

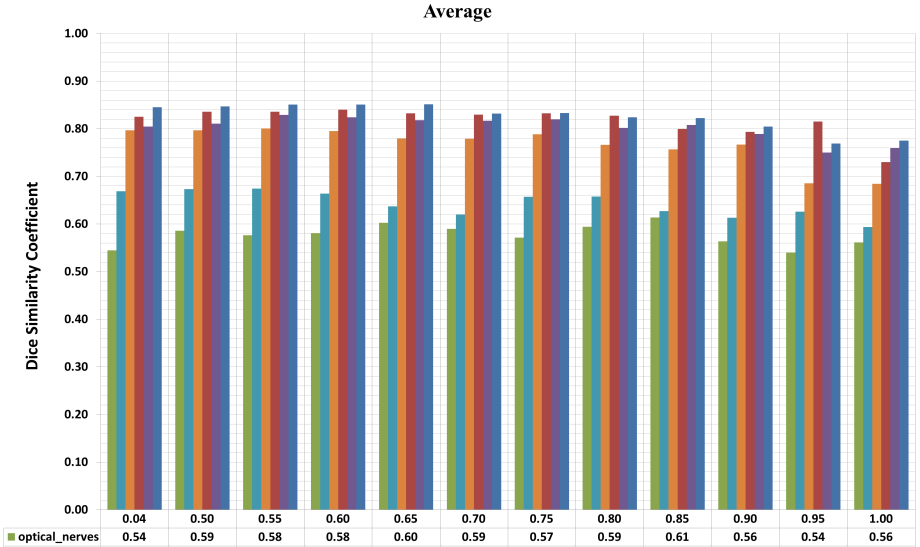
(a)



	0.04	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
optical_nerves	0.46	0.48	0.48	0.48	0.50	0.50	0.50	0.50	0.49	0.51	0.51	0.52
parotidis	0.65	0.64	0.64	0.62	0.61	0.61	0.60	0.58	0.60	0.59	0.59	0.59
brainstem	0.79	0.79	0.79	0.78	0.78	0.76	0.77	0.76	0.75	0.76	0.72	0.68
spine cord	0.80	0.80	0.79	0.78	0.79	0.78	0.78	0.76	0.76	0.78	0.77	0.73
eyes	0.78	0.79	0.79	0.80	0.79	0.78	0.78	0.76	0.77	0.77	0.75	0.76
mandible	0.83	0.82	0.83	0.83	0.83	0.83	0.82	0.82	0.81	0.80	0.80	0.78

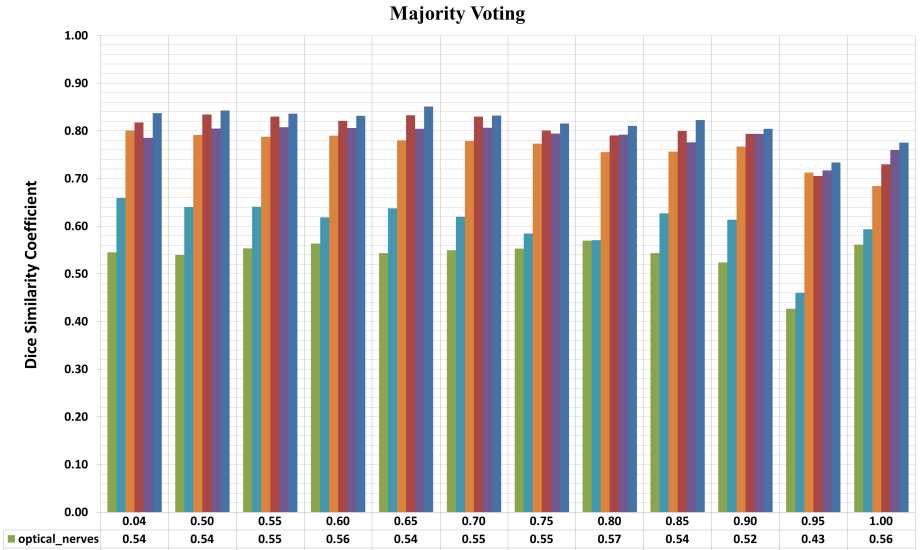
(b)

Figure 5.15: DSC for all the structures at different FN threshold according to GW fusion rules are illustrated in Panels (a) and (b). The significant degradation of DSC with decreasing number of chosen subjects discourages single-atlas approach. No significant change in segmentation performance was instead seen when selecting more than 10 atlases.



	0.04	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
optical_nerves	0.54	0.59	0.58	0.58	0.60	0.59	0.57	0.59	0.61	0.56	0.54	0.56
parotidis	0.67	0.67	0.67	0.66	0.64	0.62	0.66	0.66	0.63	0.61	0.63	0.59
brainstem	0.80	0.80	0.80	0.79	0.78	0.78	0.79	0.77	0.76	0.77	0.69	0.68
spine cord	0.83	0.84	0.84	0.84	0.83	0.83	0.83	0.83	0.80	0.79	0.82	0.73
eyes	0.80	0.81	0.83	0.82	0.82	0.82	0.82	0.80	0.81	0.79	0.75	0.76
mandible	0.85	0.85	0.85	0.85	0.85	0.83	0.83	0.82	0.82	0.80	0.77	0.78

(c)



	0.04	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
optical_nerves	0.54	0.54	0.55	0.56	0.54	0.55	0.55	0.57	0.54	0.52	0.43	0.56
parotidis	0.66	0.64	0.64	0.62	0.64	0.62	0.58	0.57	0.63	0.61	0.46	0.59
brainstem	0.80	0.79	0.79	0.79	0.78	0.78	0.77	0.76	0.76	0.77	0.71	0.68
spine cord	0.82	0.83	0.83	0.82	0.83	0.83	0.80	0.79	0.80	0.79	0.71	0.73
eyes	0.79	0.80	0.81	0.81	0.80	0.81	0.79	0.79	0.78	0.79	0.72	0.76
mandible	0.84	0.84	0.84	0.83	0.85	0.83	0.82	0.81	0.82	0.80	0.73	0.78

(d)

Figure 5.16: DSC for all the structures at different FN threshold according to mean and majority voting fusion rules are illustrated in Panels (a) and (b). Also in this case, single-atlas approach is the worst performing, whereas no significant difference was found when selecting more than 10 atlases.

Table 5.5: Comparison between literature inter-rater variability and GW results when all the atlases are selected and $k = 0.5$

Structure	FN = 0	TH = 0.9	Inter-rater variability
Mandible	0.85 (0.07)	0.86 (0.04)	$0.85 - 0.90$ [94] 0.78 ± 0.06 [167]
Spinal Cord	0.81 (0.10)	0.81 (0.10)	$0.70 - 0.80$ [94]
Brainstem	0.76 (0.10)	0.77 (0.11)	0.825 ± 0.099 [166] 0.58 ± 0.2 [167]
Right Optical Nerve	0.50 (0.12)	0.50 (0.10)	0.499 ± 0.174 [166]
Left Optical Nerve	0.52 (0.13)	0.52 (0.11)	0.499 ± 0.174 [166]
Right Eye	0.80 (0.12)	0.80 (0.12)	0.831 ± 0.094 [166]
Left Eye	0.80 (0.09)	0.80 (0.06)	0.831 ± 0.094 [166]
Right Parotid	0.67 (0.12)	0.69 (0.12)	0.853 [109] 0.66 ± 0.1 [167]
Left Parotid	0.69 (0.16)	0.70 (0.14)	0.853 [109] 0.66 ± 0.1 [167]

Table 5.6: Comparison between literature inter-rater variability and AVG results when all the atlases are selected

Structure	FN = 0	TH = 0.9	Inter-rater variability
Mandible	0.83 (0.08)	0.85 (0.08)	$0.85 - 0.90$ [94] 0.78 ± 0.06 [167]
Spinal Cord	0.83 (0.06)	0.83 (0.08)	$0.70 - 0.80$ [94]
Brainstem	0.79 (0.11)	0.80 (0.12)	0.825 ± 0.099 [166] 0.58 ± 0.2 [167]
Right Optical Nerve	0.54 (0.23)	0.52 (0.23)	0.499 ± 0.174 [166]
Left Optical Nerve	0.60 (0.21)	0.57 (0.16)	0.499 ± 0.174 [166]
Right Eye	0.80 (0.11)	0.80 (0.11)	0.831 ± 0.094 [166]
Left Eye	0.81 (0.08)	0.81 (0.05)	0.831 ± 0.094 [166]
Right Parotid	0.66 (0.14)	0.66 (0.12)	0.853 [109] 0.66 ± 0.1 [167]
Left Parotid	0.67 (0.16)	0.68 (0.15)	0.853 [109] 0.66 ± 0.1 [167]

Table 5.7: Comparison between literature inter-rater variability and MV results when all the atlases are selected

Structure	FN = 0	TH = 0.9	Inter-rater variability
Mandible	0.82 (0.08)	0.84 (0.09)	$0.85 - 0.90$ [94] 0.78 ± 0.06 [167]
Spinal Cord	0.81 (0.06)	0.82 (0.07)	$0.70 - 0.80$ [94]
Brainstem	0.79 (0.11)	0.80 (0.11)	0.825 ± 0.099 [166] 0.58 ± 0.2 [167]
Right Optical Nerve	0.52 (0.22)	0.52 (0.21)	0.499 ± 0.174 [166]
Left Optical Nerve	0.60 (0.18)	0.57 (0.16)	0.499 ± 0.174 [166]
Right Eye	0.77 (0.11)	0.77 (0.09)	0.831 ± 0.094 [166]
Left Eye	0.81 (0.08)	0.80 (0.06)	0.831 ± 0.094 [166]
Right Parotid	0.63 (0.16)	0.66 (0.12)	0.853 [109] 0.66 ± 0.1 [167]
Left Parotid	0.64 (0.16)	0.66 (0.15)	0.853 [109] 0.66 ± 0.1 [167]

Table 5.8: Comparison between literature inter-rater variability and GW results when just one atlas is selected

Structure	FN = 0.99	TH = 1.0	Inter-rater variability
Mandible	0.78 (0.06)	0.78 (0.06)	$0.85 - 0.90$ [94] 0.78 ± 0.06 [167]
Spinal Cord	0.73 (0.14)	0.73 (0.15)	$0.70 - 0.80$ [94]
Brainstem	0.68 (0.14)	0.68 (0.15)	0.825 ± 0.099 [166] 0.58 ± 0.2 [167]
Right Optical Nerve	0.51 (0.22)	0.54 (0.16)	0.499 ± 0.174 [166]
Left Optical Nerve	0.51 (0.19)	0.51 (0.19)	0.499 ± 0.174 [166]
Right Eye	0.76 (0.30)	0.76 (0.33)	0.831 ± 0.094 [166]
Left Eye	0.76 (0.09)	0.76 (0.17)	0.831 ± 0.094 [166]
Right Parotid	0.61 (0.17)	0.60 (0.14)	0.853 [109] 0.66 ± 0.1 [167]
Left Parotid	0.63 (0.16)	0.59 (0.16)	0.853 [109] 0.66 ± 0.1 [167]

5.2 ON-LINE *virtual*CT IN HEAD AND NECK ADAPTIVE RADIATION THERAPY

Daily setup variations, internal organ motion and deformation have long been a concern for external beam radiation therapy treatments. The selectivity onto the target and the sparing of surrounding structures requires accurate localization and monitoring of the patient throughout the treatment course. While optimized and sophisticated techniques (e.g. Intensity Modulated Radiation Therapy, IMRT) have been developed to enhance the dose distribution conformity around the target volumes, the actual delivery of the intended treatment is still an open issue. In ART, anatomical structures, imaged during daily patient set-up routine, are compared with those of the planning phase, for individuation and targeting of the tumor and OARs sparing. In most of the clinical workflows, the adaptation of the plan occurs on a replanning CT on the basis of relevant dosimetric discrepancies, while in-room imaging is mainly used for patient setup. Therefore, current clinical protocols require an expert physician to re-draw all of the structures, and to check carefully the dose volume histograms. As such, ART is still mostly an offline time-consuming procedure (see §2.3), which offers the potential to be, at least partially, automated. Deformable registration algorithms (see §3) stand as state of the art tools to adapt OARs and target contours, as well as to evaluate the actual delivered dose, on the basis of a displacement field obtained by image co-registration. Several DIR algorithms have been proposed, but so far no complete experimental validation of CT to CBCT registration has been provided. Nonetheless, the employment of such vector fields could lead to non-consistent (i.e. not deliverable) plans. In most of the current clinical HN protocols, CBCT images are acquired to monitor tissue changes, but the acquisition of a replanning CT (CTrepl) is still preferred.

Here, we implemented an on-line procedure to generate a *virtual*CT from the CBCT scan used for patient setup, thus dropping the need of a new replanning CT (CTrepl). We demonstrate its robustness and feasibility in a retrospective study on twenty patients treated at the European Institute of Oncology with IMRT in a conventional ART protocol. The core element of our proposed method is a deformable registration algorithm that is employed to compensate the non-rigid changes occurred between the treatment planning simulation CT (CTsim) and the control CBCTs. We assessed the equivalence between virtualCT and CTrepl by computing residual deformation, and ART strategy performance according to multiple metrics.

5.2.1 *Methods*

5.2.1.1 *Dataset*

The study was retrospectively performed on twenty HNCP treated at European Institute of Oncology (Milan, Italy). All the patient underwent IMRT with a planned dose of 60-70Gy delivered in 2-2.12Gy per fraction, one fraction per day. Current protocol entails a planning CT (GE Medical System Light Speed, Fairfield, CT) with or without contrast medium injection, verification CBCTs (On-Board Imager, OBI, Varian, CA) and replanning CTs at 40Gy and 50Gy. CBCT and CTrepl are acquired generally one day apart, with 3 days distance as the worst case. Image

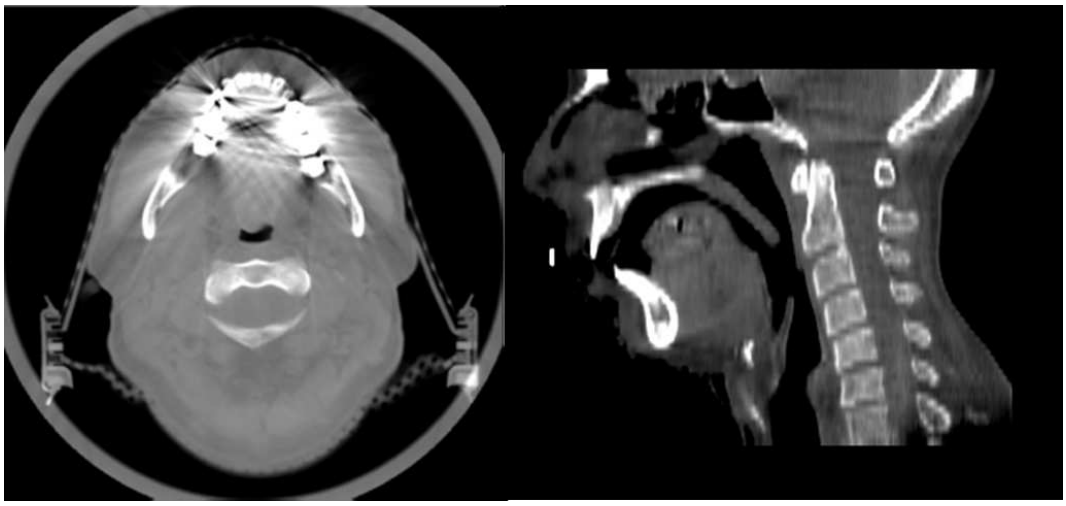


Figure 5.17: Initial CBCT volume (left) with fixes and ring effect versus pre-processed image (right). We can clearly note the contrast enhancement, as well as the removed tube and couch.

resolution was 1x1x3 mm in AP-LL-SI directions, on average. Patient position reproducibility during imaging and treatment phases is guaranteed by customized head-neck and shoulders thermoplastic masks fixed onto the couch, bite block, and three radiopaque markers attached to the mask for initial alignment. Main indication for replanning is represented by the over dosage of OARs, as highlighted by Dose Volume Histograms on CTrepl, but no beam modification and/or plan reoptimization was employed in our population. Offline adjustment of the structures of interest and dose re-calculation was performed on the CTrepl. For each patient, we selected one or more triplets (CTsim, CBCT40Gy/50Gy and CTrepl40Gy/50Gy), raising the total number of cases to twenty-eight. Mandible, parotid glands, nodal Gross Tumor Volumes (nGTV) were retrospectively drawn on CTrepl by a radiation oncologist. In seven patients, it was not possible to contour nGTV, as these were either post-operative or post responsive induction chemotherapy patients.

5.2.1.2 *Image pre-processing*

The pre-processing was necessary to prevent the DIR algorithm to be stacked into local minima and/or not to converge. CTsim, CBCT and CTrepl were preprocessed, in order to remove most of tube, treatment couch, immobilization mask and, for the CBCT, ring artifacts. Given that both fixes and treatment couch were in the same HU range of the patient, thresholding or region growing methods were not effective. CT scans were masked with the body contour drawn in the treatment planning phase. The 3D scattered points describing patient body were rendered to a binary mask, which was subsequently applied onto the original image. The outside of the patient was filled with a conventional value, chosen to be outside of the HU range (i.e. -1200 HU). The same approach could not be used for CBCT scans, which were not contoured. To mask out non-patient areas, we rigidly align the CTsim body contour to the CBCT scan and then dilate it to take into account any possible failure of the registration and/or differences in the patient. Morphological

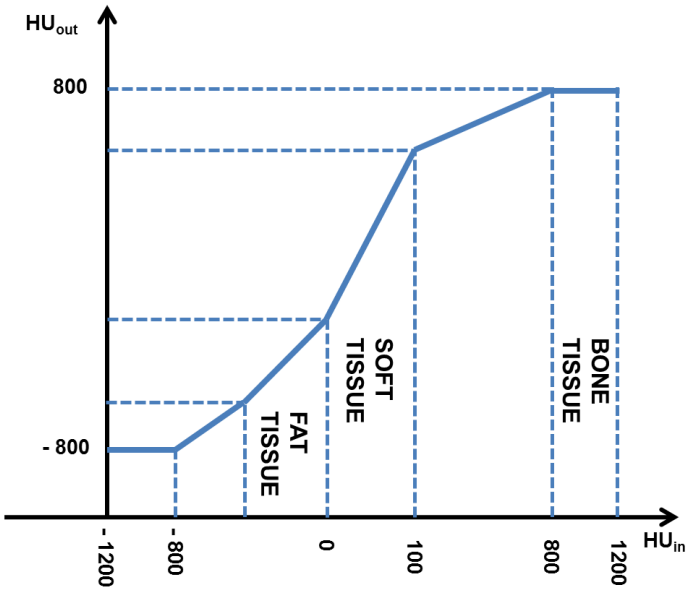


Figure 5.18: Modified HU relationship to enhance soft tissue contrast. In abscissa we report HUs of the input image, which are converted to output HUs by means of a piecewise linear relationship. Basically, we mapped any given HU value in the input image to another HU value, so that the HU range for the soft tissues is enlarged in the output pre-processed volume.

operations were then applied to further refine the contours. An example of initial versus pre-processed images is shown in Figure 5.17.

Further pre-processing includes a piecewise linear histogram modification to enlarge the HU range assigned to soft tissues. The injection of contrast media at CTsim and the presence of dental artifacts in the CBCT increased the differences between volumes as well as the histograms. The latter were filtered out by clipping the histogram to the 99.9 percentile and modifying the HU graph so that the soft tissues can be assigned more HUs (Figure 5.18).

Finally we performed an histogram matching using Slicer3D implementation [198], with 256 bins and 30 control points, between CT and CBCT volumes and histogram clipping to the 99.9 percentile to compensate for dental artifacts and differences introduced by contrast media injection between CT and CBCT volumes. The final evaluations as well as the *virtual*CT are nonetheless generated from the original images, thus avoiding pre-processing influence on the final image and on dose calculations on the virtualCT, but helping the registration to achieve more reasonable results.

5.2.1.3 *virtualCT generation*

To generate the *virtual*CT, a multistage B-Spline deformable registration (see § 3.2.3.1) between CTsim and CBCT was performed following rigid registration (Figure 5.19). For the rigid coarse alignment, we used the fast rigid registration routine based on mutual information offered in Slicer3D [198]. Non rigid registration was performed using Plastimatch [127, 129, 126, 124, 128], under the hypothesis that this ideally compensates for non-rigid modifications as detected by in-room CBCT

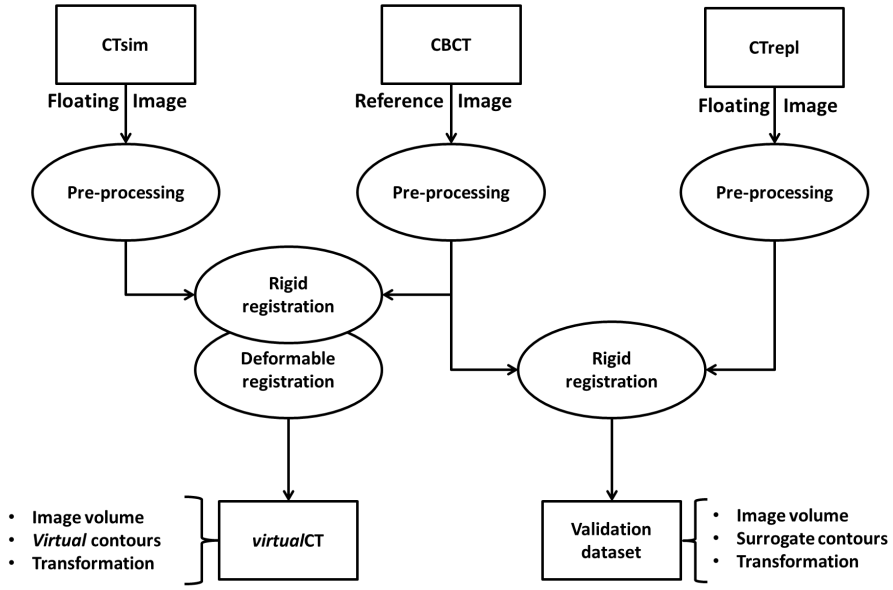


Figure 5.19: Flowchart of the proposed strategy for *virtual*CT generation. First of all we pre-process CTsim, CBCT and CTrepl, masking out all the structures outside the patient. CTsim is then deformed on CBCT volume and virtual CT is generated. CTrepl is rigidly aligned with CBCT, serving as validation dataset.

scanning. We did use one combination of parameters for this study, to make our approach independent from the tuning of the registration parameters and thus clinically feasible.

5.2.1.4 *Geometric validation*

Given the absence of a manual segmentation on CBCT, we performed a rigid registration between CTrepl and CBCT, to compensate for patient re-positioning occurred between the two scans. Only those structures whose contours were adjusted by the physicians in the replanning phase were evaluated. The contours propagated on CBCT from CTsim (i.e. virtual contours) and from CTrepl (i.e. CBCT surrogates contours) were compared in terms of Dice Coefficient (DSC), Center Of Mass (COM) distances and Surface Distances (SD). DSC is defined as in [165]:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

where A and B are the two structure volumes to be compared. Taking into account the resolution of our images, a residual absolute coordinate-wise COMs distances lower than 1 mm in LL and AP and lower than 3 mm in SI direction is indicating a good registration. To evaluate the local surface distances, after triangulation of the contour points, we computed the Root Mean Square Error (RMSE) as

$$RMSE = \frac{\|d\|}{N}$$

where $\|d\|$ is the norm of the vector of the shortest Euclidean distance between all the N correspondent closest points in the datasets without any further refinement.

The correspondence between points is defined as in [116]. For RMSE the desired level of accuracy is 3 mm.

We estimated the deformation between CTsim versus CTrepl in terms of structures volume changes as well as COM shifts, whereas the absence of deformation between CBCT and CTrepl was verified by performing an additional deformable registration to estimate the residual vector field length distribution within contoured structures. Finally, we computed Spearman correlation coefficient R (p-value = 0.05) between the residual and estimated deformation and RMSE values on the contours, looking for inaccuracies in the validation method and polarization in the algorithm, respectively.

To verify registration quality on operator-independent points, we applied Scale Invariant Feature Transform (SIFT) method illustrated in chapter 4.3. First of all, we evaluated the entity of deformation comparing CTsim with CBCT, after solely centering and resampling in the same voxel space. Than we compared *virtual*CT and CBCT, CTrepl with CBCT (CTrepl_rig) and finally CTsim rigidly registered onto CBCT (CTsim_rig) with CBCT. Theoretically, if all registration methods were equally good, the accuracy value of these last three couples would be exactly the same. We computed accuracy as the median of the residuals displacements located points on the first and on the second image.

5.2.2 *Results*

First of all, we analyzed variations in volume and COM position for nGTVs and parotid glands between CTsim and CTrepl (Table 5.9 and 5.10), without any registration. Negative variation indicates volume shrink, while positive differences in COMs to shift to the right of the patient in LL, anterior in AP and cranial in SI respectively. For parotid glands, manual contours confirm the results by Ricchetti et al [48] of volume loss and COM medial shift, while GTV volume shrinkage gives an indication of therapy response, though its direction is non-predictable.

In Figure 5.20, we present a checkerboard rendering of patient 19, demonstrating the visual agreement of *virtual*CT with CBCT images and contours (in solid blue and yellow respectively). To better highlight the difference between clinical practice and our method, we also overlaid the contour that would have been obtained by rigid registration between CTsim and CBCT (green). Critical issues raised in some patients, such as 6 and 7 (Figure 5.22a and 5.22b respectively). These were mainly due to either sub-optimal registration performances induced by external features like immobilization mask (Figure 5.22a.1) or to macroscopic modifications between CBCT and CTrepl (i.e. different jaw position, Figure 5.22b.2).

The residual median deformation between CBCT and CTrepl was between 0.13 mm and 1.8 mm in the mandible, for all the patients but patient 7. For this patient, the median deformation measured 3.80 mm, due to jaw displacement in the CTrepl scan compared to CTsim (see Figure 5.22). Also for the parotid glands and nGTV, we reported the same behavior, though the residual could step up and occasionally exceed one voxel. For nGTV, median was 0.96 mm, with maximum of 75th percentile at 3.43 mm. Median residual deformation was 0.59 mm and 0.68 mm, whereas the maximum of 75th percentile reached 2.15 mm and 1.71 mm for right and left parotid, respectively. When correlating the RMSE contour distance with

Table 5.9: Right and Left Parotids variation between CTsim and CTrepl at 40Gy (case a) and 50Gy (case b) of treatment. Negative ΔVol and ΔLL/ΔAP/ΔSI indicate volume reduction and medial shift, respectively.

	Right Parotid				Left Parotid			
	ΔVol $[cm^3]$	ΔLL [cm]	ΔAP [cm]	ΔSI [cm]	ΔVol $[cm^3]$	ΔLL [cm]	ΔAP [cm]	ΔSI [cm]
1a	−2.54	0.14	0.04	−6.75	−1.88	0.07	0.28	−6.61
1b	−1.63	0.20	−1.46	−7.28	−1.56	0.06	1.32	7.29
2	−9.18	0.01	0.48	0.84	−1.84	0.09	0.70	1.22
3a	−4.33	−0.10	−0.17	−2.38	−0.46	0.08	0.26	2.58
3b	−8.31	0.03	−0.31	−2.09	−1.76	0.14	0.42	2.29
4	−14.51	0.33	5.75	−0.41	−9.03	0.29	5.69	0.35
5a	−12.61	0.34	2.04	−4.16	−5.99	0.03	1.96	3.77
5b	−12.82	0.53	2.07	−3.89	−10.40	0.02	1.94	3.73
6a	−14.18	0.39	1.13	−4.06	−12.91	0.08	1.00	4.15
6b	−11.92	0.02	0.81	−4.06	−11.39	0.22	0.87	4.34
7	−10.49	0.31	2.93	−0.86	−8.41	0.09	2.81	0.53
8	−7.57	−0.23	2.88	−2.18	−10.42	0.39	3.12	1.93
9	−4.54	0.46	3.58	−1.18	−4.56	0.39	3.21	1.09
10a	−3.13	0.20	−1.43	−6.49	−4.42	0.14	1.31	6.32
10b	−6.56	0.04	−3.02	−6.84	−5.91	0.15	2.94	6.64
11	−13.27	−0.03	0.92	6.56	−8.24	0.22	0.88	6.36
12	−0.47	0.15	1.67	−4.46	−4.76	0.02	1.21	3.97
13	−6.58	0.78	−2.91	3.02	−7.26	0.09	3.04	3.54
14	−10.18	−0.74	−4.45	−0.36	−10.35	1.10	4.40	0.22
15	−4.89	0.33	0.84	−6.05	−5.71	0.13	0.66	6.13
16a	−7.75	0.15	0.00	−4.10	−0.44	0.23	0.05	3.68
16b	−7.23	−0.08	−0.08	−3.54	−6.00	0.06	0.00	3.36
17a	— — —	— — —	— — —	— — —	−11.97	0.47	1.93	5.62
17b	— — —	— — —	— — —	— — —	−16.93	0.54	2.00	5.42
18	−0.17	0.03	2.95	−1.25	−0.55	0.14	2.66	0.86
19	−2.79	0.24	3.58	−7.93	−4.79	0.35	3.51	7.57
20a	−9.12	0.02	2.78	−4.01	−4.24	0.51	2.64	4.26
20b	−5.02	0.10	2.91	−4.35	−2.32	0.17	2.56	4.23

Table 5.10: Gross Tumor Volume (GTV) variation between CTsim and CTrepl at 40Gy (case a) and 50Gy (case b) of treatment. If more than one GTV was present, both are displayed in the same row. Negative ΔVol and ΔLL/ΔAP/ΔSI indicate volume reduction and medial shift, respectively.

	$\Delta Vol[cm^3]$	$\Delta LL[cm]$	$\Delta AP[cm]$	$\Delta SI[cm]$
1a	−0.11	0.20	0.11	−6.90
1b	−0.34	0.01	−1.41	−7.08
2	−2.70	0.02	0.76	0.95
3a	−0.45	0.30	−0.26	−2.37
	−0.52	−0.01	0.43	−1.73
3b	−0.64	0.55	−0.58	−2.37
	−0.66	0.42	0.11	−2.08
6a	−0.70	0.21	1.20	−4.07
6b	−0.76	−0.30	1.03	−4.26
7	−0.77	0.03	2.99	−1.07
9	−6.28	0.89	3.39	−1.13
11	−0.27	−0.13	1.04	6.35
12	−0.18	−0.13	1.52	−3.90
13	−6.37	0.26	−2.66	3.26
14	−0.09	−0.92	−4.89	0.19
15	−15.58	0.27	0.98	−6.10
16a	2.23	0.23	0.45	−3.49
16b	2.53	0.02	0.27	−3.38
18	−12.28	−0.85	2.58	−0.64
19	−9.92	−0.80	3.86	−10.95
20a	−7.46	−0.04	2.28	−4.21
20b	−8.48	0.13	2.56	−4.41

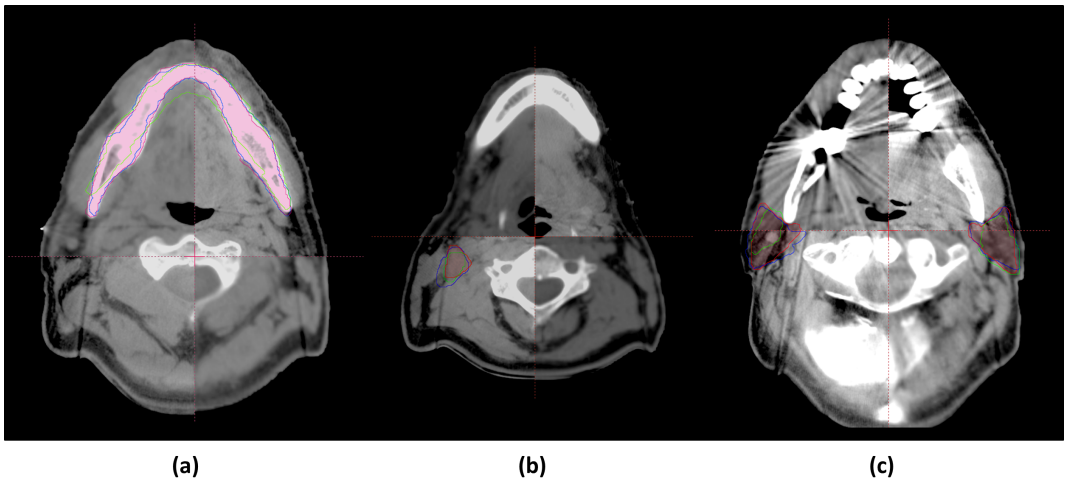


Figure 5.20: Checkerboard rendering of three axial slices of patient 19a, comparing virtualCT and CBCT scans. Manually outlined (red) and automatic (blue) mandible, nGTV and parotids are presented in panels (a), (b) and (c) respectively. In green, we report the contour that would have been obtained by rigid registration of CTsim on CBCT. nGTV contour highlights the variability of the obtained results.

the estimated CTrepl/CBCT non-rigid transform, the most evident trend can be seen for the mandible, for which we can clearly classify patient 7 as an outlier of the distribution. Three other outliers were detected, as having a higher RMSE (> 3 mm), despite the low residual deformation, as results of registration errors mainly due to metal artifacts and contouring variability. The corresponding Spearman coefficient is significant, but the correlation is rather weak (R=0.67).

In Figure 5.23, we report median, 25th and 75th percentile of DSC, COMs distance, and in Figure 5.24 RMSE between the *virtual*CT and the CBCT surrogate contours. For the mandible, we report a median DSC of 0.85 and COMs distance around 0.6 mm in all directions. Only patients 1b, 7, 12, 15 and 16a show a RMSE equal or greater than 3 mm (Figure 5.24), localized at transition between bone and teeth. As underlined by Figure 5.22 and the residual deformation estimation, the different mandible positioning causes validation method failure for patient 7. In fact, if we perform non-rigid registration between CTrepl and CBCT, DSC steps up to 0.84 and COM difference goes down to less than 0.5 mm along each dimension. For both parotid glands, median DSCs were 0.78 and 0.79 and median COMs distances were lower than 1 mm along LL and AP direction, and lower than 1.5 mm along SI direction. Looking at RMSE for both parotids, we notice very few cases, in which the distance was above the 3 mm threshold. At first sight, the most critical structure appears to be GTV, for which median DSC was just 0.58 and RMSE violated the acceptance threshold in seven patients.

We then looked at linear dependency between RMSE and volume difference or COM shift (Figure 5.25). In all cases, but for nGTV, RMSE versus volume difference, R was non-significant. Although significant, the low R value for nGTV RMSE versus volume difference confirms the difficulty to reproduce nodal volume position when subjected to large deformations, besides the variability in the results. The absence of a significant trend between RMSE results and the deformation between

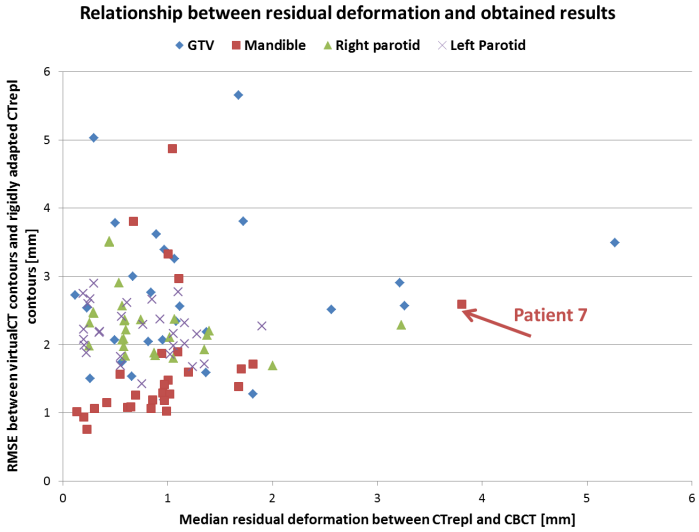


Figure 5.21: Distribution of results obtained with our strategy in terms of RMSE contour distance and CTrepl/CBCT residual deformation. The trend can be seen for mandible, where we can clearly classify patient 7 as an outlier of the distribution (red arrow). Other outliers can be ascribed to registration errors mainly due to metal artifacts and contouring variability. For GTV and parotid glands the behavior looked very similar, though more variability can be seen from analysis of GTV – residual deformation relationship.

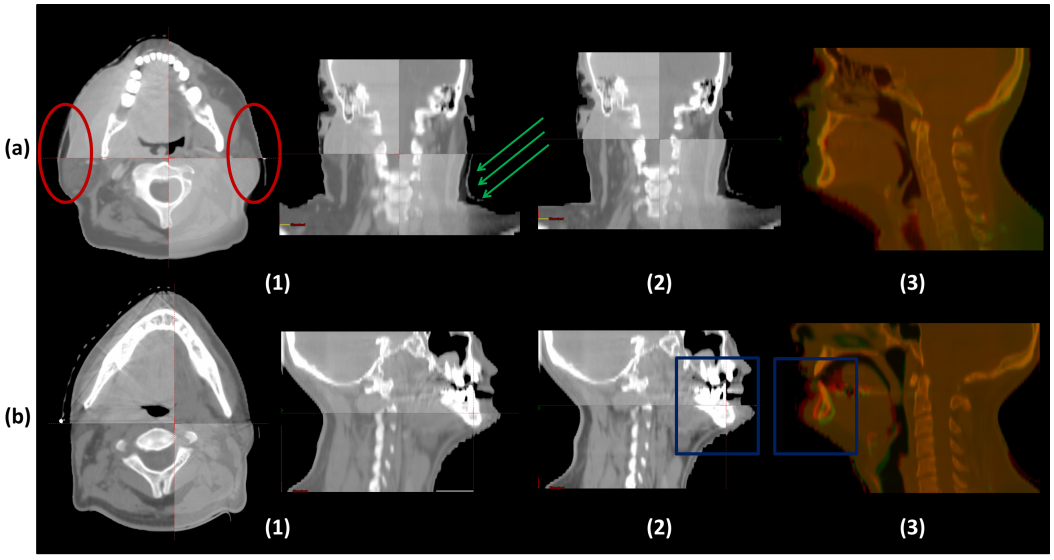


Figure 5.22: Checkerboard rendering for patient 6a and 7 virtualCT and CBCT after deformable registration are presented in panels (a.1) and (b.1) respectively, while panels (a.2) and (b.2) shows the comparison between CTrepl and CBCT after rigid correction. In the image overlay (a.3, b.3), red codes for CBCT and green for rigidly registered CTrepl. Note the effect of a sub-optimal masking of CBCT in the red circles in panel (a.1): we clearly see how the deformable registration does not detect the mask (green arrows) as external structure, because the immobilization is in the same HU of patient tissue. For patient 7, we could not use the proposed validation method, because of the different jaw position in CTrepl and CBCT (blue squares).

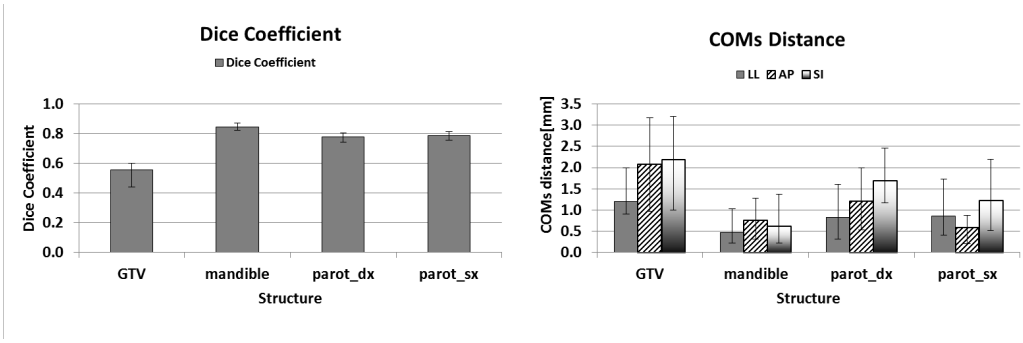


Figure 5.23: Quantitative results about the geometrical alignment between CBCT surrogate contours and structures propagated from CTsim with the same deformation used to generate virtualCT. Besides GTV, deformable registration reached a median DSC of 0.8. DSC inverse correlation with structure volume and low contrast are at the basis of a lower score for GTV. COM distances were in median in agreement with the prescribed accuracy.

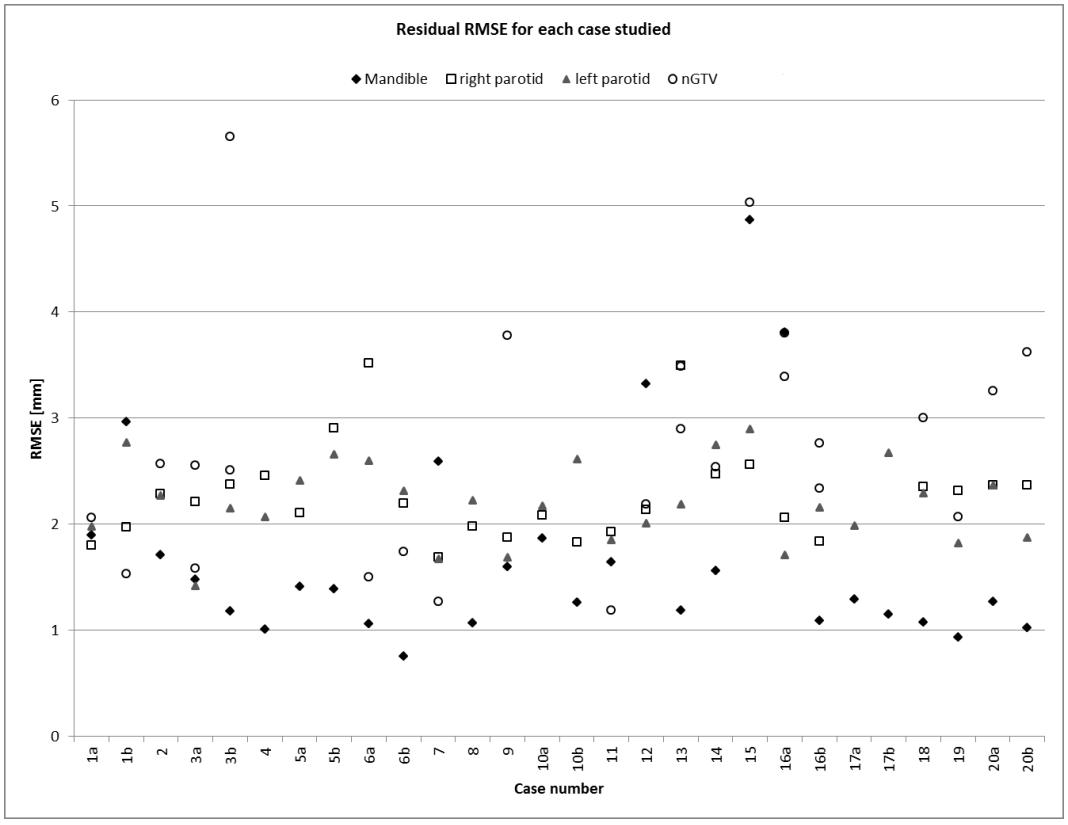


Figure 5.24: RMSE between points of the surrogate CBCT and *virtual*CT contours was used to quantify local residual surface distances. Each patient is analyzed separately, being 3 mm the acceptance threshold for this metric.

Table 5.11: Medians and variability (computed as difference between 75^{th} and 25^{th} percentiles) for all patients at different cases obtained with SIFT.

	median [mm]	inter-quartile range [mm]
CTsim/CBCT	31.52	15.953
CTsim_rig/CBCT	1.35	0.9958
*virtual*CT/CBCT	0.66	0.2744
CTrepl_rig/CBCT	0.66	0.3848

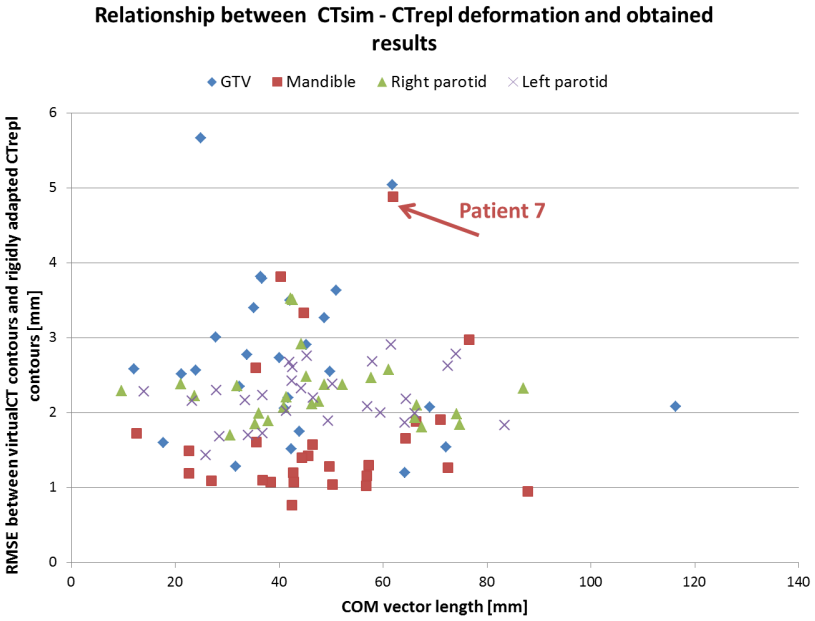
Table 5.12: Medians of DSC, COM and RMSE between *virtual*CT and CTrepl_rig for different structures.

	Mandible	Right Parotid	Left Parotid	GTV
DSC	0.85	0.78	0.79	0.56
COM [mm]	1.09	2.24	1.60	3.24
RMSE [mm]	1.4	2.21	2.23	2.55

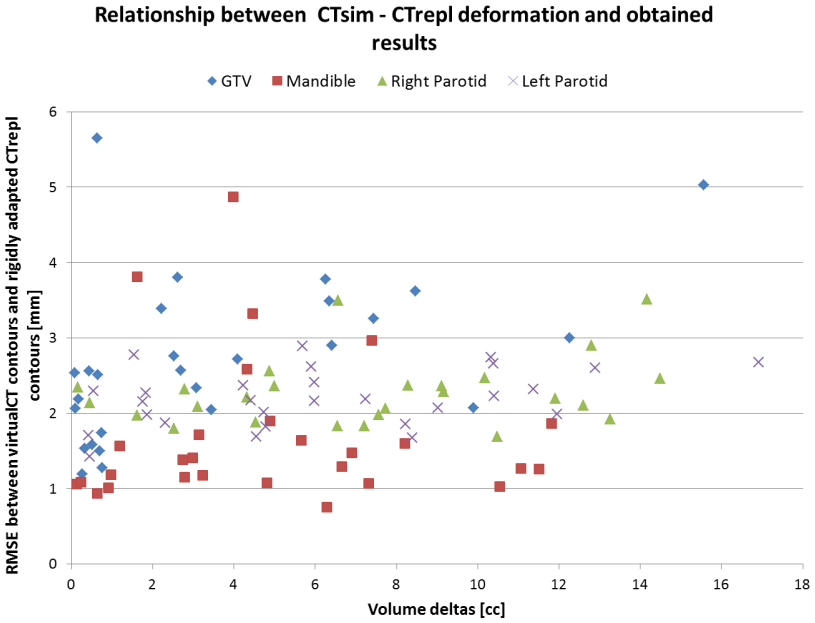
CTsim and CTrepl implicates that our strategy is able to compensate deformation in a clinical range, as represented by our patient database.

The application of SIFT to the pair CTsim/CBCT shows largest values in accuracy, thus reflecting the presence of macroscopic anatomical changes. To check whether a deformable registration is actually needed to compensate for these changes, we compare SIFT landmarks detected on CTsim_rig and *virtual*CT on CBCT respectively. The improved accuracy in the mapping of *virtual*CT features with CBCT ones, with respect to the rigid case, confirms the better performances reached by non-rigid registration. We further verified if *virtual*CT could actually be a well enough replacement for CTrepl, comparing *virtual*CT and CBCT and CTrepl_rig and CBCT accuracies. The Friedman test demonstrates the statistical difference of CTrepl_rig and CBCT and *virtual*CT and CBCT from CTsim_rig and CBCT. In Table 5.11, we report the median and the variability of the median accuracy for all patients for the different cases. Finally, by comparing results obtained from SIFT (Table 5.11) and those obtained from classical indices (Table 5.12) between *virtual*CT and CTrepl_rig for all patients, we see that SIFT is an accurate method, providing an index which expresses the accuracy in mm, in contrast with DSC, and is not bound to the sources of error related to the contours (for example manual identification), such as COM and RMSE are. Further work shall be dedicated to eliminate the normalization of the initial image, which results in a lower identification probability of soft tissues in favor of bony structures.

SIFT outliers are due to an incorrect slice attribution and to the fact that features matching is based just on image content (i.e. descriptor similarity), without further distinguishing or clustering the landmarks for example on image relative location. An example of wrong association is reported in Figure 5.26, where obviously the discriptor is not distinguishing between the mastoid in the left and right emisphere.



(a)



(b)

Figure 5.25: The relationship between RMSE results (y axis) presented in the manuscript and estimated deformation between CTsim and CTrepl (x axis). This latter is quantified in terms of COM vector length (panel a) or volume difference (panel b) of contoured volumes. Each structure is presented separately, but no trend can be noted.

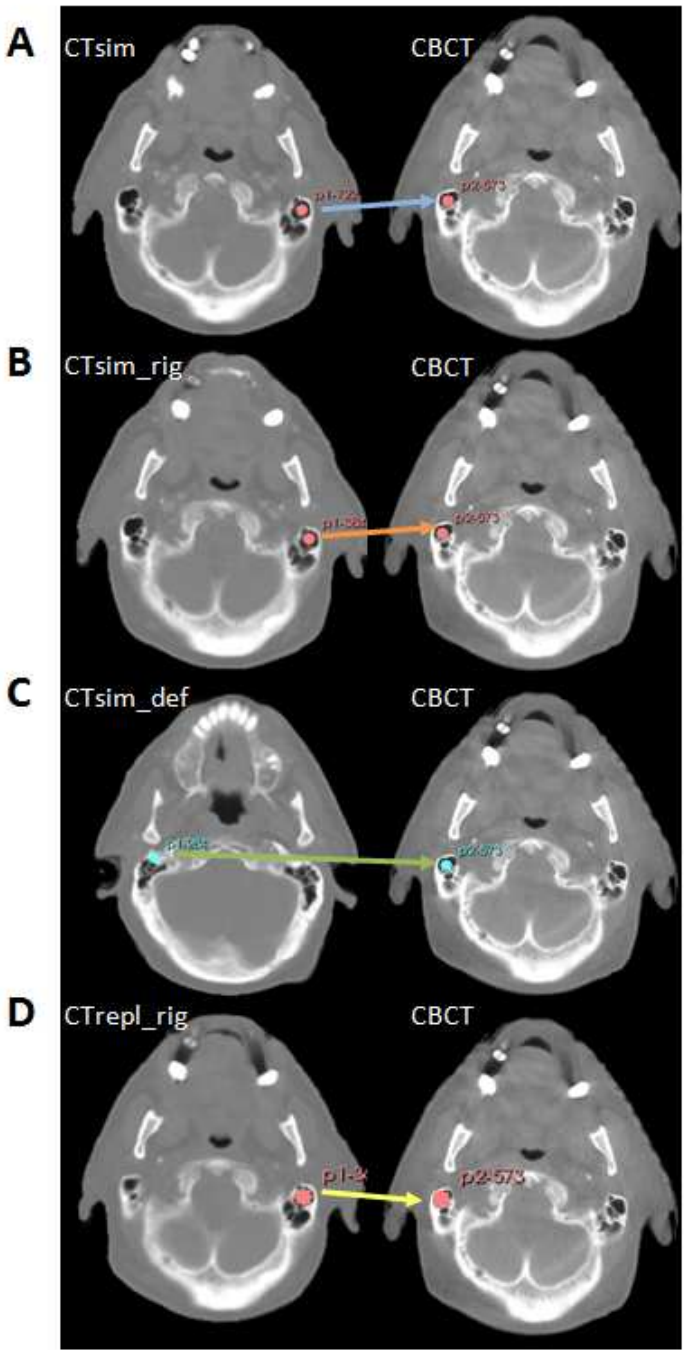


Figure 5.26: Example of wrong SIFT correspondent features matching for symmetric features (mastoid). The descriptor distance fails in correctly matching CTsim, CTsim_rig and CTrepl_rig features, whereas in Panel (c), for *virtual*CT the features are correctly associated.

As for the computational efficiency, the algorithm running times were on the order of 5 minutes with volumes of size 512x512x116 on a laptop mounting an Intel® Core™2 Duo 2.2GHz, with 2GB of RAM.

5.2.3 *Discussion*

In this work, we proposed a method to generate a virtualCT that could guide both contouring and dose adaptation in a reasonable computational time. This would boost the clinical feasibility of ART decisively, while sparing the patient CTrepl scan and preventing treatment delay. The whole concept relies on a deformable registration strategy that warps image and structures from the planning CT to the daily CBCT. Our dataset is composed by 28 triplets of CTsim, CBCT and CTrepl, which were re-contoured by an expert physician. The intra-observer variability may not be ignored in this case, but we tried to limit it by allowing only one session of volumes re-contouring per patient. Despite our efforts in masking out the ring effect and immobilization mask, residual influence was caused by comparable density range with respect to the patient body, thus hindering the successful application of threshold- or region-growing-like refinement and of the whole registration process.

The validation method we propose relies on the acquisition of CTrepl sufficiently near in time to the considered CBCT (at last within three days). In fact, if only rigid modifications occur, it is possible to define CBCT surrogates volumes from CTrepl, to be used as golden standard comparison with virtualCT structures. Otherwise, the involved structures need to be excluded from the study. For instance, patient 7 mandible drawn on CTrepl could not be used as verification tool, because of the different jaw position in the corresponding CBCT volume (Figure 5.22).

We obtained results that are comparable with previous work on monomodal registration for RT (Figure 5.23 and 5.24). Zhang et al. [107] applied deformable image registration to CT volumes, obtaining distances between manual and automatic contours below 3 mm in each patient included in the study. Shekhar et al. [199] reported an accuracy of approximately 2 mm in the thorax and 3 mm in the abdomen for a fully automated FFD-based deformable registration of 4D CT phases, while Nithiananthan et al. [169] reported a Target Registration Error (TRE) of (1.6 ± 0.9) mm in CBCT/CBCT demons deformable registration. A multimodal image registration study was recently performed by Faggiano et al. [109], who applied B-Spline based warping of head and neck between CT and Mega-Voltage CT, where the variability of manual contouring is significant. They reported an average DSC of 0.77, average surface distance of 1.66 mm and maximum surface distance of 9.47 mm. Hou et al. [110] quantified the discrepancies in terms of TRE on 9 anatomical points, showing residual mean TRE of 2.4±0.2 mm for bony targets and of 2.8±0.2 mm for soft tissue landmarks, after symmetric forces demons registration of CTsim and CBCT.

The expectation of the mandible having the best results among all the re-contoured structures, because of its high contrast on the CT scan and the absence of non-rigid deformations, was confirmed by the quantitative results. With our strategy, we are able to compensate parotid glands shifts and shape differences, as predicted in [48], with accuracy comparable to the image resolution, allowing for a constant

plan update between fractions, to avoid permanent side effects induced by irradiation, such as reduced salivary flow rate. The maximum of difference (around 1 cm) can be found at the frontal part of the parotids, where the separation between gland and masseter is not clear in CBCT, and in the lateral part of the contour, where the deformation is under-estimated because of errors induced by masking. For those patients in which the GTV was clearly contrasted the RMSE and COMs distances matched the axial voxel resolution of 1 mm (see for example patient 19, Figure 2) and DSC was greater than 0.7. Otherwise, deformable registration was not able to compensate for different shapes of GTV. In fact, for seven patients, the GTV was immerse in a rather uniform soft tissue, as it was already reported from Zhang et al. [107], and also the contoured lesion was very small, thus enhacing the probability of mis-positioning and of contouring differences as well as the negative correlation between structures volume and DSC value.

Our approach could be used to quickly generate a dataset of modified contours on the daily imaging. A quick dose re-calculation could be performed, based on which the current plan could be verified quantitatively, relying on either the measured deformation or the calculated DVHs. All these factors would allow a more robust identification of replanning indication according to a clinic specific protocol. Moreover, this work may provide an insight into the larger objective of validating CT to CBCT DIR. Local tissue contrast enhancement, penalizing the sharp transitions, and mask refinement would limit inaccuracies in registration and further refine the deformation map. This latter can be used for cumulative dose evaluation as well as for treatment plan fast update. Dosimetric evaluation will be needed to compare virtualCT and CTrepl distributions and introduce virtualCT concept into the clinic.

5.2.4 *Conclusion*

We were able to generate a virtualCT and the associated segmentation of OARs and nGTV for ART purposes, based upon the information gained from the in-room CBCT scan and the application of a deformable image registration algorithm. We aimed at sparing a CT scan (CTrepl) to the patient by modeling the inter-fractional deformations, thus avoiding any delay in the therapy as well as reducing the clinical workload. Computational time was reduced to the point that the strategy can be defined clinically feasible. Further work will anyhow be needed to improve nGTV localization. Our approach could therefore be appropriate for a clinical ART implementation, dropping the need of a CTrepl and improving the overall clinical workflow.

6 CONCLUSIONS

The aim of this work was the automatization of manual offline and online phases of a Radiation- or Hadron-Therapy treatment. We analyzed both technological issues connected with DIR applicability and clinical feasibility of innovative protocols for multi-atlas segmentation and *virtual*CT generation for Image Guided Adaptive Radiation- and Hadron-Therapy.

We studied the performaces of two of the most popular approaches, i.e. B-Spline or free-form registration and Log Domain Diffeomorphic Demons, in a clinical case of same patient CT-to-CT registration. The two CT scans were acquired at different time points along the treatment, in correspondance of physicians observation of macroscopic non-rigid changes in the patients. We compared the performances of these two algorithms with the ones of a commercial algorithm, developed by MIMvista (MIMsoftware, Cleveland, Ohio, USA) and currently clinically approved. Both implemented strategy demonstrated to be comparable to the commercial one, but B-Splines was slightly better and presented no significant difference with the MIMvista algorithm. The quantitative evaluation of these algorithms was performed evaluating contour mean overlap (i.e. Dice Similarity Coefficient) between manual and automatic contours obtained applying the computed deformation field to the manual outlines on the *floating* scans.

We then analyzed technological aspects connected with the problem of compensating the non-rigid deformations occurring throughout the treatment course. The key component here is intensity-based DIR, which aims at fusing two or more image datasets, while modelling the occurred transformation. This class of algorithms has gained popularity, but yet it is not validated for clinical applications. We addressed the problem of both ensuring full convergence of the iterative optimization and of providing a robust independent indicator of performances to the clinical user. The first question was solved studying different metrics properties on the example of Log Domain Diffeomorphic Demons algorithm. The best performing stopping condition is looking at the minimum of a set of previous iterations and it is based on the calculation of the harmonic energy of the transform, rather than just looking at the evolution of the intesity difference. Providing a final deformation field free of discontinuities and with limited amount of irregularities is primary indicator of a physically meaningful transformation and, as such, is of primary importance in a clinical context. In addition, the independence of the chosen criteria from image intensities makes it more robust to artifacts, but at the same time reliable and accurate. The study of this type of conditions is still limited to synthetic deformations applied to phantom data and/or to patients, so a patient dataset specific tuning of the acceptance threshold should be foreseen. This study was conducted on the example of Log Domain Diffeomorphic Demons, but could be easily applied to any type of registration algorithm. Attention should be dedicated also to the improvement of computational efficiency, especially for calculations on the displacement field. For B-Spline based methods and more generally

for all the parametric methods, a step forward could be to study the relationship between harmonic energy and all the other convergence criteria with the coefficients of the final transformation, so that the iteration of the cost function can be restricted onto the deformation coefficients rather than working on all the voxels in one image volume.

For technological developments, we further studied the properties of classical evaluation indices (i.e. validation and performance cost functions) and propose a new independent metric based on the extraction of salient features from the image volume itself. Although much could still be done to improve Scale Invariant Feature Transform computational efficiency, the invariance properties of the developed algorithm to translation, rotation, scaling and non-rigid deformation suggest its applicability to a clinical environment for an objective validation and assessment of DIR. This would overcome both the limits of visual inspection with its subjective judgements and of classical indices like contour or intensity based ones, which are not able to provide critical evaluation of the registration in a single number. An accurate validation and development of alternative approaches to matching of detected feature, such as adding spatial information on features locations in the image volume and/or using Iterative Closest Point algorithm, would overcome issues introduced by the descriptor solely based on features neighbooring voxel intensities and image gradients and disambiguate symmetric features.

With a reliable DIR at our disposal, we worked on the automatization of the initial planning segmentation, which in the current clinical practice is relying just on manual contouring and/or thresholding and region-growing methods. Although these are definitely feasible approaches for well contrasted structures, such as spine cord and/or mandible, they fail when facing more complex problem such as brainstem and/or parotid glands outline. We developed a flexible and robust multi-atlas based approach relying on DIR and on an efficient strategy for selecting the patient to be used as atlas for the current patient.The algorithm can be seen as a sequence of three steps:

1. Individuation of the suitable atlases for the current test patient, after its prealignment in a common reference frame.

2. Pairwise deformable registration and contour propagation of the selected atlases and test patient.

3. Fusion of the tentative segmentation to obtain the final contours for the test patient.

We analyzed the performance of different selection strategies and gave guidelines for the number of included atlases, underlining the insufficient performance of a single-subject approach. We furthermore developed an efficient and robust workflow to automatic compensate for field of view difference between the current test patient and each atlas in the database, in order to prevent the calculation of physically non meaningful deformations during the pairwise deformable registration. The developed strategy proves to be reliable and not adding further burden to the already computationally demanding registration step, while limiting user interaction to parameter tuning in the recombination step. Finally the performances of three different fusion rules were analyzed at the same time adapting the algorithm

to the need of a radiation oncology clinic and to the peculiarities of radiation planning CT volumes. We demonstrated the applicability of the method in the clinical reality achieving results comparable to clinical inter-rater variability on a database of thirty-one patients. Future work will imply the introduction of a refinement step for preventing holes and irregularities in the contouring. From clinical point of view an enlarged validation of the method is foreseen together with dosimetric evaluation, to ensure the clinical applicability and robustness.

Finally, we were able to generate a *virtual*CT and the associated segmentation of Organ At Risks and Gross Tumor Volumes (GTVs) for Image Guided Adaptive Radiation Therapy (IGART) purposes, based upon the information gained from the in-room CBCT scan and the application of a deformable image registration algorithm. We aimed at sparing a CT scan (CTrepl) to the patient by modeling the inter-fractional deformations, thus avoiding any delay in the therapy as well as reducing the clinical workload. Computational time was reduced to the point that the strategy can be defined clinically feasible. We showed that virtualCT contours are a representation of patient GTVs, parotids and mandible volumes at fixed treatment time points, as the calculated deformation is able to capture shape and position changes with an accuracy comparable to image resolution. Our approach could therefore be appropriate for a clinical IGART implementation, dropping the need of a CTrepl and improving the overall clinical workflow. The causes of failure individuated for the proposed strategy (e.g. GTV larger error due to reduced contrast) could be addressed by means of landmark-based image registration localized around tumoral lesion. Further improvements will include the use of regularization term and the study of dosimetric changes in the updated plan. The proposed strategy is nonetheless already clinically feasible. Besides the technological innovation introduced in this work, the patient undergoing head and neck Intensity Modulated Radiotherapy Treatment would have great benefit in terms of reduced damages to parotid glands (i.e. less swallowing and preservation of normal function) and of neoplastic lesion local control.

All the work of the present dissertation represent a step forward not only to the introduction of deformable registration into the clinical practice, but also to the development of a feasible strategy for online head and neck treatment adaptation, aiming at improving patient therapy outcome as well as quality of life.

BIBLIOGRAPHY

[1] M. Durante and J. S. Loeffler, "Charged particles in radiation oncology," *Nature Reviews Clinical Oncology*, vol. 7, no. 1, pp. 37–43, 2009.

[2] "Report 50. prescribing, recording and reporting photon beam therapy." International Commission on Radiation Units and Measurements (ICRU), Tech. Rep. 50, 1993.

[3] "Report 62. prescribing, recording and reporting photon beam therapy (supplement to ICRU Report 50)," International Commission on Radiation Units and Measurements (ICRU), Tech. Rep. 62, 1999.

[4] M. van Herk, "Errors and margins in radiotherapy," *Seminars in radiation oncology*, vol. 14, no. 1, pp. 52–64, 1 2004.

[5] P. Keall, G. Mageras, J. Balter, R. Emery, K. Forster, S. Jiang, J. Kapatoes, D. Low, M. Murphy, B. Murray, C. Ramsey, M. V. Herk, S. Vedam, J. Wong, and E. Yorke, "The management of respiratory motion in radiation oncology report of AAPM task group 76," *Medical physics*, vol. 33, no. 10, p. 3874, 2006.

[6] P. M. Evans, "Anatomical imaging for radiotherapy," *Physics in Medicine and Biology*, vol. 53, no. 12, p. R151, 2008.

[7] E. Rietzel, T. Pan, and G. Chen, "Four-dimensional computed tomography: Image formation and clinical protocol," *Medical physics*, vol. 32, no. 4, pp. 874–889, April 2005 2005.

[8] X. Wang, L. Li, C. Hu, J. Qiu, Z. Xu, and Y. Feng, "A comparative study of three CT and MRI registration algorithms in nasopharyngeal carcinoma," *Journal of Applied Clinical Medical Physics*, vol. 10, no. 2, 2009.

[9] T. Veninga, H. Huisman, R. W. M. Van Der Maazen, and H. Huizenga, "Clinical validation of the normalized mutual information method for registration of CT and MR images in radiotherapy of brain tumors." *Journal of applied clinical medical physics American College of Medical Physics*, vol. 5, no. 3, pp. 66–79, 2004.

[10] R. Jena, S. Price, C. Baker, S. Jefferies, J. Pickard, J. Gillard, and N. Burnet, "Diffusion tensor imaging: Possible implications for radiotherapy treatment planning of patients with high-grade glioma," *Clinical Oncology*, vol. 17, no. 8, pp. 581 – 590, 2005.

[11] V. Gregoire, K. Haustermans, X. Geets, S. Roels, and M. Lonneux, "PET-based treatment planning in radiotherapy: a new standard?" *Journal of Nuclear Medicine*, vol. 48 Suppl 1, no. 1, pp. 68S–77S, 2007.

[12] G. Lucignani, B. Jereczek-Fossa, and R. Orecchia, "The role of molecular imaging in precision radiation therapy for target definition, treatment planning optimisation and quality control." *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 31, no. 8, pp. 1059–1063, 2004.

[13] G. Baroni, C. Garibaldi, M. Scabini, M. Riboldi, G. Catalano, G. Tosi, R. Orecchia, and A. Pedotti, "Dosimetric effects within target and organs at risk of interfractional patient mispositioning in left breast cancer radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 59, no. 3, pp. 861 – 871, 2004.

[14] M. F. Spadea, G. Baroni, M. Riboldi, B. Tagaste, C. Garibaldi, R. Orecchia, and A. Pedotti, "Patient set-up verification by infrared optical localization and body surface sensing in breast radiation therapy," *Radiotherapy and Oncology*, vol. 79, no. 2, pp. 170 – 178, 2006.

[15] B. Tagaste, M. Riboldi, M. F. Spadea, S. Bellante, G. Baroni, R. Cambria, C. Garibaldi, M. Ciocca, G. Catalano, D. Alterio, and R. Orecchia, "Comparison Between Infrared Optical and Stereoscopic X-ray Technologies for Patient Setup in Image Guided Stereotactic Radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, no. 0, pp. –, 2011.

[16] V. Khoo, A. Padhani, S. Tanner, D. Finnigan, M. Leach, and D. Dearnaley, "Comparison of MRI with CT for the radiotherapy planning of prostate cancer: a feasibility study," *British Journal of Radiology*, vol. 72, no. 858, p. 590, 1999.

[17] K. Newbold, M. Partridge, G. Cook, S. Sohaib, E. Charles-Edwards, P. Rhys-Evans, K. Harrington, and C. Nutting, "Advanced imaging applied to radiotherapy planning in head and neck cancer: a clinical review," *British Journal of Radiology*, vol. 79, no. 943, p. 554, 2006.

[18] A. Padhani, K. Krohn, J. Lewis, and M. Alber, "Imaging oxygenation of human tumors." *European Radiology*, vol. 17, no. 4, pp. 525–872, 6/1 2007.

[19] J. Jonsson, M. Karlsson, M. Karlsson, and T. Nyholm, "Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions," *Radiotherapy and oncology*, vol. 5, p. 62, 2010.

[20] A. Beavis, P. Gibbs, R. Dealey, and V. Whitton, "Radiotherapy treatment planning of brain tumours using MRI alone," *British Journal of Radiology*, vol. 71, no. 845, p. 544, 1998.

[21] P. Nucifora, R. Verma, S. Lee, and E. Melhem, "Diffusion-Tensor MR Imaging and Tractography: Exploring Brain Microstructure and Connectivity," *Radiology*, vol. 245, no. 2, p. 367, 2007.

[22] C. Ling and X. Li, "Over the next decade the success of radiation treatment planning will be judged by the immediate biological response of tumor cells rather than by surrogate measures such as dose maximization and uniformity," *Medical physics*, vol. 32, no. 7, p. 2189, 2005.

[23] A. V. Baardwijk, B. Baumert, G. Bosmans, M. van Kroonenburgh, S. Stroobants, V. Gregoire, P. Lambin, and D. D. Ruysscher, "The current status of FDG-PET in tumour volume definition in radiotherapy treatment planning," *Cancer treatment reviews*, vol. 32, no. 4, p. 245, 2006.

[24] K. A. Langmack, "Portal imaging," *British Journal of Radiology*, vol. 74, no. 885, p. 789, 2001.

[25] M. Herman, J. Balter, D. Jaffray, K. McGee, P. Munro, S. Shalev, M. V. Herk, and J. Wong, "Clinical use of electronic portal imaging: report of AAPM radiation therapy committee task group 58," *Medical Physics*, vol. 28, no. 5, pp. 712–737, 5 2001.

[26] X. Artignan, M. Smitsmans, J. Lebesque, D. Jaffray, M. van Herk, and H. Bartelink, "Online ultrasound image guidance for radiotherapy of prostate cancer: impact of image acquisition on prostate displacement," *International Journal of Radiation Oncology*Biology*Physics*, vol. 59, no. 2, p. 595, 2004.

[27] F. V. den Heuvel, T. Powell, E. Seppi, P. Littrupp, M. Khan, Y. Wang, and J. Forman, "Independent verification of ultrasound based image-guided radiation treatment, using electronic portal imaging and implanted gold markers," *Medical physics*, vol. 30, no. 11, p. 2878, 2003.

[28] T. Mackie, "History of tomotherapy," *Physics in Medicine and Biology*, vol. 51, no. 13, p. R427, 2006.

[29] K. Ruchala, G. Olivera, E. Schloesser, and T. Mackie, "Megavoltage CT on a tomotherapy system," *Physics in Medicine and Biology*, vol. 44, no. 10, p. 2597, 1999.

[30] M. Murphy, J. Balter, S. Balter, J. J. BenComo, I. Das, S. Jiang, C. Ma, G. Olivera, R. Rodebaugh, K. Ruchala, H. Shirato, and F. Yin, "The management of imaging dose during image-guided radiotherapy: Report of the AAPM task group 75," *Medical physics*, vol. 34, no. 10, p. 4041, 2007.

[31] L. Feldkamp, L. Davis, and J. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America.A, Optics, image science, and vision*, vol. 1, no. 6, p. 612, 1984.

[32] X. Zhu, S. E. na, J. Daartz, N. Liebsch, J. Ouyang, H. Paganetti, T. Bortfeld, and G. E. Fakhri, "Monitoring proton radiation therapy with in-room PET imaging," *Physics in Medicine and Biology*, vol. 56, no. 13, p. 4041, 2011.

[33] G. Shakirin, H. Braess, F. Fiedler, D. Kunath, K. Laube, K. Parodi, M. Priegnitz, and W. Enghardt, "Implementation and workflow for PET monitoring of therapeutic ion irradiation: a comparison of in-beam, in-room, and off-line techniques," *Physics in Medicine and Biology*, vol. 56, no. 5, p. 1281, 2011.

[34] N. Depauw and J. Seco, "Sensitivity study of proton radiography and comparison with kV and MV X-ray imaging using GEANT4 Monte Carlo simulations," *Physics in Medicine and Biology*, vol. 56, no. 8, p. 2407, 2011.

[35] J. Seco and N. Depauw, "Proof of principle study of the use of a CMOS active pixel sensor for proton radiography," *Medical physics*, vol. 38, no. 2, p. 622, 2011.

[36] U. Schneider and E. Pedroni, "Multiple coulomb scattering and spatial resolution in proton radiography," *Medical physics*, vol. 21, no. 11, p. 1657, 1994.

[37] K. Chao, N. Majhail, C. Huang, J. Simpson, C. Perez, B. Haughey, and G. Spector, "Intensity-modulated radiation therapy reduces late salivary toxicity without compromising tumor control in patients with oropharyngeal carcinoma: a comparison with conventional techniques," *Radiotherapy and oncology*, vol. 61, no. 3, p. 275, 2001.

[38] N. Lee, P. Xia, J. Quivey, K. Sultanem, I.Poon, C. Akazawa, P. Akazawa, V. Weinberg, and K. Fu, "Intensity-modulated radiotherapy in the treatment of nasopharyngeal carcinoma: an update of the UCSF experience," *International Journal of Radiation Oncology*Biology*Physics*, vol. 53, no. 1, pp. 12–22, 5/1 2002.

[39] A. Eisbruch, J. Harris, A. Garden, C. Chao, W. Straube, P. Harari, G. Sanguineti, C. Jones, W. Bosch, and K. Ang, "Multi-institutional trial of accelerated hypofractionated intensity-modulated radiation therapy for early-stage oropharyngeal cancer (RTOG 00-22)," *International Journal of Radiation Oncology*Biology*Physics*, vol. 76, no. 5, p. 1333, 2010.

[40] K. Chao, J. Deasy, J. Markman, J. Haynie, C. Perez, J. Purdy, and D. Low, "A prospective study of salivary function sparing in patients with head-and-neck cancers receiving intensity-modulated or three-dimensional radiation therapy: initial results," *International Journal of Radiation Oncology*Biology*Physics*, vol. 49, no. 4, p. 907, 2001.

[41] A. Eisbruch, R. Haken, H. Kim, L. Marsh, and J. Ship, "Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 45, no. 3, pp. 577–587, 10 1999.

[42] F. Feng, H. Kim, T. Lyden, M. Haxer, M. Feng, F. Worden, D. Chepeha, and A. Eisbruch, "Intensity-modulated radiotherapy of head and neck cancer aiming to reduce dysphagia: Early dose-effect relationships for the swallowing structures," *International Journal of Radiation Oncology*Biology*Physics*, vol. 68, no. 5, pp. 1289–1298, 8/1 2007.

[43] D. Schwartz, K.Hutcheson, D.Barringer, S. Tucker, M. Kies, F. Holsinger, K. Ang, W. Morrison, D. Rosenthal, A. Garden, L. Dong, and J. Lewin, "Candidate dosimetric predictors of long-term swallowing dysfunction after oropharyngeal intensity-modulated radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 78, no. 5, pp. 1356–1365, 12/1 2010.

[44] J. L. Barker, A. Garden, K. Ang, J. O'Daniel, H. Wang, L. Court, W. Morrison, D. Rosenthal, K. Chao, S. Tucker, R. Mohan, and L. Dong, "Quantification of

volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system," *International Journal of Radiation Oncology*Biology*Physics*, vol. 59, no. 4, p. 960, 2004.

[45] H. Suit and A. Walker, "Assessment of the response of tumours to radiation: clinical and experimental studies." *British journal of cancer*, vol. 4, p. 1, 1980.

[46] C. Lee, K. M. Langen, W. Lu, J. Haimerl, E. Schnarr, K. Ruchala, G. Olivera, S. Meeks, P. Kupelian, T. Shellenberger, and R. Manon, "Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration," *International Journal of Radiation Oncology*Biology*Physics*, vol. 71, no. 5, pp. 1563–1571, 8/1 2008.

[47] Z. Wang, C. Yan, Z. Zhang, C. Zhang, H. Hu, J. Kirwan, and W. Mendenhall, "Radiation-induced volume changes in parotid and submandibular glands in patients with head and neck cancer receiving postoperative radiotherapy: A longitudinal study," *The Laryngoscope*, vol. 119, no. 10, p. 1966, 2009.

[48] F. Ricchetti, B. Wu, T. McNutt, J. Wong, A. Forastiere, S. Marur, H. Starmer, and G. Sanguineti, "Volumetric change of selected organs at risk during IMRT for oropharyngeal cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 80, no. 1, pp. 161–168, 5/1 2011.

[49] W. Wang, H. Yang, W. Hu, G. Shan, W. Ding, C. Yu, B. Wang, X. Wang, and Q. Xu, "Clinical study of the necessity of replanning before the 25th fraction during the course of intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma," *International Journal of Radiation Oncology*Biology*Physics*, vol. 77, no. 2, pp. 617–621, 6/1 2010.

[50] P. Ahn, C. Chen, A. Ahn, L. Hong, P. Scripes, J. Shen, C. Lee, E. Miller, S. Kalnicki, and M. Garg, "Adaptive planning in intensity-modulated radiation therapy for head and neck cancers: Single-institution experience and clinical implications," *International Journal of Radiation Oncology*Biology*Physics*, vol. 80, no. 3, pp. 677–685, 7/1 2011.

[51] K. Seet, A. Barghi, S. Yartsev, and J. V. Dyk, "The effects of field-of-view and patient size on CT numbers from cone-beam computed tomography," *Physics in Medicine and Biology*, vol. 54, no. 20, p. 6251, 2009.

[52] Y. Yang, E. Schreibmann, T. Li, C. Wang, and L. Xing, "Evaluation of on-board kV Cone Beam CT (CBCT)-based dose calculation," *Physics in Medicine and Biology*, vol. 52, no. 3, p. 685, 2007.

[53] G. X. Ding, D. Duggan, C. Coffey, M. Deeley, D. Hallahan, A. Cmelak, and A. Malcolm, "A study on adaptive IMRT treatment planning using kV cone-beam CT," *Radiotherapy and oncology*, vol. 85, no. 1, p. 116, 2007.

[54] D. Jaffray, J. Siewerdsen, J. Wong, and A. Martinez, "Flat-panel cone-beam computed tomography for image-guided radiation therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 53, no. 5, p. 1337, 2002.

[55] B. Barney, R. Lee, D. Handrahan, K. Welsh, J. Cook, and W. Sause, "Image-guided radiotherapy (IGRT) for prostate cancer comparing kV imaging of fiducial markers with cone beam computed tomography (CBCT)," *International Journal of Radiation Oncology*Biology*Physics*, vol. 80, no. 1, p. 301, 2011.

[56] W. Shi, J. Li, R. Zlotecki, A. Yeung, H. Newlin, J. Palta, C. Liu, A. Chvetsov, and K. Olivier, "Evaluation of kV cone-beam CT performance for prostate IGRT: a comparison of automatic grey-value alignment to implanted fiducial-marker alignment," *American journal of clinical oncology*, vol. 34, no. 1, p. 16, 2011.

[57] Q. Wu, T. Li, Q. Wu, and F. Yin, "Adaptive radiation therapy: technical components and clinical applications," *The cancer journal*, vol. 17, no. 3, p. 182, 2011.

[58] D. Schwartz and L. Dong, "Adaptive radiation therapy for head and neck cancer - can an old goal evolve into a new standard?" *Journal of Oncology*, vol. 2011, p. 1, 2011.

[59] J. Nijkamp, F. Pos, T. Nuver, R. de Jong, P. Remeijer, J. Sonke, and J. Lebesque, "Adaptive radiotherapy for prostate cancer using kilovoltage cone-beam computed tomography: first clinical results," *International Journal of Radiation Oncology*Biology*Physics*, vol. 70, no. 1, p. 75, 2008.

[60] A. Jensen, S. Nill, P. Huber, R. Bendl, J. Debus, and M. Münter, "A clinical concept for interfractional adaptive radiation therapy in the treatment of head and neck cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 82, no. 2, pp. 590 – 596, 2012.

[61] A. de la Zerda, B. Armbruster, and L. Xing, "Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework," *Physics in Medicine and Biology*, vol. 52, no. 14, p. 4137, 2007.

[62] M. Birkner, D. Yan, M. Alber, J. Liang, and F. Nüsslin, "Adapting inverse planning to patient and organ geometrical variation: algorithm and implementation," *Medical physics*, vol. 30, no. 10, p. 2822, 2003.

[63] C. Wu, R. Jeraj, W. Lu, and T. Mackie, "Fast treatment plan modification with an over-relaxed Cimmino algorithm," *Medical physics*, vol. 31, no. 2, p. 191, 2004.

[64] T. Nuver, M. Hoogeman, P. Remeijer, M. van Herk, and J. Lebesque, "An adaptive off-line procedure for radiotherapy of prostate cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 67, no. 5, pp. 1559–1567, 4/1 2007.

[65] D. Yan, D. Lockman, D. Brabbins, L. Tyburski, and A. Martinez, "An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 48, no. 1, pp. 289–302, 8/1 2000.

[66] Q. Wu, J. Liang, and D. Yan, "Application of dose compensation in image-guided radiotherapy of prostate cancer," *Physics in Medicine and Biology*, vol. 51, no. 6, p. 1405, 2006.

[67] C. Wu, R. Jeraj, G. Olivera, and T. Mackie, "Re-optimization in adaptive radiotherapy," *Physics in Medicine and Biology*, vol. 47, no. 17, p. 3181, 2002.

[68] C. Woodford, S. Yartsev, A. Dar, G. Bauman, and J. V. Dyk, "Adaptive radiotherapy planning on decreasing gross tumor volumes as seen on megavoltage computed tomography images," *International Journal of Radiation Oncology*Biology*Physics*, vol. 69, no. 4, pp. 1316–1322, 11/15 2007.

[69] L. Court, L. Dong, A. Lee, R. Cheung, M. Bonnen, J. O'Daniel, H. Wang, R. Mohan, and D. Kuban, "An automatic CT-guided adaptive radiation therapy technique by online modification of multileaf collimator leaf positions for prostate cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 62, no. 1, pp. 154–163, 5/1 2005.

[70] L. Court, R. Tishler, J. Petit, R. Cormack, and L. Chin, "Automatic online adaptive radiation therapy techniques for targets with significant shape change: a feasibility study," *Physics in Medicine and Biology*, vol. 51, no. 10, p. 2493, 2006.

[71] R. Mohan, X. Zhang, H. Wang, Y. Kang, X. Wang, H. Liu, K. Ang, D. Kuban, and L. Dong, "Use of deformed intensity distributions for on-line modification of image-guided IMRT to account for interfractional anatomic changes," *International Journal of Radiation Oncology*Biology*Physics*, vol. 61, no. 4, p. 1258, 2005.

[72] W. Fu, Y. Yang, N. Yue, D. Heron, and M. Huq, "A cone beam CT-guided online plan modification technique to correct interfractional anatomic changes for prostate cancer IMRT treatment," *Physics in Medicine and Biology*, vol. 54, no. 6, p. 1691, 2009.

[73] Q. Wu, D. Thongphiew, Z. Wang, B. Mathayomchan, V. Chankong, S. Yoo, W. Lee, and F. Yin, "On-line re-optimization of prostate IMRT plans for adaptive radiation therapy," *Physics in Medicine and Biology*, vol. 53, no. 3, p. 673, 2008.

[74] Y. Feng, C. Castro-Pareja, R. Shekhar, and C. Yu, "Direct aperture deformation: An interfraction image guidance strategy," *Medical physics*, vol. 33, no. 12, p. 4490, 2006.

[75] E. Ahunbay, C. Peng, S. Holmes, A. Godley, C. Lawton, and X. Li, "Online adaptive replanning method for prostate radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 77, no. 5, pp. 1561–1572, 8/1 2010.

[76] X. Geets, J. Daisne, S. Arcangeli, E. Coche, M. D. Poel, T. Duprez, G. Nardella, and V. Gregoire, "Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI," *Radiotherapy and oncology*, vol. 77, no. 1, p. 25, 2005.

[77] N. Kovacevic, N. Lobaugh, M. Bronskill, B. Levine, A. Feinstein, and S. Black, "A robust method for extraction and automatic segmentation of brain images," *NeuroImage*, vol. 17, no. 3, p. 1087, 2002.

[78] J. Sethian, "Level set methods and fast marching methods," *Journal of computing and information technology*, vol. 11, no. 1, p. 1, 2003.

[79] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation," *Optical Engineering*, vol. 34, p. 433, 1992.

[80] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky, "Mutual information in coupled multi-shape model for medical image segmentation," *Medical image analysis*, vol. 8, no. 4, pp. 429–445, 12 2004.

[81] J. Yang, L. Staib, and J. Duncan, "Neighbor-constrained segmentation with 3D deformable models," in *Proceeding of the International Conference on Information Processing in Medical Imaging (IPMI,2003)*, vol. 2732, 2003, p. 198.

[82] T. Rohlfing, R. Brandt, R. Menzel, D. Russakoff, and C. J. Maurer, *Quo vadis, atlas-based segmentation?* Kluwer Academic / Plenum Publishers, New York, NY, 2005.

[83] T. Rohlfing, R. Brandt, R. Menzel, and C. J. Maurer, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, p. 1428, 2004.

[84] P. Aljabar, R. Heckemann, A. Hammers, H. J.V., and R. D., "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, p. 726, 2009.

[85] R. Kikinis, M. Shenton, D. Iosifescu, R. McCarley, P. Saiviroonporn, H. Hokama, A. Robatino, D. Metcalf, C. Wible, C. Portas, R. Donnino, and F. Jolesz, "A digital brain atlas for surgical planning, model-driven segmentation, and teaching," *IEEE Transactions on Visualization and Computer Graphics*, vol. 2, no. 3, p. 232, 1996.

[86] D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes, and A. Evanss, "Design and construction of a realistic digital brain phantom," *Medical Imaging, IEEE Transactions on*, vol. 17, no. 3, p. 463, 1998.

[87] P. Bondiau, G. Malandain, S. Chanalet, P. Marcy, J. Habrand, F. Fauchon, P. Paquis, A. Courdi, O. Commowick, I. Rutten, and N. Ayache, "Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context," *International Journal of Radiation Oncology*Biology*Physics*, vol. 61, no. 1, p. 289, 2005.

[88] "Online: http://www.bic.mni.mcgill.ca/brainweb/."

[89] A. Rao, G. I. Sanchez-Ortiz, R. Chandrashekara, M. Lorenzo-Valdes, R. Mohiaddin, and D. Rueckert, "Construction of a cardiac motion atlas from MR using non-rigid registration," in *Proceedings of the 2nd international conference*

on Functional imaging and modeling of the heart (FIMH'03), vol. 2674, 2003, p. 141.

[90] B. Li, G. Christensen, E. Hoffman, G. McLennan, and J. Reinhardt, "Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images," *Academic Radiology*, vol. 10, no. 3, p. 255, 2003.

[91] H. Jia, G. Wu, Q. Wang, and D. Shen, "ABSORB: Atlas Building by Self-Organized Registration and Bundling," *NeuroImage*, vol. 51, no. 3, p. 2785, 2010.

[92] P. Thompson and A. Toga, "Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations," *Medical image analysis*, vol. 1, no. 4, p. 271, 1997.

[93] M. Depa, M. Sabuncu, G. Holmvang, R. Nezafat, E. Schmidt, and P. Golland., "Robust atlas-based segmentation of highly variable anatomy: left atrium segmentation," in *Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention (MICCAI 2010)*, vol. 6364, 2010, p. 85.

[94] X. Han, M. Hoogeman, P. Levendag, L. Hibbard, D. Teguh, P. Voet, A. Cowen, and T. Wolf, "Atlas-based auto-segmentation of head and neck CT images," in *11th Intl. Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI 2008)*, vol. 5242, 2008, p. 434.

[95] H. Park, P. Bland, and C. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 4, p. 483, 2003.

[96] M. Valdes, G. Sanchez-Ortiz, A. Elkington, R. Mohiaddin, and D. Rueckert, "Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm," *Medical image analysis*, vol. 8, no. 3, p. 255, 2004.

[97] C. Svarer, K. Madsen, S. Hasselbalch, L. Pinborg, S. Haugbol, V. Frokjaer, S. Holm, O. Paulson, and G. Knudsen, "MR-based automatic delineation of volumes of interest in human brain PET images using probability maps," *NeuroImage*, vol. 24, no. 4, p. 969, 2005.

[98] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in CT scans," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 7, p. 1000, 2009.

[99] L. Ramus, O. Commowick, and G. Malandain, "Construction of patient specific atlases from locally most similar anatomical pieces," in *Proceedings of the 13th International conference on Medical image computing and computer-assisted intervention (MICCAI 2010)*, vol. 6363, 2010, p. 155.

[100] E. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. Viergever, J. Pluim, and B. van Ginneken, "Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus," *Medical image analysis*, vol. 14, no. 1, p. 39, 2010.

[101] H. Jia, P. Y. PT, and D. Shen, "Iterative multi-atlas-based multi-image segmentation with tree-based registration," *NeuroImage*, 2011.

[102] R. Wolz, P. A. J. Hajnal, A. Hammers, D. Rueckert, and A. D. N. Initiative, "LEAP: Learning embeddings for atlas propagation," *NeuroImage*, vol. 49, no. 2, p. 1316, 2010.

[103] A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen, and P. Suetens, "SPARC: Unified framework for automatic segmentation, probabilistic atlas construction, registration and clustering of brain MR images," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, 2010, pp. 856–859.

[104] K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert, "Consistent groupwise non-rigid registration for atlas construction," in *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, 2004, pp. 908–911 Vol. 1, iD: 1.

[105] R. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. Hajnal, A. Hammers, and A. D. N. Initiative, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *NeuroImage*, vol. 51, no. 1, p. 221, 2010.

[106] Q. Wu, Y. Chi, P. Chen, D. Krauss, D. Yan, and A. Martinez, "Adaptive replanning strategies accounting for shrinkage in head and neck IMRT," *International Journal of Radiation Oncology*Biology*Physics*, vol. 75, no. 3, p. 924, 2009.

[107] T. Zhang, Y. Chi, E. Meldolesi, and D. Yan, "Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 68, no. 2, p. 522, 2007.

[108] S. Nithiananthan, K. Brock, M. Daly, H. Chan, J. Irish, and J. Siewerdsen, "Demons deformable registration for CBCT-guided procedures in the head and neck: Convergence and accuracy," *Medical physics*, vol. 36, no. 10, p. 4755, 2009.

[109] E. Faggiano, C. Fiorino, E. Scalco, S. Broggi, M. Cattaneo, E. Maggiulli, I. Dell'Oca, N. D. Muzio, R. Calandrino, and G. Rizzo, "An automatic contour propagation method to follow parotid gland deformation during head-and-neck cancer tomotherapy," *Physics in Medicine and Biology*, vol. 56, no. 3, p. 775, 2011.

[110] J. Hou, M. Guerrero, W. Chen, and W. D'Souza, "Deformable planning CT to cone-beam CT image registration in head-and-neck cancer," *Medical physics*, vol. 38, no. 4, p. 2088, 2011.

[111] B. Fischer and J. Modersitzki, "Ill-posed medicine - an introduction to image registration," *Inverse Problems*, vol. 24, no. 3, p. 034008, 2008.

[112] B. Zitova, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[113] Z. Yaniv, *Rigid Registration*, ser. Image-Guided Interventions Technology and Applications. Springer-Verlag, 2008, ch. 6.

[114] T. Liu, D. Shen, and C. Davatzikos, "Deformable registration of cortical structures via hybrid volumetric and surface warping." *NeuroImage*, vol. 22, no. 4, pp. 1790–801, 2004.

[115] J. B. Maintz and M. A. Viergever, "A survey of medical image registration." *Medical image analysis*, vol. 2, no. 1, pp. 1–36, 1998.

[116] P. J. Besl and H. D. McKay, "A method for registration of 3-d shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, p. 239, feb 1992.

[117] C. A. Pelizzari, G. T. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen, "Accurate three-dimensional registration of CT, PET, and/or MR images of the brain," *Journal of computer assisted tomography*, vol. 13, no. 1, pp. 20–26, Jan-Feb 1989.

[118] D. Levin, C. A. Pelizzari, G. T. Chen, C. T. Chen, and M. D. Cooper, "Retrospective geometric correlation of MR, CT, and PET images." *Radiology*, vol. 169, no. 3, pp. 817–823, 1988.

[119] W. M. W. III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical image analysis*, vol. 1, no. 1, p. 35, 1996.

[120] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Non-rigid multimodality image registration," in *Proceedings of SPIE*, M. Sonka and K. M. Hanson, Eds., vol. 4322, 2001, pp. 1609–1620.

[121] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Interpolation artefacts in mutual information-based image registration," *Computer Vision and Image Understanding*, vol. 77, no. 2, p. 211, 2000.

[122] ——, "Mutual-information-based registration of medical images: a survey," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 8, p. 986, aug. 2003.

[123] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 6, pp. 567–585, 1989.

[124] J. A. Shackleford, N. Kandasamy, and G. C. Sharp, "On developing b-spline registration algorithms for multi-core processors," *Physics in Medicine and Biology*, vol. 55, no. 21, p. 6329, 2010.

[125] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *Medical Imaging, IEEE Transactions on*, vol. 18, no. 8, p. 712, aug. 1999.

[126] Z. Wu, E. Rietzel, V. Boldea, D. Sarrut, and G. C. Sharp, "Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations," *Medical physics*, vol. 35, no. 2, pp. 775–781, 2008.

[127] G. C. Sharp, M. Peroni, R. Li, J. Shackleford, and N. Kandasamy, "Evaluation of plastimatch B-Spline registration on the EMPIRE10 dataset," in *Workshop Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention*, ser. Medical Image Analysis for the Clinic: A Grand Challenge, 2010, pp. 99–108.

[128] G. C. Sharp, N. Kandasamy, H. Singh, and M. Folkert, "GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration," *Physics in Medicine and Biology*, vol. 52, no. 19, p. 5771, 2007.

[129] G. C. Sharp, R. Li, J. Wolfgang, G. T. Y. Chen, M. Peroni, M. F. Spadea, S. Mori, J. Zhang, J. Shackleford, and N. Kandasamy, "Plastimatch - An Open Source Software Suite for Radiotherapy Image Processing," in *Proceedings of the XVI International Conference on the Use of Computers in Radiation Therapy*, 2009.

[130] K. K. Brock and Deformable Registration Accuracy Consortium, "Results of a multi-institution deformable registration accuracy study (midras)," *International Journal of Radiation Oncology*Biology*Physics*, vol. 76, no. 2, p. 583, 2010.

[131] D. Paquin, D. Levy, E. Schreibmann, and L. Xing, "Multiscale image registration." *Mathematical biosciences and engineering MBE*, vol. 3, no. 2, pp. 389–418, 2006.

[132] H. Lester, "A survey of hierarchical non-linear medical image registration," *Pattern Recognition*, vol. 32, no. 1, pp. 129–149, 1999.

[133] J. Thirion, "Image matching as a diffusion process: an analogy with maxwell's demons," *Medical image analysis*, vol. 2, no. 3, p. 243, 1998.

[134] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: efficient non-parametric image registration." *NeuroImage*, vol. 45, no. 1 Suppl, pp. S61–S72, 2009.

[135] F. Dru, P. Fillard, and T. Vercauteren, "An itk implementation of the symmetric log-domain diffeomorphic demons algorithm," *The Insight Journal*, 09 2010.

[136] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*, 2nd ed. Kitware, Inc. ISBN 1-930934-15-7, 2005.

[137] K. K. Brock, M. B. Sharpe, L. A. Dawson, S. M. Kim, and D. A. Jaffray, "Accuracy of finite element model-based multi-organ deformable image registration," *Medical physics*, vol. 32, no. 6, pp. 1647–1659, 2005.

[138] M. Chao, Y. Xie, and L. Xing, "Auto-propagation of contours for adaptive prostate radiation therapy," *Physics in Medicine and Biology*, vol. 53, no. 17, p. 4533, 2008.

[139] T. Liu, D. Shen, and C. s. Davatzikos, "Deformable registration of cortical structures via hybrid volumetric and surface warping." *NeuroImage*, vol. 22, no. 4, pp. 1790–1801, 2004.

[140] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 1, p. 120, jan. 2003.

[141] T.-N. Nguyen, J. L. Moseley, L. A. Dawson, D. A. Jaffray, and K. K. Brock, "Adapting population liver motion models for individualized online image-guided therapy," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, aug. 2008, p. 3945.

[142] E. Schreibmann and L. Xing, "Image registration with auto-mapped control volumes," *Medical physics*, vol. 33, no. 4, pp. 1165–1179, 2006.

[143] R. W. van der Put, E. M. Kerkhof, B. W. Raaymakers, I. M. Jürgenliemk-Schulz, and J. J. W. Lagendijk, "Contour propagation in MRI-guided radiotherapy treatment of cervical cancer: the accuracy of rigid, non-rigid and semi-automatic registrations," *Physics in Medicine and Biology*, vol. 54, no. 23, p. 7135, 2009.

[144] M. Beauchemin and K. P. B. Thomson, "The evaluation of segmentation results and the overlapping area matrix," *International Journal of Remote Sensing*, vol. 18, no. 18, pp. 3895–3899, 1997.

[145] S. Kabus, T. Klinder, K. Murphy, B. van Ginneken, C. Lorenz, and J. Pluim, *Evaluation of 4D-CT Lung Registration*, ser. Proceedings of the 12th international conference on Medical image computing and computer-assisted intervention (MICCAI 2009). Springer Berlin / Heidelberg, 2009, vol. 5761, pp. 747–754.

[146] K. Murphy, B. van Ginneken, S. Klein, M. Staring, B. J. de Hoop, M. A. Viergever, and J. P. W. Pluim, "Semi-automatic construction of reference standards for evaluation of image registration," *Medical image analysis*, vol. 15, no. 1, p. 71, 2011.

[147] K. Murphy, B. van Ginneken, J. Reinhardt, S. Kabus, K. Ding, X. Deng, K. Cao, K. Du, G. Christensen, V. Garcia, T. Vercauteren, N. Ayache, O. Commowick, G. Malandain, B. Glocker, N. Paragios, N. Navab, V. Gorbunova, J. Sporring, M. de Bruijne, X. Han, M. Heinrich, J. Schnabel, M. Jenkinson, C. Lorenz, M. Modat, J. McClelland, S. Ourselin, S. Muenzing, M. Viergever, D. De Nigris, D. Collins, T. Arbel, M. Peroni, R. Li, G. Sharp, A. Schmidt-Richberg, J. Ehrhardt, R. Werner, D. Smeets, D. Loeckx, G. Song, N. Tustison, B. Avants, J. Gee, M. Staring, S. Klein, B. Stoel, M. Urschler, M. Werlberger, J. Vandemeulebroucke, S. Rit, D. Sarrut, and J. Pluim, "Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 11, pp. 1901 –1920, nov. 2011.

[148] L. Kitchen and A. Rosenfeld, "Gray-level corner detection," *Pattern Recognition Letters*, vol. 1, no. 2, p. 95, 1982.

[149] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," ser. ISPRS Intercommission Workshop Interlaken, 1987, pp. 281–305.

[150] C. Harris and M. Stephens, *A combined corner and edge detector*, ser. Alvey vision conference. Manchester, UK, 1988, vol. 15, ch. Manchester, pp. 147–151.

[151] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, p. 1150.

[152] ——, "Distinctive image features from scale-invariant keypoints," 2003.

[153] W. Cheung and G. Hamarneh, "N-SIFT: N-dimensional scale invariant feature transform for matching medical images," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, april 2007, p. 720.

[154] G. Christensen, J. Kuhl, T. Grabowski, I. Pirwani, M. Vannier, J. Allen, and H. Damasio, "Introduction to the non-rigid image registration evaluation project (nirep)," in *Proceedings of SPIE*, 2006, pp. 128–135.

[155] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *Medical Imaging, IEEE Transactions on*, vol. 25, no. 11, p. 1451, nov. 2006.

[156] G. Gerig, M. Jomier, and M. Chakos, "Valmet: A new validation tool for assessing and improving 3D object segmentation," *Computer*, vol. 2208, no. 2208, pp. 516–523, 2001.

[157] P. Rogelj, S. Kovacic, and J. C. Gee, "Validation of a nonrigid registration algorithm for multimodal data," in *Proceedings of SPIE*, M. Sonka and J. M. Fitzpatrick, Eds., vol. 4684, 2002, pp. 299–307.

[158] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 2, pp. 153 –163, feb. 2012.

[159] W. Crum, T. Hartkens, and D. Hill, "Non-rigid image registration: theory and practice," *British Journal of Radiology*, vol. 77, no. 2, pp. S140–S153, 2004.

[160] P. Castadot, J. A. Lee, A. Parraga, X. Geets, B. Macq, and V. Gregoire, "Comparison of 12 deformable registration strategies in adaptive radiation therapy for the treatment of head and neck tumors," *Radiotherapy and oncology*, vol. 89, no. 1, pp. 1–12, Oct 2008.

[161] A. Klein, J. Andersson, B. Ardekani, J. Ashburner, B. Avants, M. Chiang, G. Christensen, D. Collins, J. Gee, P. Hellier, J. Song, M. Jenkinson, C. Lepage,

D. Rueckert, P. Thompson, T. Vercauteren, R. Woods, J. Mann, and R. Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, p. 786, 2009.

[162] D. Yang, H. Li, D. Low, J. Deasy, and I. El Naqa, "A fast inverse consistent deformable image registration method based on symmetric optical flow computation," *Physics in Medicine and Biology*, vol. 53, no. 21, p. 6143, 2008.

[163] M. Holden, "A review of geometric transformations for nonrigid body registration," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 1, pp. 111 –128, jan. 2008.

[164] X. Gu, H. Pan, Y. Liang, R. Castillo, D. Yang, D. Choi, E. Castillo, A. Majumdar, T. Guerrero, and S. Jiang, "Implementation and evaluation of various demons deformable image registration algorithms on a gpu," *Physics in Medicine and Biology*, vol. 55, no. 1, p. 207, 2010.

[165] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.

[166] M. A. Deeley, A. Chen, R. Datteri, J. H. Noble, A. J. Cmelak, E. F. Donnelly, A. W. Malcolm, L. Moretti, J. Jaboin, K. Niermann, E. S. Yang, D. S. Yu, F. Yei, T. Koyama, G. X. Ding, and B. M. Dawant, "Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study," *Physics in Medicine and Biology*, vol. 56, no. 14, p. 4557, 2011.

[167] R. Sims, A. Isambert, V. Gregoire, F. Bidault, L. Fresco, J. Sage, J. Mills, J. Bourhis, D. Lefkopoulos, O. Commowick, M. Benkebil, and G. Malandain, "A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck," *Radiotherapy and Oncology*, vol. 93, no. 3, pp. 474 – 478, 2009.

[168] A. Schmidt-Richtberg, J. Ehrhardt, R. Werner, and H. Handels, *Diffeomorphic Diffusion Registration of Lung CT Images*, ser. Workshop Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention, 2010.

[169] S. Nithiananthan, K. K. Brock, M. J. Daly, H. Chan, J. C. Irish, and J. H. Siewerdsen, "Demons deformable registration for CBCT-guided procedures in the head and neck: Convergence and accuracy," *Medical Physics*, vol. 36, no. 10, pp. 4755–4764, 2009.

[170] F. J. Anscombe, "Graphs in statistical analysis," *The American Statistician*, vol. 27, no. 1, pp. pp. 17–21, 1973.

[171] C. Caldwell, K. Mah, Y. Ung, C. Danjoux, J. Balogh, S. Ganguli, and L. Ehrlich, "Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: the impact of 18FDG-hybrid PET fusion," *International Journal of Radiation Oncology*Biology*Physics*, vol. 51, no. 4, pp. 923 – 931, 2001.

[172] M. Eliasziw, S. L. Young, M. G. Woodbury, and K. Fryday-Field, "Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example." *Physical Therapy*, vol. 74, no. 8, pp. 777–788, 1994.

[173] X. Geets, J. Daisne, S. Arcangeli, E. Coche, M. D. Poel, T. Duprez, G. Nardella, and V. Gregoire, "Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: Comparison between CT-scan and MRI," *Radiotherapy and Oncology*, vol. 77, no. 1, pp. 25 – 31, 2005.

[174] M. Murphy, Z. Wei, M. Fatyga, J. Williamson, M. Anscher, T. Wallace, and E. Weiss, "How does CT image noise affect 3D deformable image registration for image-guided radiotherapy planning?" *Medical Physics*, vol. 35, no. 3, pp. 1145–1153, 2008.

[175] P. Remeijer, C. Rasch, J. Lebesque, and M. van Herk, "A general methodology for three-dimensional analysis of variation in target volume delineation," *Medical Physics*, vol. 26, no. 6, pp. 931–940, 1999.

[176] N. Aspert, D. Santa-Cruz, and T. Ebrahimi, "MESH: measuring errors between surfaces using the hausdorff distance," in *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, vol. 1, 2002, pp. 705 – 708 vol.1.

[177] C. Bert, K. Metheany, K. Doppke, A. Taghian, S. N. Powell, and G. Chen, "Clinical experience with a 3D surface patient setup system for alignment of partial-breast irradiation patients," *International Journal of Radiation Oncology*Biology*Physics*, vol. 64, no. 4, pp. 1265 – 1274, 2006.

[178] C. Bert, K. G. Metheany, K. Doppke, and G. T. Y. Chen, "A phantom evaluation of a stereo-vision surface imaging system for radiotherapy patient setup," *Medical Physics*, vol. 32, no. 9, pp. 2753–2762, 2005.

[179] K. Jankauskas and A. Noreika, "Measuring error between similar surfaces." *Information Technology And Control, Kaunas, Technologija*, vol. 37, no. 8, pp. 198–204, 2008.

[180] B. Rodriguez-Vila, F. Gaya, F. Garcia-Vicente, and E. J. Gomez, "Three-dimensional quantitative evaluation method of nonrigid registration algorithms for adaptive radiotherapy," *Medical Physics*, vol. 37, no. 3, pp. 1137–1145, 2010.

[181] V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *Medical Imaging, IEEE Transactions on*, vol. 16, no. 5, pp. 642 –652, oct. 1997.

[182] P. Danielsson, "Euclidean distance mapping," *Computer Graphics and Image Processing*, vol. 14, pp. 227–248, 1980.

[183] M. A. Föstner and E. Gülch, "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features," in *ISPRS Intercommission Workshop*, 1987.

[184] W. Förstner, "A feature based correspondence algorithm for image matching," *Archives of Photogrammetry and Remote Sensing*, vol. 26, no. 3, pp. 150–166, 1986.

[185] R.Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. Garg, and T. Guerrero, "A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets," *Physics in Medicine and Biology*, vol. 54, no. 7, p. 1849, 2009.

[186] K. Rohr, H. S. Stiehl, S. Frantz, and T. Hartkens, "Performance characterization of landmark operators," in *Proceedings of the Theoretical Foundations of Computer Vision, TFCV on Performance Characterization in Computer Vision*. Deventer, The Netherlands, The Netherlands: Kluwer, B.V., 2000, pp. 285–297.

[187] R. Pielot, M. Scholz, K. Obermayer, E. Gundelfinger, and A. Hess, "3D edge detection to define landmarks for point-based warping in brain imaging," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, oct 2001, pp. 343 –346 vol.2.

[188] K. G. Derpanis. (2004) The Harris Corner Detector.

[189] J. L. Crowley and A. C. Parker, "A representation for shape based on peaks and ridges in the difference of low-pass transform," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-6, no. 2, pp. 156 –170, march 1984.

[190] Y. Xie, M. Chao, P. Lee, and L. Xing, "Feature-based rectal contour propagation from planning CT to cone beam CT," *Medical Physics*, vol. 35, no. 10, pp. 4450–4459, 2008.

[191] M. Chao, Y. Xie, E. G. Moros, Q.-T. Le, and L. Xing, "Image-based modeling of tumor shrinkage in head and neck radiation therapy," *Medical Physics*, vol. 37, no. 5, pp. 2351–2358, 2010.

[192] S. Allaire, J. Kim, S. Breen, D. Jaffray, and V. Pekar, "Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis," in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, june 2008, pp. 1 –8.

[193] K. Mikolajczyk, "Detection of local features invariant to affines transformations," Ph.D. dissertation, INPG, Grenoble, juillet 2002.

[194] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proceedings of the 7th European Conference on Computer Vision-Part I*, ser. ECCV '02. London, UK, UK: Springer-Verlag, 2002, pp. 128–142.

[195] R. Lemuz-Lopez and M. Arias-Estrada, "Iterative closest SIFT formulation for robust feature matching." in *ISVC'06 Proceedings of the Second international conference on Advances in Visual Computing - Volume Part II*, ser. Lecture Notes in Computer Science, vol. 4292. Springer, 2006, pp. 502–513.

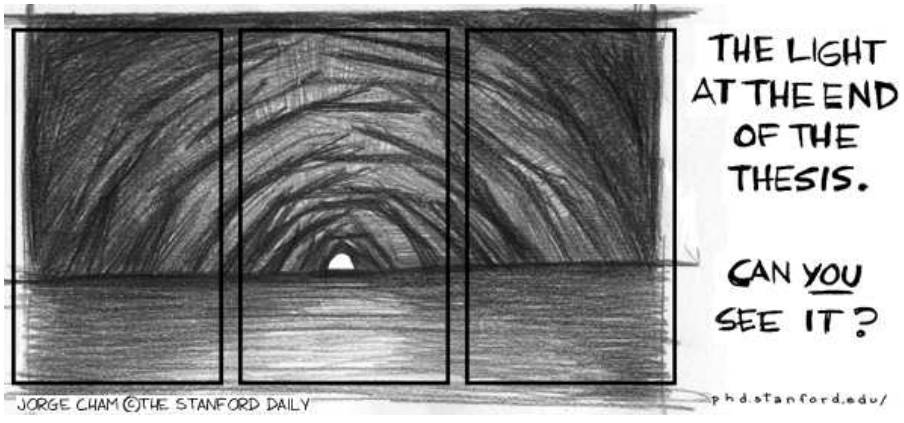
[196] M. Peroni, P. Golland, G. Sharp, and G. Baroni, "Ranking of stopping criteria for log domain diffeomorphic demons application in clinical radiation therapy," in *EMBC 2011. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011*, sept. 2011.

[197] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 10, pp. 1714 –1729, oct. 2010.

[198] S. Pieper, B. Lorensen, W. Schroeder, and R. Kikinis, "The NA-MIC Kit: ITK, VTK, pipelines, grids and 3DSlicer as an open platform for the medical image computing community," in *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*, april 2006, pp. 698 –701.

[199] R. Shekhar, P. Lei, C. Castro-Pareja, W. Plishker, and W. D'Souza, "Automatic segmentation of phase-correlated CT scans through nonrigid image registration using geometrically regularized free-form deformation," *Medical Physics*, vol. 34, no. 7, pp. 3054–3066, 2007.


...HERE'S A COPY OF THE STATEMENT OF PURPOSE YOU WROTE FOR YOUR GRAD SCHOOL APPLICATION...
HMM...
...I THINK YOU PICKED A VERY EXCITING TOPIC TO DO YOUR THESIS ON BECAUSE ...
HMPH... SNICKER...
HAHA HAHA HAHA HA HA HA HA HA HAHAHA HA HA HA HA HA HA HAHA!!
...AHEM...HMPH... SORRY, IT'S JUST SOOO FUNNY... NO, REALLY, YOU DIDN'T ACTUALLY BELIEVE I COULD DO ANY OF THIS, DID YOU...?
JORGE CHAM ©THE STANFORD DAILY


THE LIGHT AT THE END OF THE THESIS.
CAN YOU SEE IT?
JORGE CHAM ©THE STANFORD DAILY
phd.stanford.edu/