Propagation of registration uncertainty during multi-fraction cervical cancer brachytherapy

A Amir-Khalili¹, G Hamarneh², R Zakariae³,⁴, I Spadinger⁴, R Abugharbieh¹

¹ Biomedical Signal and Image Computing Lab, University of British Columbia, Vancouver, BC, Canada
² Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, Canada
³ Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada
⁴ BC Cancer Agency Vancouver Center, Vancouver, BC, Canada

E-mail: alborza@ece.ubc.ca

August 2017

Abstract.

Multi-fraction cervical cancer brachytherapy is a form of image-guided radiotherapy that heavily relies on 3D imaging during treatment planning, delivery, and quality control. In this context, deformable image registration can increase the accuracy of dosimetric evaluations, provided that one can account for the uncertainties associated with the registration process. To enable such capability, we propose a mathematical framework that first estimates the registration uncertainty and subsequently propagates the effects of the computed uncertainties from the registration stage through to the visualizations, organ segmentations, and dosimetric evaluations. To ensure the practicality of our proposed framework in real world image-guided radiotherapy contexts, we implemented our technique via a computationally efficient and generalizable algorithm that is compatible with existing deformable image registration software. In our clinical context of fractionated cervical cancer brachytherapy, we perform a retrospective analysis on 37 patients and present evidence that our proposed methodology for computing and propagating registration uncertainties may be beneficial during therapy planning and quality control. Specifically, we quantify and visualize the influence of registration uncertainty on dosimetric analysis during the computation of the total accumulated radiation dose on the bladder wall. We further show how registration uncertainty may be leveraged into enhanced visualizations that depict the quality of the registration and highlight potential deviations from the treatment plan prior to the delivery of radiation treatment. Finally, we show that we can improve the transfer of delineated volumetric organ segmentation labels from one fraction to the next by encoding the computed registration uncertainties into the segmentation labels.

Keywords: deformable image registration, uncertainty, multi-fraction brachytherapy

1. Introduction

Deformable image registration (DIR) techniques play an integral part in the field of medical image processing. Medical image analysts rely on DIR to transform different image volumes into a single coordinate system, allowing them to assess differences across pathologies, subjects, groups, time, and imaging modalities. Unlike simpler rigid and affine transformation models, DIR is fraught with challenges stemming from the complexity of parameter optimization, choice of a similarity metric, and evaluation of accuracy and precision of resulting transformations (Jannin et al. 2006, Tang & Hamarneh 2013a, Sotiras et al. 2013). Such challenges are exacerbated as transformation models become more complex to accommodate for more realistic tissue deformations.

1.1. Motivation

Despite the challenges associated with DIR, deformable models generally outperform simpler registration models since most organs inside the human body undergo complex elastic deformations. Inevitably, however, errors that occur during DIR would propagate through to all subsequent analyses performed post-registration. Examples include errors in augmented reality visualization during computer-assisted surgery (Holloway 1997), biased estimation of head movement in fMRI time-series analysis (Friston et al. 1996), and geometric uncertainties in localizing organ shape or motion during image-guided radiotherapy (Dawson & Sharpe 2006). Furthermore, established methods for estimating the overall accuracy of registration processes from fiducial or target registration errors cannot be simply generalized from rigid to deformable transformation models (Yetik & Nehorai 2006). Instead, in order to calculate the accuracy of DIR, the ground truth (GT) transformation, a dense 3D-3D mapping, is required.

Unfortunately, GT data for DIR are either scarce, non-existent, or impossible to obtain for most clinical applications. Obtaining GT data is particularly difficult in clinical applications where volumetric medical images from complementary imaging modalities are fused for the purpose of image-guided interventions or therapies. Such applications widely vary from multimodal imaging for radiation treatment (Dawson & Sharpe 2006) to computer assisted surgery (Markelj et al. 2012). The lack of a GT thus restricts the validation process to testing on synthetically generated data or simplified approximations of DIR errors (Tang & Hamarneh 2013a, Brock et al. 2017).

Additionally, DIR errors also affect the decisions made by the end-users of DIR, which include both image analysts and physicians. An erroneous DIR may result in catastrophic outcomes for a patient during image guided therapies where DIR is often used for the purpose of guidance (Fedorov et al. 2014). Without a registration quality measure, clinicians are rendered blind to performance of the registration stage and it is therefore crucial to quantify the quality of DIR.

Even though errors cannot be equated to uncertainty, the two have been found to be strongly correlated (Lotfi et al. 2013, Hub & Karger 2013) and registration uncertainty (RU) thus remains as a valuable measure of quality for DIR. The importance
of understanding and communicating the uncertainty associated with DIR software is a timely issue. Recently, the Therapy Physics Committee of the American Association of Physicists in Medicine (AAPM) commissioned a task group to review the current approaches and solutions for image registration in radiotherapy. Among their clinical recommendations, Brock et al (2017) advocated for a better understanding of the basic components of the registration algorithm; end-to-end tests of imaging, registration, and treatment systems using a physical phantom; and comprehensive commissioning of image registration using digital phantom data. Although we support these recommendations, we emphasize that phantom-based studies shed limited insights on the validity of DIR software, primarily due to the fact that phantom-based analyses are often an oversimplification of real world situations where noise, distortion, and complex anatomical variations typically occur.

Given the challenges described above, our work is inspired by the clear need for a RU estimation and propagation solution that is supported by rigorous mathematical formalism while being implementable as an end-to-end framework in real world applications. In this paper, we address this crucial need with a mathematical framework that is capable of estimating DIR quality using RU and propagating the effects of the RU from the registration stage through to the visualizations, delineations, and dosimetric analysis results that are provided to the end-user. Our framework is computationally efficient and is designed to be deployed on top of existing DIR software such that it can be easily extended to applications other than those specifically considered in this work.

1.2. Multi-fraction cervical cancer brachytherapy

Our application choice for demonstrating the effects of RU on DIR and therapy delivery is image-guided multi-fraction cervical cancer brachytherapy (MFCCB). MFCCB is a form of radiotherapy that relies heavily on 3D imaging for treatment planning, delivery, and quality control. A patient undergoing MFCCB is exposed to multiple sessions (fractions) of radiation separated by one or more days to maximize the effectiveness of radiation on cancerous cells and to minimize the damage to healthy cells. In MFCCB, an updated set of 3D patient-specific planning images is acquired and segmented prior to the delivery of radiation at each fraction to identify changes in the clinical target volume and in the organs at risk (OAR).

Similar to other image-guided radiotherapy applications, many of the challenges faced in MFCCB are rooted in the fact that hollow elastic OAR, such as the urinary bladder, may have a significantly different size, shape and orientation across treatment fractions mainly due to differences in fill volumes and brachytherapy applicator positioning (Hellebust et al 2001). In other similar applications, the use of RU quantification in DIR-based assessment has been presented as a potential solution to the following challenges: (i) computing the accumulation of radiation doses across multiple fractions for quality control and post-treatment analysis (Risholm et al 2011, Murphy et al 2012, Tilly et al 2013); (ii) transferring the delineation of organs from one
fraction to the next to expedite the planning stage (Foskey et al 2005, Simpson et al 2011, Iglesias et al 2013, Heinrich et al 2016); and finally (iii) visualizing localized RU to assist clinicians when decisions are made based on registered image data (Risholm et al 2010).

1.3. Related works

Our survey of the prior works relating to computation and propagation of RU is naturally divided into two parts. In Section 1.3.1, we survey and outline the limitations of current mathematical methods for estimating RU and, in Section 1.3.2, we discuss notable works that propose application specific end-to-end frameworks for propagating the effects RU.

1.3.1. RU computation: Current approaches to estimating RU may be divided into three groups: 1) characterizing uncertainty from contextual image information (Wang et al 2015, Tang & Hamarneh 2013b), 2) frequentist approaches involving multiple registrations (Hub et al 2009, Kybic 2010, Watanabe & Scott 2012, Hub & Karger 2013), and 3) Bayesian approaches involving model inference on the posterior of the deformation parameters (Risholm et al 2010) and at the regularization level (Simpson et al 2015). Among these methods, Watanabe & Scott (2012) stands out as a suitable base for a solution in the context of MFCCB, and similar real world applications, since: it can be implemented on top of existing DIR software; it is computationally more efficient than the Bayesian approach of Simpson et al (2011); and it is capable of representing RU through ellipsoidal spatial confidence regions in the pixel-domain of the target image, which facilitates intuitive visualization and RU propagation. Despite these important advantages, some associated limitations exist which we address below.

It may be argued that a notable limitation of the cited frequentist methods, in contrast to the Bayesian approach of Simpson et al (2015), is that the computation of RU is evaluated from changes in image similarity subject to local random deformations without an explicit consideration for the global influence of regularization. Though the effects of regularization on RU have been shown to be significant, e.g., especially the case for complex intra-subject brain registration (Simpson et al 2012), the amount of computation required for inference on regularization parameters is costly even when efficient variational models are used. This cost hinders the applicability of such Bayesian methods in many real world situations, such as brachytherapy interventions. On the other hand, the frequentist RU computation methods are often pleasingly parallel and extendable to commonly used medical image registration software such as elastix (Klein et al 2010). We elaborate on this point further in Section 3.5.

Another limitation of the frequentist approach of Watanabe & Scott (2012) and the context-based method of Wang et al (2015) is that they are designed for unimodal registration tasks; wherein two images from the same modality, or even from the same patient, are registered together. In Watanabe & Scott (2012) for example, RU is estimated from intensity information of the moving image alone, which may not be
representative of the true RU. This approach to estimating RU is not suitable for MFCCB, even though DIR is being performed on different CT scans of the same patient, due to occasional observable differences in how the contrast agent presents within the bladder across different fractions. Although the protocol dictates that the amount of contrast agent must not be variable among different fractions of the same patient, noticeable variations occur due to both human error, which results in possible variations in the water-contrast mixture used for each fraction, as well as visible variation of contrast within the same image due to settling of the contrast material inside the bladder (Kim & Han 1991). This settling effect is further complicated by the fact that the bladder fill level and balloon catheter position are intentionally altered across fractions in some treatment centers to reduce the risk of urinary toxicity (Zakariaee et al 2016). A multimodal RU estimation method would therefore be better suited to MFCCB and other similar interventional applications.

1.3.2. RU propagation: Compared to the amount of literature on RU estimation methodologies, research into RU propagation in an end-to-end fashion is surprisingly scarce. Within the broad context of fractionated radiotherapy, which also includes external beam radiotherapy, the effects of DIR error propagation have been mainly studied on dose accumulation. Most notably, the Bayesian RU estimation method of Risholm et al (2010) was applied to oropharyngeal radiotherapy in Risholm et al (2011), while Murphy et al (2012) and Tilly et al (2013) implemented a frequentist approach to quantify RU during radiotherapy of the prostate. In Murphy et al (2012), principal component analysis was applied to multiple registrations of CT image pairs to obtain decorrelated modes of error, from which sample deformations (used to compute RU) were drawn. In Tilly et al (2013), a synthetic simulation framework was used to study the sensitivity of planning parameters to DIR. Radiation planning parameters are typically derived from dose-volume-histograms (DVH) that provide a summary of absorbed radiation over the entire volume of a structure, which include the target volume and the OAR (Nag et al 2004, Pöttner et al 2006). Cumulative DVH is important for planning and postoperative analysis as it has been shown to correlate with patient toxicity outcomes (Georg et al 2011). For hollow elastic OAR, in addition to DVH parameters, it is also important to study the spatial distribution of the accumulated doses as it is indicative of the formation of radiation hot spots and, thus, potential resulting complications (Zakariaee et al 2016). To the best of our knowledge, the effects of RU have only been studied on the entire dose volume (Risholm et al 2010) or the resulting DVH. There is still a need for planning parameters that can capture the spatial distribution of total radiation dose while accounting for the quality of DIR.

1.4. Contributions

In this paper, we propose a mathematical framework for multimodal RU computation and propagation that extends the RU estimation method of Watanabe & Scott (2012)
to address the limitations outlined in Section 1.3.1. Our framework is designed to interface with existing multimodal DIR software to extract RU, which is then used to generate uncertainty encoded visualization of DIR quality and to quantify the effects of RU on radiation dose accumulation and transfer of organ delineations from one fraction to the next. We also present evidence that our proposed framework is beneficial to the three challenges associated with image guided MFCCB outlined in Section 1.2. Specifically, we show that (i) the effect of RU on dose accumulation provides useful insights for quality control and post-treatment analysis; (ii) RU propagation improves the transfer of delineations from one fraction to the next; and (iii) RU can be used to generate intuitive visualizations that reflect the quality of DIR and thus can assist physicians in making decisions based on registered image data. We emphasize that our proposed methodology, though presented here to address the challenges of MFCCB, can be generalized to other similar clinical application as it is applicable to different DIR software, is faster than current Bayesian approaches (detailed in Section 4.4), and has been designed to handle multimodal registration tasks.

2. Methods

The details of our proposed framework are organized into four components and are presented as follows. We start by detailing our RU estimation technique in Section 2.1. In Section 2.2, we present a parametric representation of the computed RU along with two different methods for visualizing the computed RU. In Section 2.3, we present our uncertainty propagation method, which encodes the parametric representation of the computed RU into associated dosimetric maps and segmentation masks representing delineation of OAR. Finally, in Section 2.4, we take the analysis of RU propagation one step further by demonstrating how RU encoded dosimetric maps can be accumulated into an RU encoded accumulated dose map, which can then be summarized into a RU encoded DVH representation.

2.1. Computing registration uncertainties

Given a pair of volumetric images \( f_t, f_m : \mathbb{R}^3 \to \mathbb{R} \) henceforth referred to as target and moving image, the objective of parametric DIR models is to obtain an optimal set of transformation parameters \( \theta \) that describe a transformation function \( \mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^3 \). The function \( \mathbf{T}(x; \theta) \), defined over the domain of pixels \( x \in \Omega_t \subset \mathbb{R}^3 \) of the target image, maps a deformation field to the domain of the moving image \( \Omega_m \subset \mathbb{R}^3 \) in order to maximize its overall correspondence or similarity with the target image, measured with the similarity metric \( \Psi \). This optimization is written as

\[
\hat{\theta} = \arg\max_{\theta} \Psi(f_t(x), f_m \circ \mathbf{T}(x; \theta)).
\]  

If the solution to (1) is guaranteed to be correct every time (which is not the case in many real world applications), we should then be able to perturb (geometrically deform) the target volume in a controlled manner and expect the solution to (1) to
correctly identify exactly how the target image was perturbed. If the DIR process cannot correctly predict the perturbations, there must be uncertainties in the process, and we can therefore compute the difference between the controlled perturbations and the perturbations estimated with (1) to quantify the uncertainty in registration.

The process of applying controlled perturbations and estimating the errors in DIR is computationally expensive. Rather than applying random and unrealistic perturbations to the target image, we can leverage the initial DIR solution \( \hat{\theta} \) and prior knowledge about how we expect the registration to go wrong to induce these perturbations in a more informed and computationally efficient manner. As such, we compute RU by treating \( \theta \) as a normally distributed random variable \( \mathcal{N}(\mu_\theta, \Sigma_\theta) \), in which \( \mu_\theta \) is estimated by \( \hat{\mu}_\theta = \hat{\theta} \) and \( \Sigma_\theta \) is estimated depending on the deformation model used. \( \Sigma_\theta \) is estimated using an \emph{a priori} baseline \( \Sigma_0 \) in a convex combination

\[
\Sigma_\theta = (1 - \rho)\Sigma_0 + \rho \hat{\theta}\hat{\theta}^T \tag{2}
\]

weighted by \( \rho \in [0, 1] \) forming a shrinkage estimator. In context of 3D B-spline DIR, this \emph{a priori} baseline \( \Sigma_0 \) is computed by taking the Kronecker product of a 3D first order autoregressive model \( \Sigma_{AR} \) with a 3 \( \times \) 3 matrix \( c \) that encodes the covariance of the \( x, y, z \) components of the transformations parameters \( \theta \), such that \( \Sigma_0 = c \otimes \Sigma_{AR} \).

The 3D autoregressive kernel

\[
\Sigma_{AR}(i, j) = i_x^{\lfloor x(i) - x(j) \rfloor} r_x^{y(i) - y(j)} r_z^{z(i) - z(j)}, 1 \leq i, j \leq K \tag{3}
\]

is computed subject to smoothness parameters \( r = (r_x, r_y, r_z) \) and constraints \( |r_x|, |r_y|, |r_z| < 1 \) that control the smoothness between sampled neighboring B-spline control points, where \( K \) is the total number of control points. In this formulation, \( x(i) = \text{mod}(i - 1, n_x), y(i) = \text{mod}([(i - 1)/n_x], n_y), z(i) = [(i - 1)/(n_xn_y)] \) represent the mapping from index \( i \) to the corresponding \( x, y, z \) coordinate of the \( i^{th} \) control point, assuming an \( (n_x, n_y, n_z) \) grid of B-spline control points. The derivation of this prior and efficient techniques for sampling it are detailed in Watanabe & Scott (2012). The selection of the free parameters \( r \) and \( c \) indeed depends on the chosen DIR hyperparameters, e.g., number of B-spline control points, transformation regularization, optimization strategy, extent of possible deformable motion in clinical context, etc. A good strategy for selecting \( r \) and \( c \) would be based on initial clinical phantom studies, as suggested by Brock et al (2017), or from clinical experiments as explained in Section 3.

Once the distribution of \( \theta \) has been estimated, the next steps for computing RU are to: (i) draw samples \( \theta_i \sim \mathcal{N}(\hat{\mu}_\theta, \Sigma_\theta) \) from the distribution of parameters, (ii) apply the sampled transformation to the target image to generate synthetic images \( f^{(i)}_t = f_t \circ T^{-1}(x; \hat{\theta}) \circ T(x; \theta_i) \), (iii) register the moving image \( f_m \) to each and every synthetic image to obtain

\[
\hat{\theta}_i = \arg\max \Psi \left( f_t \circ T^{-1}(x; \hat{\theta}) \circ T(x; \theta_i), f_m \circ T(x; \theta) \right), \tag{4}
\]

and finally (iv) to compute the errors in the simulated transformations

\[
e_i(x) = (e_{ix}(x), e_{iy}(x), e_{iz}(x)) = T(x; \hat{\theta}_i) - T(x; \theta_i). \tag{5}
\]
These errors represent the performance of the DIR method, in the spatial domain of the target image, subject to a set of random but structured perturbations (governed by prior $\Sigma_o$) applied to the target image. These errors implicitly account for overall end-to-end performance of a given a pair of images and a DIR method with fixed hyperparameters. The overall end-to-end performance accounts for both the contextual geometric uncertainties that exist in the image pair (e.g., lack of salient image information) and the behaviour of the DIR method (e.g., correctness of the deformation model and performance of the optimization strategy used) subject to the contextual geometric uncertainties. The spatial distribution of these errors are therefore the RU associated to the DIR process. The estimation of RU is improved by drawing more samples from $\theta_i$, however this process is expensive as computing an additional $e_i$ requires a computationally expensive 3D registration. In the next section, we describe how this expensive process can be simplified by assuming a parametric model over the distribution of the RU computed for DIR parameter $\theta$, and by extension, at every voxel $x$.

### 2.2. Parametric representation and visualization of RU

The benefits of assuming a parametric model for the distribution of RU are three-fold. This parametric representation reduces memory requirements for storing $\{e_i(x)\}$, reduces the number of $e_i$ samples required to estimate RU, and simplifies the process of propagating the RU onto the next processing steps, i.e., using the spatial confidence to diffuse segmentations or dose maps. Assuming that $e(x) \sim \mathcal{N}(\mu_e(x), \Sigma_e(x))$, spatial confidence region $\Phi(x)$ is defined by

$$\Phi(x) = \{x' : (x - \mu_e(x))^\top \Sigma_e^{-1}(x - \mu_e(x)) < \chi^2_3(1 - \gamma)\},$$

(6)

where $\gamma$ is the confidence level and $\{\mu_e(x), \Sigma_e(x)\}$ are the sample mean and covariance of $\{e_i(x)\}$. Sample covariance calculation is sensitive to outliers and becomes a problem if the synthetic samples are few. Student’s t-distribution is a better fit but it is computationally complex as its parameters have to be solved numerically. We propose a two-step estimation of the sample covariance by omitting the outliers that fall outside of the $\gamma = 0.95$ level set of the first estimate of $\Phi(x)$. It should be noted that even if $\theta$ is sampled normally, there is no guarantee that the errors $\{e_i(x)\}$ will follow a normal distribution. This matter is further discussed in the Results section.

The confidence regions $\Phi(x)$ are effectively structural tensors parameterized by $\{\mu_e(x), \Sigma_e(x)\}$ at every voxel $x$. To visualize the quality of DIR at each voxel, these structural tensors may be rendered as ellipsoids in 3D using tensor rendering toolkits available for medical images. An alternative visualization may also be produced by transforming these structural tensors to scalar quantities that represent RU by computing the differential entropy of each tensor

$$E(x) = \frac{3}{2} \left(1 + \ln(2\pi)\right) + \frac{1}{2} \ln |\Sigma_e(x)|.$$

(7)

Note that differential entropy $E$ is a logarithmic measure, which depends on the log-determinant of the covariance matrix $\Sigma_e(x)$ and as a result $E \in (-\infty, \infty)$.
In the next section, we show how this parametric representation is used within our framework to propagate RU beyond the registration step.

2.3. Diffusion of image information using RU

The parametric representation of RU, in essence, presents the likelihood of possible locations where a voxel from the domain of the moving image would be mapped to in the domain of the target image. As such, RU also presents the likelihood that a voxel in the target image domain may have the values of its neighbors due to potential errors in registration. We can propagate the effects of potential DIR errors onto associated mappings, e.g., segmentation labels and dose maps, using a weighted averaging scheme that encodes the likely contributions of neighboring voxels.

Let $I_m = (D_m, S_{m1}^1, S_{m2}^2, ..., S_{mN}^N) : \mathbb{R}^3 \rightarrow \mathbb{R}^{N+1}$ represent a vector valued volume that contains $N$ binary segmentation labels $S_{m} \in \{0, 1\}$ of different OAR and radiation dose map $D_m$ in the domain of the moving image $f_m$. This image can be transformed into the domain of the target image using the parameters $\hat{\theta}$ such that $I'_m(x) = I_m \circ T(x; \hat{\theta})$, $x \in \Omega_t$. To compute the posterior distribution $P'_m \sim \mathcal{N}(\mu_P, \sigma_P^2)$ of $I'_m$, RU is encoded using a convex combination characterized by the weighted mean and variance

$$
\mu_P(x) = \frac{\sum_{x' \in \Omega_t} W(x| x') I'_m(x')}{\sum_{x' \in \Omega_t} W(x| x')},
$$

$$
\sigma_P(x)^2 = \frac{\sum_{x' \in \Omega_t} W(x| x')(I'_m(x') - \mu_P(x))^2}{\sum_{x' \in \Omega_t} W(x| x')},
$$

where the weights

$$
W(x| x') = (2\pi)^{-\frac{3}{2}} |\Sigma_e(x')|^{-\frac{1}{2}} e^{-(\frac{1}{2}(x-\mu_e(x'))'\Sigma_e(x')^{-1}(x-\mu_e(x')))}
$$

are obtained by evaluating the probability density of the trivariate normal distribution $\mathcal{N}(\mu_e(x'), \Sigma_e(x'))$ at $x$. The computational performance of this diffusion process can be increased by assuming that the contribution of $I'_m(x')$ to $P_m(x)$ is negligible, i.e. $W(x| x') \sim 0$, when $x$ and $x'$ are more than a certain distance apart; effectively reducing the space of $x' \in \Omega_t$, over which (8) is evaluated.

It is worth noting that a possible simplification of (8), which we do not perform, is to assume that $W(x| x') = W(x'| x)$, then the computation of the mean simplifies to $\mu_P = W * I'_m$. This allows for the mean of the posterior to be calculated by convolution or, in other words, by blurring $I'_m$ with an adaptive Gaussian kernel parameterized by $\{\mu_e(x), \Sigma_e(x)\}$. This approach was used in Simpson et al. (2011) to diffuse segmentation labels in a study involving registration of inter-subject brain MRI scans.

Having computed a parametric distribution of associated dose maps $\{\mu_{D_m}, \sigma_{D_m}\}$ as part of $P'_m$, we can now extend the propagation of RU though to the accumulated dose map and resulting DVH representation, which is detailed in the following section.
2.4. Accounting for RU in DVH representation of accumulated dose

DVH parameters have been used in numerous published toxicity analyses, which are the basis for recommended dose constraints used in modern-day treatment planning. It is important to compute and visualize the effects of RU on the DVH curves as the RU information may be used to (i) simulate potential errors in how the DVH was computed, and to (ii) provide uncertainty metrics for the DVH-derived parameters that are used in practice for evaluating the treatment plans. We propose to use the propagation technique presented in Section 2.3 to compute the distribution of accumulated dose at each pixel after dose information from different fractions have been probabilistically aligned onto the target (reference) image volume. An RU encoded probabilistic DVH curve is then computed from the generated probabilistic accumulated dose map.

The DVH for each \( n \in N \) OAR, denoted \( H_n : d \in \mathbb{R} \rightarrow \mathbb{R} \), is defined as the cumulative distribution function of the number of pixels within an OAR that receive an accumulated radiation dose above a certain dose value \( d \). Let \( x_n \subset x \) define the subset of pixels that fall within a certain OAR indexed as \( n \), as indicated by the corresponding segmentation label, such that \( S_n(x_n) = 1, \forall x_n \). The DVH is formally defined as

\[
H^n(d) = \sum_{\forall x_i \in x_n} \mathbb{I}(D_{\text{acc}}(x_i) > d),
\]

where \( \mathbb{I}(\cdot) \) is the indicator function. \( D_{\text{acc}}(x) \) represents dose map containing the summation of the total amount of radiation received throughout all fractions accumulated onto one fraction (selected as the target volume) such that

\[
D_{\text{acc}}(x) = D_t(x) + \sum_{\forall m} D_m \circ T(x; \hat{\theta}_m),
\]

where \( D_t \) is the dose map of the fraction selected as the target and \( D_m \) are the dose maps of all \( m \in M \) moving fractions. The DIR alignment of each moving fraction onto the target fraction is parametrized by \( \hat{\theta}_m \) obtained by registering each \( f_m \) to \( f_t \).

Having defined how the DVH is computed using (10), we can now extend the formulation to account for RU by leveraging the parametric representation of probabilistic dose maps \( \{\mu_{D_m}, \sigma_{D_m}\} \) computed in the previous section. Assuming that \( \{\mu_{D_m}, \sigma_{D_m}\} \) computed for each moving fraction \( m \in M \) are independent random variables, the sum of them is also normal, with its mean being the sum of the means, and its variance being the sum of the variances. Summing the aligned RU encoded dose results in the parametric representation of the accumulated dose map

\[
\mu_{D_{\text{acc}}}(x) = D_t(x) + \sum_{\forall m} \mu_{D_m},
\]

\[
\sigma_{D_{\text{acc}}}(x)^2 = \sum_{\forall m} \sigma_{D_m}^2.
\]

The DVH may now be computed at different \( \alpha \) standard deviations from the mean accumulated dose value such that

\[
H^n(d; \alpha) = \sum_{\forall x_i \in x_n} \mathbb{I}((\mu_{D_{\text{acc}}}(x_i) + \alpha \sigma_{D_{\text{acc}}}(x_i)) > d).
\]
In summary of the methodologies presented in this section, we have proposed a framework that is capable of (i) estimating the quality of a B-spline DIR method via RU, (ii) visualizing the RU, and (iii) propagating the effects of RU onto associated segmentation labels and dose maps using a weighted averaging technique. We then went a step further to propagate the effects of RU on the accumulated dose from multiple fractions. The corresponding DVH representation, which may be used for interventional planning and quality assessment purposes, was also computed. We discuss the potential benefits of RU encoded DVH analysis in further detail in Section 4.2. In the following section, we demonstrate the utility of our framework on a retrospective study of MFCCB interventions.

3. Data and Experimental Setup

We applied the methods outlined in the previous section to a retrospective 37 patient dataset of MFCCB collected at BC Cancer Agency, Vancouver, Canada, following applicable Research Ethics Board review. Each of the 37 instances within the dataset is comprised of five fractions and the set of planning images acquired at each fraction contains a computed tomography (CT) scan $f$, a detailed segmentation of the interior $S^i$ and exterior $S^e$ surfaces of the bladder wall, and a dosimetric map $D$, representing the amount of radiation absorbed at every voxel. The segmentation and dose maps in each instance are in the same coordinate space as the CT volume. Also, three landmarks on the urinary bladder that make up the Trigone (bladder neck and ureteral orifices) were identified in all CT scans to enable the calculation of target registration error (TRE). Landmarks were manually localized with respect to the origin of the corresponding CT scan. Landmark locations were not used to facilitate the alignment of the image volumes and no further processing was done to the image volumes prior to registration.

The clinical goal in MFCCB analysis is to account for changes in patient anatomy between fractions, thus we need to perform intrapatient registration. The CT volume of the first fraction is selected as the target volume $f_t$ to which the CT volumes of all other fractions are registered. Automatic registration is performed in two stages: (i) a simple affine image registration to roughly align the volumes followed by (ii) a DIR to account for elastic deformations. The elastix toolbox (Klein et al. 2010), which is based on the popular Insight Segmentation and Registration Toolkit (ITK, Kitware Inc.), was used to perform the registrations and the same pair of affine and DIR parameter files were used for all registrations. Corresponding registration parameter files are included in the supplementary materials. To speed up computation, each image volume was downsampled by a factor of two; to $256 \times 256$ axial resolution.

The error in landmark alignment is computed before and after every stage of registration by computing the euclidean distance between the Trigone landmarks of the target and moving image volumes. The landmark alignment error is averaged across the four fractionated moving volumes of each patient and is presented in Fig. 1. The automatic affine image registration step of our framework reduced the mean landmark
alignment errors across four moving fractions and 37 patients from 29.57 mm to a mean TRE of 10.67 mm. This mean TRE is further reduced from 10.67 mm to 7.71 mm following the initial DIR. All resulting reductions in landmark alignment errors following each registration step were found to be statistically significant ($p < 0.001$) subject to a Wilcoxon signed-rank test.

RU is computed only for the DIR step of the registration. To compute the RU after the initial DIR, our method requires an a priori baseline $\Sigma_0$ in order to realize 40 samples of the transformation parameters. We parametrize the free parameters of the first order 3D autoregressive model $\Sigma_{AR}$ by setting the value of $c$ to the covariance of absolute $x,y,z$ error components of the TRE computed post-DIR. We set the smoothness parameter $r$ to 0.9 for all $x,y,z$ directions, and $\rho$ to 0.1 as recommended in the original paper (Watanabe & Scott 2012).

To increase the computational performance of diffusion process, we restricted the space over which (8) is evaluated to a $80 \times 80 \times 80$ mm neighborhood about $x$. Effectively, we assume that $W(x|x') \sim 0$, when the distance between $x$ and $x'$ is more than 40 mm, i.e., twice the maximum measured TRE post-DIR.

4. Results and Discussion

4.1. Registration results and RU visualizations

Applying the transformations obtained with the initial DIR to align the segmentations of the entire bladder, up to the external bladder wall $S_{e}'$, with the target volume results in an average Dice similarity coefficient (DSC) of 0.8599 with a standard deviation (STD)
Propagation of registration uncertainty during fractionated brachytherapy

Figure 2. Scalar (left) and tensor (right) representations of registration uncertainty overlayed on CT. In the scalar visualization, the colors of the overlay represent uncertainty ranging from most certain (blue) to most uncertain (red). This uncertainty corresponds to the volume of the ellipsoids depicted in the tensor visualization. In the tensor representation, the colors indicate the orientation of the major axis of the ellipsoid or direction of maximum uncertainty. Note that this exemplar registration result is more certain at rigid structures (bones) compared to deformable ones (bladder, bowels, and air) and the tensor orientations follow the edge information in the image.

In the methodology, we presented a parametric representation for $\Phi(x)$, which can be interpreted as structure tensors. After computing RU, these tensors can be rendered as 3D ellipsoids using existing tensor rendering toolkits available for medical images (Toussaint et al. 2007) or, alternatively, transformed into scalar quantities using (7). An exemplar visualization of the computed RU as scalar and tensor overlays is provided in Fig. 2. To generate the color mapping for the scalar visualizations,
we first normalized the computed differential entropy values to $E_N \in [0, 1]$ such that
$$E_N = \frac{(E - \min(E))}{(\max(E) - \min(E))}$$
and mapped color values to the normalized $E_N$ values to represent relative uncertainty in DIR. In Fig. 2, $\min(E) = -1.64$ and $\max(E) = 7.91$. Qualitatively, the proposed visualizations of RU are as expected; RU is more certain at rigid structures (bones) compared to deformable ones (bladder, bowels, and air) and the tensors follow the edge information in the image. Such visualizations may be useful in many similar image guided applications but its utility to clinicians needs to be thoroughly validated.

4.2. Propagation of RU onto segmentation labels

We encode RU by diffusing the aligned binary segmentation labels $S_{m^{'}}$ using (8) and observed improvements in the alignment of the bladder labels following the diffusion (Fig. 3). The resulting mean of the diffused labels $\mu_{S_{e}}$ are fuzzy and in order to quantify the improvement over the initial crisp alignment, we first threshold $\mu_{S_{e}}$ at 0.5 and compute the DSC. In this scenario, the average DSC increases to $0.8619 \pm 0.0925$. The improvement may seem marginal but this is due to the fact that precious information is discarded during thresholding.

The fuzzyness of $\mu_{S_{e}}$ itself conveys useful information as well. The diffusion of $\mu_{S_{e}}$ is based on local RU; if $\mu_{S_{e}}$ is thresholded at values other than 0.5, it follows a trade-off between true positive and true negative rates. This behavior is evident in the

‡ In this article, we use the adjective ‘crisp’ to refer to deterministic or non-probabilistic variables and processes that do not encode uncertainty information.
Propagation of registration uncertainty during fractionated brachytherapy

Figure 4. Quantitative effects of DIR and RU based diffusion on segmentation labels of the entire bladder summarized using mean ROC curves comparing the mean performance of bladder alignment using different RU diffusion methods. The results of a discrete classifier (red) and an isotropically diffused version of it (black) are used as baselines for comparison. The isotropic diffusion was generated by performing 3D morphological dilation and erosion operations on the discrete segmentation. Our diffusion technique based on the proposed RU method (green) outperforms diffusion using Watanabe & Scott (2012) (blue) and the baselines. The associated area under the mean ROC curves are 0.9442 (red), 0.9844 (black), 0.9872 (blue), and 0.9909 (green).

Receiver Operator Characteristic (ROC) space (Fawcett 2006) shown in Fig. 4. In the ROC space, the correctness of the alignment can be measured by the area under the ROC curve (AUC). To perform this comparison, we interpreted the crisp alignment results as a discrete pixel classifier and $\mu_{SE}$ as a probabilistic one, and compared their performance in classifying the ground truth bladder pixels in the target image. With the $\mu_{SE}$ classifier, the average AUC (across 4 fractions and 37 patients) was increased from 0.9422 to 0.9909. Note that these reported numbers are high due to the fact that the relative number of true negative responses outnumber the misclassified ones in the CT volume. By comparison, the Watanabe & Scott (2012) method for computing RU, resulted in an average AUC of 0.9872. Furthermore, to demonstrate that diffusion based on our RU estimation is more representative of true registration error than an isotropic diffusion, and that the improvements in segmentation are not due to under-or over-segmentation, rather than using an adaptive RU diffusing kernel, we eroded and dilated the crisp segmentations with a 3x3x3-voxel structuring element and observed an average AUC of 0.9844.

A summary of the quantitative improvements in segmentation accuracy are provided in Fig. 5 and Table 1. We tested the significance of the reported AUC improvements using the Wilcoxon signed-rank test and observed that the differences between the Watanabe & Scott (2012) method and isotropic diffusion were not significant; on the other hand, the differences between all other AUC comparisons
Crisp Isotropic Watanabe proposed

Figure 5. Boxplots summarizing labelling performance following the proposed and Watanabe & Scott (2012) RU based diffusion of registered segmentation labels compared to the standard and isotropically diffused baselines. Performance results reflect performance across all 148 registrations and are measured using the Dice similarity coefficient (DSC) and the area under the ROC curve (AUC). Note that the DSC results for standard and isotropically diffused baselines are identical as the isotropic diffusion is derived from the crisp results.

<table>
<thead>
<tr>
<th>Method</th>
<th>AUC</th>
<th>DSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crisp baseline</td>
<td>0.9422 ± 0.0603</td>
<td>0.8599 ± 0.0921</td>
</tr>
<tr>
<td>Isotropically diffused</td>
<td>0.9844 ± 0.0343†</td>
<td>0.8599 ± 0.0921</td>
</tr>
<tr>
<td>Watanabe &amp; Scott (2012)</td>
<td>0.9872 ± 0.0281†</td>
<td>0.8545 ± 0.0997</td>
</tr>
<tr>
<td>Proposed RU method</td>
<td>0.9909 ± 0.0207</td>
<td>0.8619 ± 0.0925</td>
</tr>
</tbody>
</table>

Table 1. Summary of the labelling performance across all 148 registrations measured using the Dice similarity coefficient (DSC) and the area under the ROC curve (AUC). Note that the DSC results for crisp and isotropically diffused baselines are identical as the isotropic diffusion is derived from the crisp results. †: The differences between the computed AUC for the isotropically diffused and Watanabe & Scott (2012) approach were not found to be significant.

were significant ($p < 0.001$). This insignificant improvement over an isotropic diffusion implies that the standard unimodal Watanabe & Scott (2012) method may not be suitable for RU estimation in context of MFCCV, which is further corroborated by the the associated drop in mean DSC when the Watanabe & Scott (2012) method is used. Note that all differences in DSC were significant ($p < 0.001$), with the obvious exception of the crisp and isotropic diffusion approach that are identical in terms of DSC performance. From these observations we conclude that, as expected, our proposed multimodal RU estimation approach can better estimate the DIR errors as it is capable of improving segmentation performance through propagation of RU.
A critically important, yet subtle, point that merits repeating and further discussion at this point is in regards to the direction of the RU diffusion. This behaviour can only be seen in the tensor (vector) visualization presented in Fig. 2. From this figure, one can easily observe that the diffusion of RU is greater in homogeneous regions (e.g., inside of the bladder) compared to regions with salient image texture such as observable tissue boundaries. More importantly, it can be seen that the diffusion is greater along the direction of a tissue boundary than across it. In other words, the major axis of the ellipsoidal tensors are aligned with tissue boundaries, e.g., the bladder wall. This is logical as the DIR algorithm can align two edges with relative ease, however perfect correspondence between two points along these edges is so far impossible to obtain, without the aid of perfectly accurate tissue deformation models, due to the lack of salient image information along the edge. Compared to the abdominal cavity and surrounding soft tissues, the boundaries of the bladder wall are easy to identify in most of the images within our dataset due to the use of contrast agents. It is thus of no surprise that the proposed RU diffusion results in a seemingly incremental improvement in segmentation alignment. By contrast the effects of RU on dose accumulation analysis, presented in the following section, is more significant.

4.3. Effects of RU on accumulation of dose volumes

The aligned dose volumes \( D' \) acquired during the planning stages are also diffused using (8). The resulting uncertainty in accumulated diffused dose maps are first computed using (12) and then used to compute the DVH (13). A visual comparison between the crisp and probabilistic accumulated dose maps over the bladder walls (intersection of the interior and exterior bladder wall segmentation labels \( S_i \cap S_e \)) are provided in Fig. 6.

As shown in Fig. 6, the effects of RU on the resulting DVH is significant. In this exemplar case, the mean of the RU encoded accumulated-dose map (Fig. 6b) is slightly higher than that of the crisp DIR aligned dose maps (Fig. 6a), especially at the radiation hot spot that is formed on the bladder wall. This effect can also be observed from the small discrepancy between the blue and dotted black line in the DVH curves of Fig. 6. Considering a widely used clinical dose metric such as \( D_{2cc} \) (the minimum dose received by the most exposed 2 cm\(^3\) of tissue), which characterizes the intensity of hot spots that have the potential to cause significant morbidity (Pötter et al 2006), the DVH curves for this example yield a \( D_{2cc} \) value of 28.5 Gy using the crisp approach (dashed black line), whereas the probabilistic method (blue line) computes a more conservative value of 29.2 Gy. The mean of the RU encoded dose map \( \mu_{D_{acc}} \) is more conservative (in terms of total absorbed radiation) than that obtained from a crisp DIR alignment and, building on our conclusions from the experiments with RU encoded segmentation label propagation in Section 3.3, we hypothesize that our RU encoded accumulated-dose map and corresponding DVH curves are more representative of the true accumulated dose compared to the naïve crisp DIR approach.

Furthermore, it is evident that there exists a large uncertainty in the accumulated-
Figure 6. Effects of RU on accumulated dose map for one patient. Top: comparison between accumulation of crisp aligned $D_{acc}$ dose (a) versus the mean $\mu_{D_{acc}}$ (b) and standard deviation $\sigma_{D_{acc}}$ (c) of the accumulated RU-based diffusion of $D_{acc}$, which encodes the effects of RU at every voxel. Bottom: the effects of RU on the DVH. The volume of the bladder wall presented in this example is 49 cm$^3$.

dose DVH curves. More importantly, this uncertainty is greatest at high accumulated dose thresholds that constitute a smaller percentage of the total volume of the bladder. In the above example, the probabilistic DVH curves indicate that within six standard deviations (99.7% confidence interval) around the mean of the computed probabilistic dose values $\mu_{D_{acc}}$, the true $D_{2cc}$ may lie anywhere within the range of 21.6 Gy (green line) to 37.8 Gy (red line). In Fig. 6c, this region of uncertainty is shown to be spatially localized in the proximity of the small-volume/high-dose region, i.e., the radiation hot
spot, which is of more radiobiological interest. The implications of this phenomena are crucial as we believe that this observed localized variability in accumulated dose casts doubts on the usefulness of planning parameters that are derived from the DVH alone, and consequently leans in favor of methods that also account for the spatial distribution of the accumulated dose such as Zakariae et al (2016). The demonstrated effects of RU on the accumulated dose corroborates the surveyed prior art in other applications, yet their impact on MFCCB analyses is to be explored.

### 4.4. Computational Performance

Our pipeline is well suited to applications where image volumes from different modalities need to be registered due to our choice to use the established multimodal mutual information similarity metric readily implemented in elastix. Furthermore, our algorithm is fast; our unoptimized MATLAB plus elastix implementation takes, on average, 37 minutes to compute RU and 12 minutes to diffuse both the segmentation and dose volumes per fraction on an 8-core workstation with two Intel 3GHz Xeon x5472 processors and 8GB of RAM, which is faster than the Bayesian approach of Risholm et al (2011) that takes 11 hours per fraction on an 8-core machine. In our RU computation for every fraction, the initial registration and sample generation takes less than one minute and the following 40 registrations are independent and may be performed in parallel. Similarly, the diffusion equations (8) may be evaluated at each voxel location independently, and therefore higher performance gains are expected with a larger computing cluster. It should be noted that the computational performance of Risholm et al (2011) can be improved with the help of modern sampling techniques, i.e., as was proposed in Simpson et al (2012) and Simpson et al (2015) with variational techniques for performing Bayesian inference. We were however unable to quantify the performance improvements proposed in Simpson et al (2012) and Simpson et al (2015) as performance times were not reported in their publications.

### 5. Conclusions

In this paper, we proposed a mathematical framework for estimating RU and propagating the effects of the computed uncertainties from the registration stage through to the dosimetric evaluation and visualization stages that are critical to the physician performing radiation therapy. Specifically, we proposed a novel and computationally efficient RU computation technique that enables a parametric representation of RU as structure tensors, as well as two different methods for visualizing the computed RU. We further presented an uncertainty propagation method that encodes the RU into dosimetric maps and organ segmentation masks. Finally, we demonstrated how the RU encoded dosimetric maps computed for different fractions can be accumulated into an RU encoded accumulated dose map, which can then be summarized into a RU encoded DVH representation.
We designed our proposed framework to compute RU from existing B-spline DIR software such as the popular elastix software package. Our framework can therefore be easily integrated within other image-guided radiotherapy applications as it generally does not require major changes to the existing medical image analysis workflow.

In the context of our showcased clinical application, we provided preliminary evidence indicating that the computation and propagation of RU results in improvements during treatment planning and quality control stages of MFCCB. Specifically, we demonstrated how our proposed RU estimation and propagation can be used to visualize potential errors in DIR and, further, to transfer the delineation of organs from a prior fraction to the next more accurately. We also substantiated the utility of RU propagation in evaluating the potential errors and deviations in accumulated dose at a specific voxel, as well as in the DVH. More importantly, we observed that the resulting accumulated dose uncertainty is greatest in the proximity of the small-volume/high-dose region, i.e., the radiation hot spot, which is of significant radiobiological interest. Our next objectives include assessing the effects of RU on the correlation between accumulated dose and MFCCB patient toxicity outcomes. We also plan to extend our framework to group-wise or symmetric registration frameworks to mitigate uncertainties and potential biases associated with manual selection of one image volume as the reference image.

References


Dawson L A & Sharpe M B 2006 Image-guided radiotherapy: rationale, benefits, and limitations The lancet oncology 7(10), 848–858.

Fawcett T 2006 An introduction to ROC analysis Pattern Recognition Letters 27(8), 861–874.


Hellebust T P, Dale E, Skjonsberg A & Olsen D R 2001 Inter fraction variations in rectum and bladder volumes and dose distributions during high dose rate brachytherapy treatment of the uterine cervix investigated by repetitive CT-examinations Radiotherapy and Oncology 60(3), 273–280.

Holloway R L 1997 Registration error analysis for augmented reality Presence: Teleoperators and Virtual Environments 6(4), 413–432.
Propagation of registration uncertainty during fractionated brachytherapy

Hub M & Karger C 2013 Estimation of the uncertainty of elastic image registration with the demons algorithm Physics in Medicine & Biology 58(9), 3023.

Hub M, Kessler M L & Karger C P 2009 A stochastic approach to estimate the uncertainty involved in B-spline image registration Transactions on Medical Imaging 28(11), 1708–1716.


Jannin P, Krupinski E & Warfield S 2006 Validation in medical image processing Transactions on Medical Imaging 25(11), 1405.


Kybic J 2010 Bootstrap resampling for image registration uncertainty estimation without ground truth Transactions on Image Processing 19(1), 64–73.


Tilly D, Tilly N & Almesjö A 2013 Dose mapping sensitivity to deformable registration uncertainties in fractionated radiotherapy—applied to prostate proton treatments BMC medical physics 13(1), 2.


Watanabe T & Scott C 2012 *in* ‘Biomedical Image Registration’ Springer pp. 120–130.


Appendix A. Registration errors

Figure A1. Exemplar case depicting the effects of registration on bladder and landmark alignment. 3D contours of the bladder presented in this figure are extracted from the target image volume (black) and the moving image at different stages of alignment: before registration (red), after affine registration (blue), and after affine+DIR (green). Landmark locations (two ureter orifices and the bladder neck) are plotted as spheres and are colored according to their corresponding 3D bladder contour. Correspondence between landmark locations are indicated with a black line. In this example, the mean landmark alignment error is reduced from 36.4 mm before registration to a mean TRE of 12.9 mm after affine and 5.4 mm after affine+DIR.

Figure A2. Independent and joint histograms of the $x, y, z$ components of displacements $E$ between corresponding landmarks. Displacements are computed for all three corresponding pairs of landmarks (two ureter orifices and the bladder neck) with respect to their locations in the target image volume and the moving image volume before and after every stage of image registration.