



HAL
open science

Semi-automatic segmentation of whole-body images in longitudinal studies

Eloïse Grossiord, Laurent Risser, Salim Kanoun, Richard Aziza, Harold Chiron, Loïc Ysebaert, François Malgouyres, Soléakhéna Ken

► **To cite this version:**

Eloïse Grossiord, Laurent Risser, Salim Kanoun, Richard Aziza, Harold Chiron, et al.. Semi-automatic segmentation of whole-body images in longitudinal studies. *Biomedical Physics & Engineering Express*, 2020, 7, pp.015014. 10.1088/2057-1976/abce16 . hal-03234347

HAL Id: hal-03234347

<https://hal.science/hal-03234347>

Submitted on 25 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

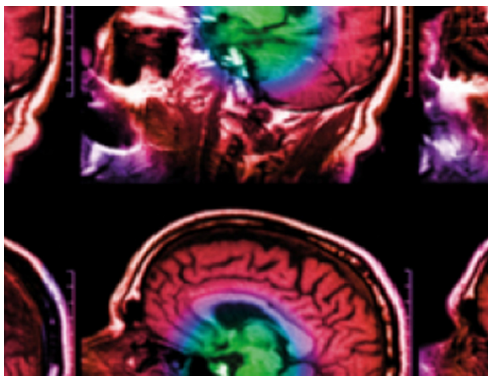
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PAPER • OPEN ACCESS

Semi-automatic segmentation of whole-body images in longitudinal studies

To cite this article: Eloïse Grossiord *et al* 2021 *Biomed. Phys. Eng. Express* 7 015014

View the [article online](#) for updates and enhancements.



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.

Biomedical Physics & Engineering Express



PAPER

Semi-automatic segmentation of whole-body images in longitudinal studies

OPEN ACCESS

RECEIVED

18 August 2020

REVISED

2 November 2020

ACCEPTED FOR PUBLICATION

26 November 2020

PUBLISHED

8 December 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Eloïse Grossiord^{1,2}, Laurent Risser¹, Salim Kanoun³, Richard Aziza³, Harold Chiron³, Loïc Ysebaert³, François Malgouyres¹ and Soléakhéna Ken²

¹ Institut de Mathématiques de Toulouse ; UMR5219 ; Université de Toulouse ; CNRS UPS IMT F-31062 Toulouse Cedex 9, France

² Institut Universitaire du Cancer Toulouse Oncopôle (IUCT-O) - Institut Claudius Regaud ; Département d'Ingénierie et de Physique Médicale, F-31059 Toulouse Cedex 9, France

³ Institut Universitaire du Cancer Toulouse Oncopôle (IUCT-O) - Institut Claudius Regaud ; Département d'Imagerie Médicale, F-31059 Toulouse Cedex 9, France

E-mail: Ken.Soleakhena@iuct-oncopole.fr

Keywords: image segmentation, fast marching, longitudinal segmentation, whole-body MRI, clinical trials

Abstract

We propose a semi-automatic segmentation pipeline designed for longitudinal studies considering structures with large anatomical variability, where expert interactions are required for relevant segmentations. Our pipeline builds on the regularized Fast Marching (rFM) segmentation approach by Risser *et al* (2018). It consists in transporting baseline multi-label FM seeds on follow-up images, selecting the relevant ones and finally performing the rFM approach. It showed increased, robust and faster results compared to clinical manual segmentation. Our method was evaluated on 3D synthetic images and patients' whole-body MRI. It allowed a robust and flexible handling of organs longitudinal deformations while considerably reducing manual interventions.

1. Introduction

In the context of clinical trials, assessing the disease burden at different therapy stages is essential to evaluate the efficacy of a drug treatment. For each patient, such evaluations may require the segmentation of a given set of structures from 3D images acquired during serial medical examinations. Due to the potentially high anatomical variability of these structures between patients, their segmentation is not straightforward to automatize. Another source of complexity when dealing with longitudinal studies comes from the strong evolution of a given structure from an exam to the follow-up exams, e.g. when dealing with tumour control and treatment response (see figure 1). In addition, the boundaries of the structures to segment may not always be clearly visible in the acquisitions. Also, the number of exploitable data can be limiting in the development of segmentation strategies because clinical trials are often retrospectively analyzed with heterogeneous images acquisitions and/or because the study is performed at a preliminary stage. As a result, machine-learning or fully automatic strategies are not adapted as a consistent ground-truth training dataset with such

non-standard structures with high variability is not always available. Manual intervention by experts are therefore often required to obtain clinically exploitable segmentations, making such studies particularly time consuming.

The goal of this paper is to present a new semi-automatic and clinically relevant segmentation technique for longitudinal images, aiming at reducing the time dedicated to manual interventions. Precisely, this pipeline is intended to be generic to integrate manually-dependent segmentation and extend it longitudinally alleviating the necessary manual intervention as much as possible.

The proposed generic segmentation pipeline was developed in the context of a clinical trial evaluating Chronic Lymphocytic Leukemia (CLL), which is the most common leukemia among adults (Noone *et al* 2016), for patients under Ibrutinib treatment. From a cohort of 10 patients with several follow-up time points (at baseline (M0), after one month (M1) and twelve months of treatment (M12)), we extracted several in-vivo and non-invasive clinical imaging biomarkers and followed their evolution along treatment. All biomarkers were associated to specific organs and extracted from whole-body MR images. Organs

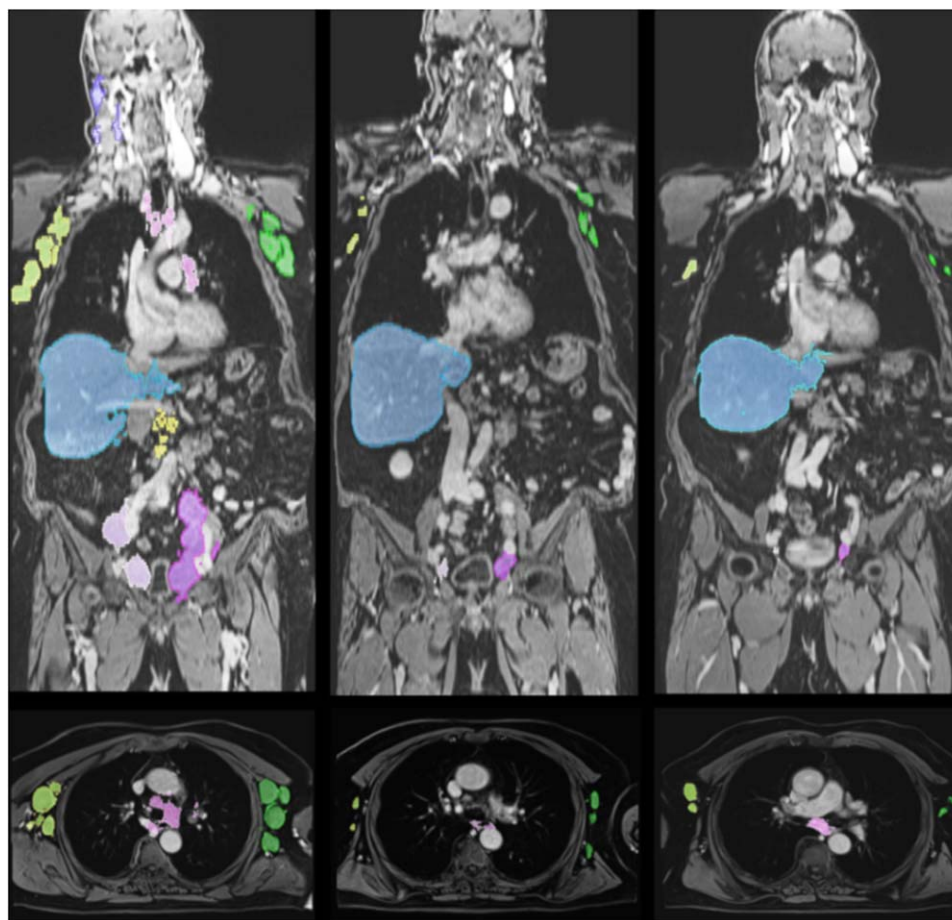


Figure 1. Longitudinal whole-body MR images of a CLL patient with M0, M1 and M12 time of follow-up. Coronal views (top) present the segmentation results for the regions of interest wide spread all over the body; highlighted in color are the cervical, axillary, mediastinal, retroperitoneal, iliac lymph nodes and liver. Focus on variability of axillary lymph nodes (right and left, contoured in light green and green, respectively) is given along the different acquisition times on the axial views (bottom). Notice the important volumes of the axillary nodes at baseline before Ibrutinib treatment; their significant shrinkage at M1 follow-up and then the slight volume increase at M12 follow-up. These illustrations are from (Ysebaert *et al* 2017).

manual delineation was especially tedious for CLL as it is a diffuse disease affecting numerous organs all over the body, with a great variability (figure 1) among patients and among the image acquisitions to consider; adding significant inter-operator variability to the process. Therefore, the development of a specific algorithm to handle whole-body segmentation and adapted to follow-up exams was identified as very beneficial for time, accuracy and robustness gain.

Longitudinal segmentation has been widely explored in the different medical imaging modalities. In this article, we focus on the segmentation of follow-up MR images representing lymph nodes and different organs. Tracking tumorous livers and lymph nodes was however only addressed by a few approaches, mainly in CT images (Moltz *et al* 2009, Vivanti *et al* 2018). In MRI, longitudinal studies strongly focused on brain segmentation, where the anatomy is well known (Gordon *et al* 2018), or on MS lesion segmentation (Schmidt *et al* 2019). To the best of our knowledge, the issue of longitudinal segmentation for the follow-up of multiple organs in whole-body MRI has never been studied.

A classical strategy for longitudinal studies is to register baseline exam on follow-up exam and report the registered baseline segmentation mask. We propose instead to use a segmentation pipeline that consists of manipulating the inputs of the baseline segmentation (e.g experts priors, segmentation initialization, etc.) of multiple organs to propagate them longitudinally to perform longitudinal segmentation. Hence, a new segmentation is performed instead of directly propagating segmentation mask on follow-up exams.

Countless techniques have been proposed in the literature for the automatic segmentation of medical images (McInerney and Terzopoulos 1996, Pham and Prince 2000, Sharma and Aggarwal 2010), most of them being dedicated to the segmentation of specific organs such as the brain or the heart. It is noticeable that their originality, compared with broad segmentation techniques, often deals with a prior knowledge about the anatomy of the structures to segment or their representation in the acquired images. Techniques that deal with a simultaneous segmentation of multiple organs on MRI can be subdivided into 3 main

categories: (1) Atlas-based methods have been proposed, such as patch-based multi-atlas label fusion (Bai *et al* 2013) for cardiac segmentation, keypoints transfer segmentation (Wachinger *et al* 2018) and label transfer segmentation (Lavdas *et al* 2017) for abdominal and thoracic organs. These approaches require the definition of multiple atlases and can also involve a computationally expensive registration process of each new image on reference database. (2) A second category is model-based approaches. Among them, (Chen and Bagci 2011) combines the multi-object recognition with an iterative trained graph-cuts with active shape model for abdominal organs in CT and bones in MR. In (Lay *et al* 2013), the authors proposed a shape model combining global and local context for the simultaneous segmentation of abdominal regions. Finally, statistical shape models were proposed in (Cerroloza *et al* 2016) for brain segmentation on MR images and abdominal organs on CT images. (3) Recent methods based on deep learning have been developed with Convolutional Neural Networks (CNN) for the 3D segmentation of liver and spleen on multi-contrast MR data of the body trunk (Küstner *et al* 2018), or with two-stage weighted-Fully Convolutional Neural Network (FCN) using multi-atlas spatial priors and auto-context for small organs segmentation (Valindria *et al* 2018). Another study proposed to consider dual pathway CNN to process multiple levels of resolution simultaneously for abdominal organs segmentation (Lavdas *et al* 2017). Although deep learning approaches are competitive, such techniques require an adequate training database or multiple atlases/models (for each organ), which are not always available in clinical applications. In our context, the potential organ variability in terms of anatomy, shape, and locations all over the body also makes such approaches inappropriate.

So, in this paper, we focus on segmentation methods where manual interventions by experts are required due to this strong anatomical variability of the segmented structures from one image to the other. In clinical practice, it is first common to achieve this by manually segmenting regularly-sampled 2D slices out of the 3D images and then interpolating the segmentations over the remaining slices. This straightforward method is however particularly time-consuming. Hence, we consider interactive segmentation approaches for the proposed pipeline.

In this context, an important and pioneering family of algorithms performing interactive segmentation algorithms are those based on graph-cuts (Boykov and Jolly 2001). These methods quickly minimize binary Markov Random Fields (MRF) to segment the images (Kolmogorov and Zabini 2004). They provide excellent results and have linear complexity. However, despite accelerations (Li *et al* 2004, Lermé *et al* 2010), they still require too much resources in terms of memory for applications in 3D medical imaging (Lermé *et al* 2010). Let's notice that methods based on MRF

have also been tackled using Belief Propagation (Wang and Cohen 2005). Random walk algorithms (Grady 2006) propose another way for propagating the seeds information. It is however relatively demanding in terms of computations, although it shows a very good capability to limit the contamination of the seed information to neighboring regions connected to the seed by narrow bridges. Finally, a third type of very fast algorithms is the Fast Marching (FM) (Sethian 1996). The FM quickly computes the distance between any voxel and the nearest seed point and relies on the notion of distance between neighboring voxels (Falcão *et al* 2004, Bai and Sapiro 2009). The latter can be mostly freely designed. The FM algorithm has already been used for segmentation tasks in CT, combined with contours refinement strategies (such as shape model fitting (Losnegård *et al* 2010), geodesic voting (Zuluaga *et al* 2014)) and for tumour growth modeling in brain MRI to estimate tumour density (Unkelbach *et al* 2014).

A potential issue with the FM is that it tends to propagate the segmentation to neighboring regions connected to the seed with similar intensities, in particular by narrow paths (contrary to random walk algorithms). This leaking issue led us to propose in (Risser *et al* 2018) a semi-automatic multi-label Fast Marching (FM) segmentation method optimized with a regularization term that penalized the crossing of narrow bridges. We have shown in (Risser *et al* 2018) a benefit of this regularized Fast Marching (rFM), in terms of processing time, when segmenting whole body MR images of patients with CLL.

In this article, we propose to build our pipeline extending (Risser *et al* 2018) to fully consider longitudinal segmentation. The proposed workflow corresponds to using the rFM segmentation for the baseline image, and then to benefit from the pre-defined seeds by automatically transporting and selecting the relevant ones for follow-up images. Note that the concept of the proposed pipeline is generic enough to be applicable with any of the pre-mentioned interactive segmentation method. In addition to the proposed workflow, the main contributions of this paper are then to describe how the FM seed points are transported and selected in the longitudinal regularized Fast Marching (LrFM) pipeline. Further evaluation tests and discussions are also given. We also slightly updated the baseline algorithm of (Risser *et al* 2018) to compute simultaneously the distances of all seeds labels, as explained in (Falcão *et al* 2004).

The article is organized as follows. Section 2 of this paper focuses on the description of the regularized multi-label FM segmentation methodology (rFM). Section 3 describes the proposed longitudinal pipeline (LrFM) with the seeds transformation and selection workflow for the follow-up segmentations. Both the proposed semi-automatic rFM segmentation and LrFM pipeline are validated on 3D synthetic data and a clinical dataset composed of 3D whole-body MR

images in section 4. Finally, we discuss the results in section 5, and conclude in section 6.

2. Regularized multi-label fast marching segmentation

2.1. Introduction

The goal of the previously proposed regularized multi-label Fast Marching (rFM) strategy (Risser *et al* 2018) is to define L regions in an image I , each region corresponding to an organ of interest identified with a label R_l , $l = 1, \dots, L$. As in standard fast-marching (FM), it consists in propagating a list of seeds until a maximum distance to the seeds is reached. For each considered region R_l , we denote by S_l the list of these point seeds.

The key idea of rFM was to incorporate local intensity changes into this distance.

Therefore, an additional regularization term taking into account region-wise intensity changes is also integrated in rFM.

The multi-label propagation algorithm is described in SubSection 2.2 and the regularization term, which is the cornerstone of this paper, is developed in section 2.3. We also present in section 2.4 the use of rFM to segment a single medical image in a clinical context. Extension to the segmentation of follow-up images is developed in section 3.

2.2. Multi-label fast marching

Algorithm 1. Simultaneous Multi-label Fast-Marching

Require: Image I and the seeds S_l , $\forall l = 0, \dots, L$
Ensure: Segmentation P and global min distance map d
 Initialize $P \equiv 0$ and $d \equiv +\infty$
 Initialize regularizing term R used in the cost c
For $l = 1 \rightarrow L$ **do**
 Initialize $mask_l$
 Initialize $d(p) = 0$ and $P(p) = l$, $\forall p \in S_l$ Initialize $Q = Q \cup \{S_l\}$
end for
while $Q \neq \emptyset$ **do**
 Take $p \in \arg \min_{q \in Q} d(q)$
 $l = P(p)$
for $q \sim p$ with $q \notin Q$ **do**
if $q \in mask_l$ **then**
 $dist = d(p) + c(p, q)$
if $(dist \leq d(q))$ and $(d(q) \neq +\infty)$ **then**
 $Q.remove(q)$
end if
if $(dist \leq d(q))$ and $(dist \leq thresh)$ **then**
 $d(q) = dist$
 $P(q) = P(p)$
 $Q = Q \cup \{q\}$
end if
end if
end for
end while

In order to define the multi-label FM segmentation, we assume that any pair of neighboring voxels $p \sim q$ in the image I is endowed with a non-negative cost $c(p, q)$. This cost is the most important modeling element of the proposed segmentation algorithm.

Considering the cost function, we define the length of any path $l(p = p_1 - p_2 - \dots - p_K = q)$ between voxels p and q as the sum of the cost of each neighboring pair $p_k - p_{k+1}$ from this path:

$$l(p_1 - p_2 - \dots - p_K) = \sum_{k=1}^{K-1} c(p_k, p_{k+1}) \quad (1)$$

The geodesic distance $dist(p, q)$ between any pair of voxels p and q is then defined as the length of the shortest path going from p to q . The simultaneous multi-label FM, based on the IFT algorithm (Falcão *et al* 2004), computes in a single distance map the geodesic distance from each voxel to the closest seed, considering simultaneously seeds from all labels. We denote by d this global distance map to the union of all seeds:

$$d(p) = \min_{q \in \bigcup S_l} dist(p, q), \forall p \quad (2)$$

Given these notations and definitions, a Voronoi diagram is such that all voxels that are associated with the same closest label l define the Voronoi cell R_l . We also denote by P the image describing the partition defined by the regions R_l

$$P(p) = l(\arg \min_{q \in \bigcup S_l} dist(p, q)), \forall p \quad (3)$$

where $l(q)$ is the function giving the label of seed q , $l = 1 \dots L$. Thus, we set $P(p) = l \Leftrightarrow p \in R_l$. For each label l , we restrict the computation of $d(p)$ to voxels in a large rectangular regions $mask_l$ surrounding S_l extended from a geodesic distance $thresh$, which limits the computational burden. We also impose $d(p) = +\infty$ at any voxel p further away than $thresh$ from the seed S_l . The simultaneous multi-label Fast-Marching algorithm is described in algorithm 1.

2.3. Regularizing cost

The fast marching method is particularly sensitive to narrow bridges between two organs with similar intensities. Without any regularization, the seeds belonging to different organs or the background must be equidistant from the expected boundary to avoid one region to spread on the other side of the bridge. This makes the segmentation result very sensitive to seed placement.

In order to avoid using the equidistance constraint for narrow bridges, we introduce the regularizing cost:

$$c(p, q) = \sqrt{(I(p) - I(q))^2 + \gamma R(p)^2}, \quad (4)$$

where the regularization map R penalizes bridge crossing, and γ is its weight. Given a structuring element \mathcal{N} , the cost R is defined by

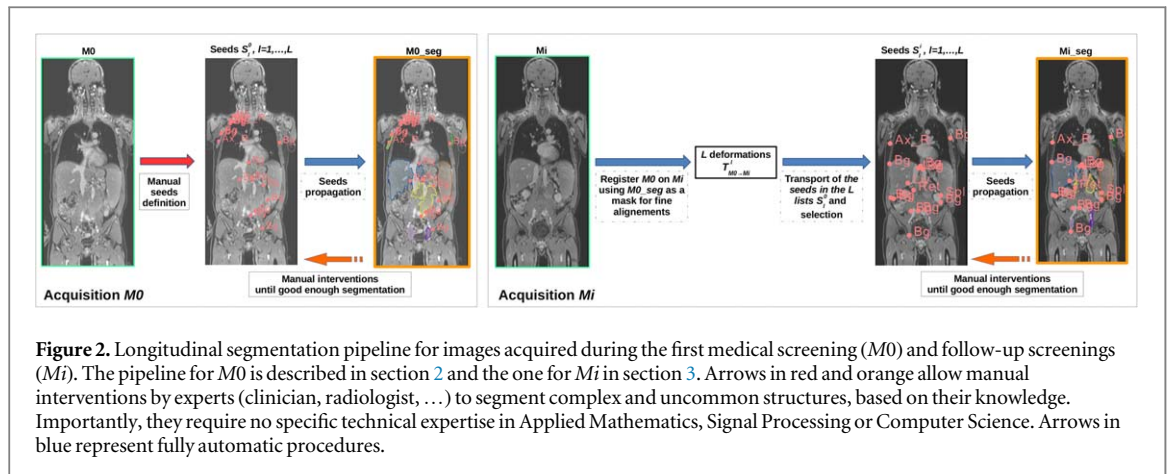


Figure 2. Longitudinal segmentation pipeline for images acquired during the first medical screening (M_0) and follow-up screenings (M_i). The pipeline for M_0 is described in section 2 and the one for M_i in section 3. Arrows in red and orange allow manual interventions by experts (clinician, radiologist, ...) to segment complex and uncommon structures, based on their knowledge. Importantly, they require no specific technical expertise in Applied Mathematics, Signal Processing or Computer Science. Arrows in blue represent fully automatic procedures.

$$R(p) = \max_{q \in \mathcal{N}} I(p + q) - \min_{q \in \mathcal{N}} I(p + q), \forall p \quad (5)$$

Bridges that are thinner than \mathcal{N} are then penalized by R . It is interesting to remark that $\max_{q \in \mathcal{N}} I(p + q)$ and $\min_{q \in \mathcal{N}} I(p + q)$ are a dilation and an erosion of I with structuring element \mathcal{N} , respectively. Hence, γ weights the distance allowed between local extremas in the neighborhood \mathcal{N} of the voxel. For algorithmic reasons, we recommend to model \mathcal{N} with a cuboidal structuring element (s.e.) of size $r_x \times r_y \times r_z$. This indeed makes it possible to split the 3D dilations/erosions into a sequence of 1D dilations/erosions with first a $r_x \times 1 \times 1$ s.e., then a $1 \times r_y \times 1$ s.e., and finally a $1 \times 1 \times r_z$ s.e., for exactly the same result. For each voxel of I , the algorithmic cost of a dilation or erosion is then $2(r_x + r_y + r_z) + 3$ instead of $(2r_x + 1) \times (2r_y + 1) \times (2r_z + 1)$.

2.4. Application of rFM in a clinical context

As mentioned in introduction, the motivation of this work is to make the semi-interactive segmentation of complex 3D structures as efficient as possible in a clinical context. To segment a single image, the rFM-based segmentation procedure is then as follows. For each considered region R_l , the users (clinician, expert, ...) first define a list of point seeds S_l . The rFM propagation algorithm is then run to segment I based on these seeds. The resulting segmentation is displayed to the user. Based on their expertise they can improve the segmentation by adding, removing or moving seeds and then re-running the propagation algorithm. This iterative process can be repeated multiple times until a satisfactory visual segmentation is reached according to the expert.

This pipeline opens a first practical remark, the fact that there are bridges between different regions to segment motivated the regularization strategy of section 2.3. There may be also bridges between a structure to segment and another one which has not to be segmented. To address this case, we consider an additional background region denoted R_0 and use a list of seeds S_0 when running the rFM propagation. This set gathers the background seeds associated to all

considered regions. The resulting segmentation of region R_0 is obviously not displayed to the users.

Another remark deals with the estimation of the regularization map R in equation (5). This map depends on the image I and the structuring element \mathcal{N} but not on the seeds. It can therefore be computed once for all before the first seeds propagation and re-used later, with an algorithmic cost of M times $2(D_x + D_y + D_z) + 3$, where M is the number of voxels of I , and (D_x, D_y, D_z) the dimensions of I . Once computed, the computational cost of the seeds propagation is then almost the same with or without regularization, which is clearly a plus from a user experience point of view. The computational complexity of algorithm 1 is then $2^* N_L \log(N_L)$, with $N_L = \sum_l N_l$ and N_l the number of voxels in $mask_l$ of organ l . Indeed, the complexity is $N_l \log(N_l)$ for the heap data structure construction (points storage) plus $N_l \log(N_l)$ for FM propagation in the worst case. Finally, global minimum distance and associated label are obtained in M times $o(1)$.

3. Longitudinal segmentation pipeline

3.1. Segmentation pipeline

We have presented in section 2 the regularized FM segmentation (rFM) method and its application to the segmentation of a single image in a clinical context. We now propose to extend the rFM method to the longitudinal case (LrFM), where several follow-up images of the same patient are segmented. In what follows, M_0 refers to the first acquired image and M_i refers to the i^{th} follow-up image, with $i \geq 1$. The segmentation of image M_0 follows the procedure of section 2 and we denote S_l^0 the manually defined seeds used to segment region l in image M_0 . The methodological contribution of this section then deals with how taking advantage of the information in image M_0 and the seeds of the S_l^0 to segment a follow-up image M_i with as little manual interactions as possible.

The segmentation pipeline is summarized in figure 2. After segmenting M_0 with the procedure of section 2, the seeds of the list S_l^0 are first transported

using a displacement field obtained by registering $M0$ on Mi , as developed in section 3.2. A subset of the transported seeds is then selected using the procedure of section 3.3. The user finally iteratively refines the seeds as in section 2.4 and re-runs segmentation. The benefit of this procedure is therefore to alleviate the need for manual interventions with an automatic definition of initial seeds.

3.2. Longitudinal Seeds propagation

As illustrated in figure 2, the seeds are propagated based on a displacement field $T_{M0 \rightarrow Mi}$ which is computed by registering $M0$ on Mi . Various registration techniques could be applied depending on the application. In our context, i.e. the segmentation of several organs with a potentially high anatomical variability across follow-up images, we use a two-steps registration technique.

3.2.1. Global alignment

A global alignment of the two volumes within the same coordinate system is first performed to coarsely align the body envelopes. To do so, an affine deformation model was used. In our application, we maximized the Mattes Mutual information (MMI) (Mattes *et al* 2003) between $M0$ and the deformation of Mi with $T_{M0 \rightarrow Mi}^g$.

3.2.2. Region-wise registration

Finer registration is then performed for each of the L segmented regions R_l . This task is particularly challenging as the environment of the segmented organs is heterogeneously deformed in space due to respiratory movements, patients positioning, and the strong variability of the anatomical regions and of their deformation. It therefore appears as reasonable to solve this registration problem using a strongly constrained deformation model. As developed in A.2.2, we first tested BSpline and affine registration in this context, which led to equivalent segmentation results. As affine registration was faster, we later used it to specifically register each organ of interest. For organ l , the deformation $T_{M0 \rightarrow Mi}^l$ is then computed as follows. The segmentation computed in first place for organ l at $M0$ is dilated. This dilated segmentation will mask the region-wise registration in the $M0$ image domain, which enables us to focus on the region of interest and its local neighborhood only. The affine registration is then performed by initiating $T_{M0 \rightarrow Mi}^l$ as equal to $T_{M0 \rightarrow Mi}^g$. In our application, the circular structuring element for the dilation has a radius of 10 mm and the normalized cross-correlation metric (NCC) (Avants *et al* 2008) was maximized during the registration. We adopted at each step of the registration process the metrics with the best segmentation performances among the three tested (SSD, MMI, NCC).

3.2.3. Region-wise seeds propagation

We recall that the seeds used to segment region l in $M0$ are contained in a list S_l^0 . The initial list of seeds S_l^i to segment region l in Mi is then $T_{M0 \rightarrow Mi}^l(S_l^0)$. Note that a list of *background seeds* S_0^0 in $M0$ also exists, as mentioned in section 2.4. When propagating the seeds S_l^0 on Mi , we also propagate the subset of seeds in S_0^0 for which the nearest non-background seed is in S_l^0 .

3.3. Seeds selection

It is important to emphasize that the seeds of S_l^i used to segment region l in Mi were propagated using an affine deformation model. The propagation technique described in section 3.2 is likely to propagate some seeds of S_l^i outside the region to segment in Mi . Although these undesired seeds can be simply erased by the user (see section 2.4), we then propose to minimize this burden with an automatic seeds selection procedure.

We denote μ and σ the mean and the standard deviation of the intensities at the location of the seeds S_l^i in Mi . We make the hypothesis that organ l has homogeneous intensities and most seeds of S_l^i are properly transported in its spatial domain. The seeds with intensities at Mi outside of the interval $[\mu - 1.25^* \sigma, \mu + 1.25^* \sigma]$ are then considered as outliers and discarded from S_l^i . Therefore, the seeds which are obviously propagated outside of organ l do not have to be manually erased. As for background seeds, they are considered into subgroups defined from their nearest organ.

3.4. Implementation details

The segmentation and registration procedures are performed using 3D Slicer. In particular, our rFM segmentation algorithm is integrated to 3D Slicer (Fedorov *et al* 2012) as a C++ loadable module. The registration is based on Slicer BRAINS registration module (Johnson *et al* 2007). The fiducial editor of 3D Slicer is used for the seeds definition. The seeds transformation is processed with the SimpleITK toolkit⁴.

4. Experiments and results

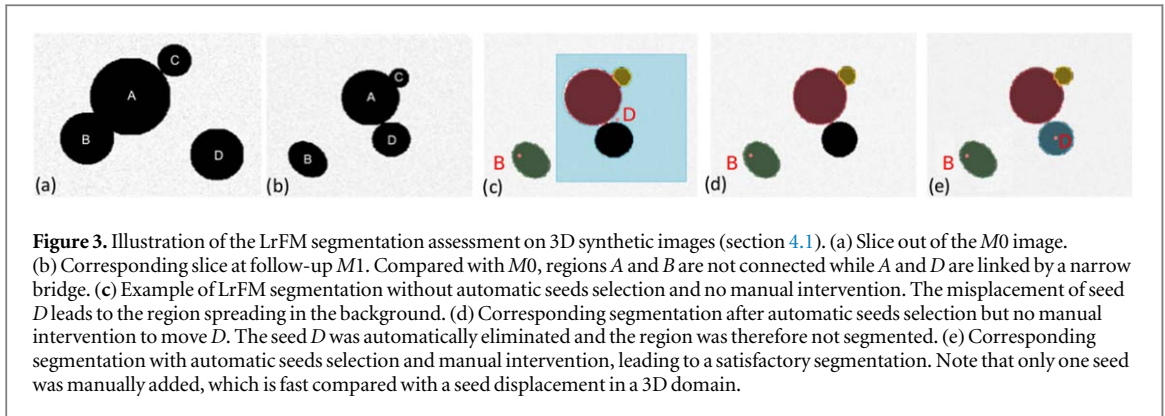
We now evaluate the LrFM pipeline on both 3D synthetic (section 4.1) and clinical data (section 4.2).

4.1. Results on 3D synthetic data

4.1.1. Methodology

We validate hereafter the LrFM segmentation pipeline on 3D synthetic data by evaluating its robustness towards both the seeds location and the variability of structures deformation. To do so, we first defined a 3D synthetic image composed of four spherical regions, denoted A , B , C and D , modelling tumorous regions

⁴ www.simpleitk.org/



which present a lot of variability along follow-up, as illustrated between figure 3-(a) and -(b). This image represents the information at time M_0 and is designed as follows. Region A is surrounded by both regions B and C , linked by a large bridge and a narrow bridge respectively. Region D is not connected to any other regions. The 3D binary synthetic image was slightly deteriorated with a Gaussian noise, to simulate a blur e.g. due to breathing motion. In order to later assess the stability of our method, 10 seed sets were defined on this M_0 image. One seed was randomly placed in each region B , C and D , whereas a random amount of one to four seeds were randomly placed in region A . Note that the evaluation of the rFM strategy on 3D synthetic images at M_0 is given in appendix A.1 and leads to very similar results as those obtained with 2D synthetic images in (Risser *et al* 2018).

We then transformed these regions to model their evolution from M_0 to M_1 . In order to evaluate LrFM pipeline robustness to different levels of deformation, we applied a specific transform to each region, using distinct zoom and shear parameters. In this transformed image (see figure 3(b)), regions A and B are no longer connected while regions A and D are now linked with a narrow bridge, and A and C by a large bridge. The transformed synthetic image was also slightly deteriorated with a Gaussian noise.

We then assessed the LrFM segmentation pipeline without manual correction of the seeds on this 3D synthetic image, following the subsequent steps. We first registered M_0 image on M_1 image, using our two-steps registration procedure described in section 3.2. For these synthetic data, the Mean Square Error was used as the registration metric, since the regions intensity were comparable between M_0 and M_1 images. This step defined a distinct local transform $T_{M_0 \rightarrow M_1}^l$ for each region. The 10 distinct sets of seeds, defined on the M_0 image, were then transported on the M_1 image, using $T_{M_0 \rightarrow M_1}^l$ accordingly. Automatic seeds selection was finally performed.

Figure 3 illustrates an example of the LrFM segmentation pipeline on the 3D synthetic data at M_1 follow-up. It displays the different segmentations obtained after each step of the pipeline. We compared the result obtained with the LrFM segmentation

pipeline with the discrete sphere volume of each transformed region, using the local Dice coefficient (DSC). The DSC quantifies the overlap agreement between two segmentations; it varies from 0 (i.e. no overlap) to 1 (i.e. complete overlap). This computation is performed locally, i.e. for each region specifically. The local DSC, obtained for each region with LrFM segmentation at M_1 with and without regularization, are summarized in the boxplots in figure 4. The boxplots were computed after seeds selection, no manual correction was performed to replace inaccurate seeds or add missing ones.

4.1.2. Results

On average, only 13.8% of the seeds fell outside the regions of interest after the registration and the LrFM pipeline segmented of the different regions with a good accuracy. The performances of the FM-based segmentation pipeline are related to the presence of bridges of similar intensity and the position of seeds around them. Our pipeline demonstrated a better robustness towards these factors with the regularization. Mean local DSC reached 0.95 ± 0.06 and 0.92 ± 0.11 over all four regions with and without regularization, respectively. Median local DSC increased from 0.94 without regularization to 0.98 with regularization. In addition, the mean minimum local DSC was 0.80 over the four regions with regularization versus 0.66 without regularization.

The regularization therefore leads to more stable and accurate segmentation results with a smaller variability for regions that are connected by a narrow bridge, such as regions A and D here. Indeed, the proposed regularization enables to limit region leaking across the bridge, improving local dice coefficients. Since region A is the largest region transformed with the smoothest deformation, possibly one to four seeds fell inside this region, and thus, if one seed (or more) is eliminated during the selection, there are potentially other remaining seeds to conduct a correct segmentation. Hence, mean local dice of A increases from 0.96 ± 0.04 without regularization to 0.98 ± 0.03 with regularization. Compared to A , region D presents a lower variability. Indeed, the propagation of D into A is not equally widespread due to the potential larger

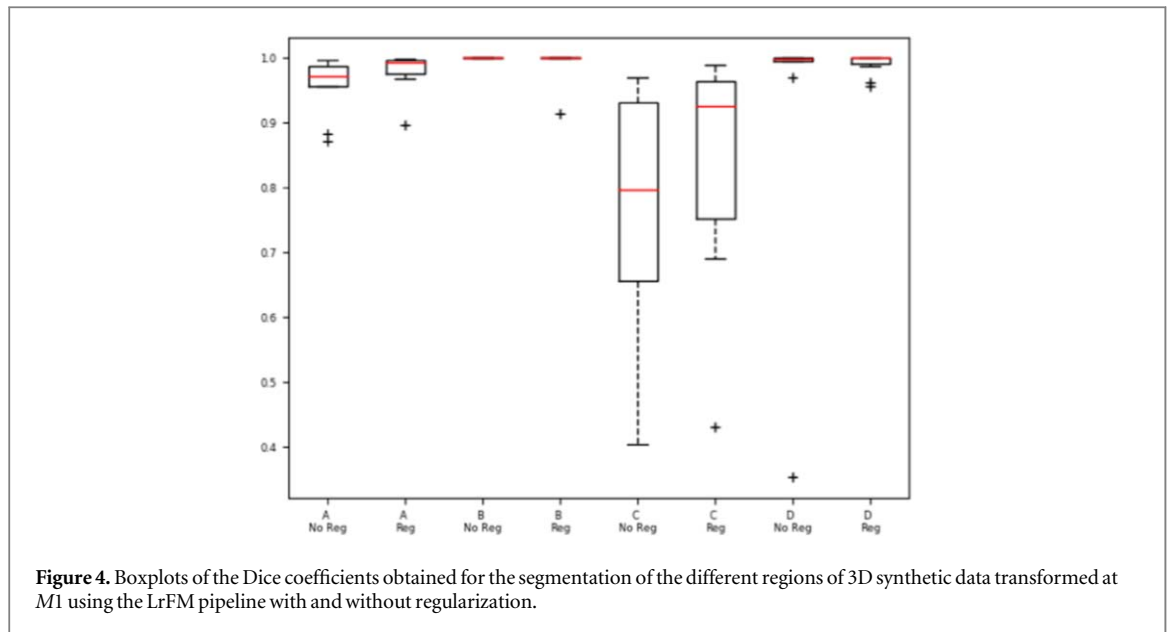


Figure 4. Boxplots of the Dice coefficients obtained for the segmentation of the different regions of 3D synthetic data transformed at $M1$ using the LrFM pipeline with and without regularization.

number of seeds remaining in A . Mean local DSC of D increases as well from 0.93 ± 0.20 to 0.99 ± 0.02 with the regularization.

The regularization also allows to slightly improve segmentation results within regions that are connected by a larger bridge, such as regions A and C . The FM algorithm requires the presence of equidistant seeds from the expected border to limit the propagation in either region. Since this condition is not always respected, the level of spreading of region C into region A depends on the position of seeds in A around the border, and the regularization contributes to slightly constrain this propagation, mean DSC of C increases from 0.78 ± 0.18 to 0.84 ± 0.18 .

Region B is always well segmented in this transformed image with a median dice of 1.0 with and without regularization. Only an outlier Dice of 0.9 was observed (see Boxplot 4).

Importantly, we finally emphasize that we obtained similar segmentation accuracies by using directly manually placed seeds (as for the rFM segmentation at $M0$) or by re-using and transporting the already-defined seeds for a new follow-up image (as for the LrFM pipeline). This shows the relevance of our pipeline to reduce user-interactions in longitudinal studies.

4.2. Results on clinical data

In (Risser *et al* 2018), we have shown that the rFM segmentation of clinical data was more robust and more accurate with the proposed FM regularization, compared to the complete manual delineation of regions. In this article, we demonstrate that, when considering longitudinal segmentation, the propagation of seeds is very relevant to reduce user-interaction and limit the time dedicated to the segmentation task, compared to directly applying the rFM procedure on follow-up exams. We also show that the transformation of seeds

appears more robust and accurate than the transformation of pre-defined baseline segmentation masks, especially with the regularization which is very tractable to handle regions deformation.

4.2.1. Dataset

The LrFM segmentation technique was validated on a clinical dataset of 10 consecutive MR images of patients with CLL using $M0/M1$ data. Axial Volumetric Interpolated Breath-hold Examination (VIBE) MR images using Dixon fat-water separation were acquired on a 1.5 Tesla scan using one head coil, two 18-channel design body flex coils and one 36-channel design peripheral angio coil used in combination of the 32-channel design spine coil in order to perform whole-body imaging with the following parameter settings TR/TE = 6.76/2.39 ms and Flip Angle = 10° . Typical image size is $250 \times 250 \times 200$ voxels for a resolution of $1.5 \times 1.5 \times 4.0$ cubic millimeter per voxel. All images were pre-processed with bias field correction and a grey-level homogenization on whole-body MR composing. An histogram-matching of $M1$ exam on $M0$ was also performed. Segmentation parameters γ and $thresh$ were set empirically at first. γ was fixed to 0.0025 and $thresh$, defined from the diameter of the largest organ, was equal to 100 mm. They were not modified during the experiments.

4.2.2. Methodology

The semi-automatic rFM segmentation was validated at $M0$ for the delineation of 12 anatomical structures disseminated over the whole-body MRI, namely the cervical, axillary, inguinal, iliac, mediastinum and retroperitoneum lymph nodes along with liver and spleen organs. These regions present a large anatomical variability, and not always clear anatomical contours, thus they require interactive segmentation. MR images at $M0$ were used to assess our rFM method

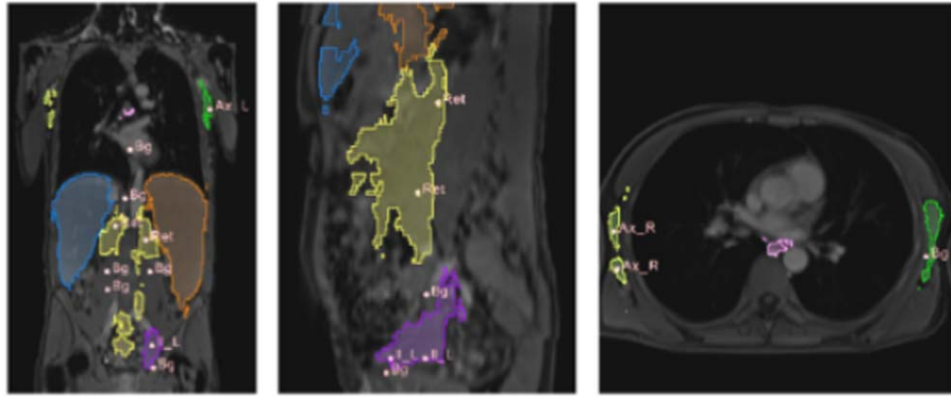


Figure 5. Segmentation obtained on a whole-body MR image at M_0 by using the rFM approach, as already shown in (Risser *et al* 2018).

stability, and compared to the manual delineation performed by clinical experts. Figure 5 illustrates a rFM segmentation obtained on one of these whole-body MR exams, as already shown in (Risser *et al* 2018). Extensions of the quantitative results of (Risser *et al* 2018) for the rFM segmentation are given in A.2.1. The resulting sets of seeds defined on M_0 exams are the initial material to evaluate in this section the accuracy and the stability of the LrFM pipeline on the same regions in the M_1 exam.

Different variants of the LrFM pipeline are assessed to quantify the impact of its different steps: (S_{All}) LrFM segmentation using all seeds, i.e. without automatic seeds selection and no manual intervention. (S_{Sel}) LrFM segmentation after the automatic selection of relevant seeds but no manual intervention. (S_{CorReg}) Full LrFM pipeline, as shown in figure 2. ($S_{CorNoReg}$) Same as (S_{CorReg}) but with no regularization during the Fast Marching propagation. We also denote by (S_{Mask}) the propagation of the rFM segmentation masks obtained at M_0 on the M_1 exam, after region-specific registration with $T_{M_0 \rightarrow M_1}^l$. These different tests are compared to (S_{rFM}) rFM segmentation directly computed on M_1 exam, considered as the reference hereafter.

The agreement between our LrFM segmentation results and reference segmentations (here, S_{rFM}), is then quantified by considering seven metrics, classically used to evaluate the performances of a segmentation algorithm (Lavdas *et al* 2017): (**Global and Local DSC**) The dice coefficients, already introduced in section 4.1, are computed globally and locally. Global computations consider all voxels over the whole-body images and local computations focus on specific organs of interest. (**RE**) The Recall corresponds to the sensitivity (i.e. the true positive rate) and reaches 1 when all voxels of the reference are correctly retrieved in the segmentation. (**PR**) The Precision is the positive predictive value, it illustrates how many of the segmented voxels are actually positive in the reference (it ranges from 0 (no true positive) to 1 (no false positive)). (**HD**) The Hausdorff Distance is the maximal distance

from a voxel in the segmentation to a nearest voxel in the reference segmentation. (**MSD**) The Mean symmetric Surface Distance is the average of all the distances between the boundary voxels of the evaluated segmented region and the nearest voxel on the boundary of the reference region. (**RMSSD**) The Root Mean Square symmetric Surface Distance is similar to the MSD but considers the squared distances between the two sets of boundaries, the root of the squared distances leads to the RMSSD. For both the MSD and the RMSSD, a value of 0 corresponds to a perfect overlap between the two segmentations. These three distances are expressed in millimeters and are unbounded. Here distance metrics are only computed locally.

Note that the evaluation metrics are averaged on a set of 9 patients out of the 10 original ones. Indeed, one patient did not performed correctly breath holding during MR acquisition at M_0 , which led to poor image quality with a strong movement artifact which induces poor segmentations.

4.2.3. Results

LrFM Segmentation evaluation metrics are described in table 1. For each considered metric, all tested segmentation strategies were shown to perform with significantly different accuracies ($p < 0.0005$, Kruskal-Wallis test).

4.2.3.1. LrFM versus S_{Mask} :

Results show that registering M_0 seeds to perform the segmentation at M_1 globally offers better performances compared to registering the M_0 segmentation mask (S_{Mask}). Indeed, considering S_{rFM} at M_1 as reference, our S_{CorReg} strategy was the most accurate strategy (mean global DSC of 0.93 ± 0.01 , mean local DSC of 0.81 ± 0.05). In comparison, S_{Mask} presented a mean global DSC of 0.79 ± 0.05 and a mean local DSC of 0.43 ± 0.12 . This means that the organs are better segmented regardless of their volumes. Our S_{CorReg} strategy enables to segment organs of interest with better precision and better sensitivity compared to S_{Mask} (mean local PR of 0.78 ± 0.06 , compared to

Table 1. Evaluation of the different segmentation strategies on 9 patients of the CLL cohort. Results are presented for different variants of the proposed LrFM pipeline (S_{All} , S_{Sel} , S_{CorReg} and $S_{CorNoReg}$), and compared with S_{Mask} , considering S_{rFM} at M1 as reference, as explained in section 4.2. Evaluation metrics are Global and local DSC, precision (PR), recall (RE), Hausdorff distance (HD), mean surface distance (MSD), root mean square surface distance (RMSSD). The last line indicates the average number of seeds required for each patient, in comparison to the 386 seeds used by S_{rFM} at M1. The first line indicates the maximum p-value obtained with the Mann-Whitney statistical test for each considered segmentation compared to S_{CorReg} over the 7 evaluation metrics. The last column gives the p-value according to the Kruskal-Wallis statistical test comparing altogether the 5 segmentation techniques for each of the 7 evaluation metrics.

	S_{rFM} at M1 (reference)	S_{Mask}	LrFM pipeline with transformed seeds				<i>p-value</i> Kruskal-Wallis
			S_{All}	S_{Sel}	S_{CorReg}	$S_{CorNoReg}$	
<i>p-value</i> Mann-Whitney	—	≤ 0.001	≤ 0.001	≤ 0.009	—	≤ 0.022	
Global DSC	—	0.79 ± 0.05	0.86 ± 0.03	0.86 ± 0.03	0.93 ± 0.01	0.90 ± 0.03	< 0.0001
Local DSC	—	0.43 ± 0.12	0.56 ± 0.09	0.59 ± 0.14	0.81 ± 0.05	0.72 ± 0.09	< 0.0001
Local PR	—	0.43 ± 0.13	0.53 ± 0.06	0.61 ± 0.14	0.78 ± 0.06	0.67 ± 0.09	< 0.0001
Local RE	—	0.50 ± 0.13	0.66 ± 0.11	0.63 ± 0.16	0.84 ± 0.04	0.81 ± 0.09	< 0.0001
Local HD	—	38.89 ± 9.77	42.12 ± 10.54	44.24 ± 6.70	23.03 ± 5.17	39.45 ± 9.79	< 0.0005
Local MSD	—	7.21 ± 4.07	6.88 ± 3.02	7.17 ± 3.30	1.48 ± 0.81	2.79 ± 1.24	< 0.0001
Local RMSSD	—	10.18 ± 4.75	10.45 ± 3.80	10.77 ± 3.90	2.92 ± 1.16	5.65 ± 1.79	< 0.0001
Nb of seeds	386	—	524	331	401	401	—

0.42 ± 0.13 for S_{Mask} , and mean local RE 0.84 ± 0.04 , which was equal to 0.50 ± 0.13 for S_{Mask}). S_{CorReg} also showed a good robustness by obtaining the lowest distance metrics with low variability. It showed a mean local HD of 23.03 ± 5.17 , a mean MSD of 1.48 ± 0.81 and a mean local RMSSD of 2.92 ± 1.16 whereas these metrics reached 38.89 ± 9.77 , 7.21 ± 4.07 and 10.18 ± 4.75 respectively for the S_{Mask} strategy. Particularly, S_{CorReg} was shown significantly different from all the other strategies for all metrics ($p < 0.022$, Mann-Whitney test).

4.2.3.2. Number of seeds and time gain:

The LrFM pipeline also allowed to reduce user-interaction for the longitudinal segmentation task compared to S_{rFM} . Indeed, the set of seeds automatically registered on M1 is composed of 524 seeds per patient in average. The seeds selection reduces this set to 331 seeds (-37%) in average. The manual correction, which is the only step of the pipeline requiring user interaction, leads to the addition of 75 new seeds in average per patient. In comparison, S_{rFM} at M1 requires a mean of 386 seeds. Hence, the longitudinal pipeline leads to a reproducible segmentation result with a reduction of 80% of manually corrected seeds (deleted or displaced). Besides, this manual correction takes 36 minutes in average on the database of 9 patients, whereas it takes 50 minutes to define all seeds directly on M1 and segment it. The majority of the corrected seeds concerned the background. For the smallest organs (e.g. axillary, inguinal and cervical lymph nodes), the transported seeds were more likely to fall outside the region to segment (because of nodes shrinkage), so the manual correction consisted in repositioning them. For the biggest organs, the robustness of the segmentation required less correction, the

minimal manual intervention essentially consisted in adding background seeds.

4.2.3.3. Interest of the seeds selection:

The seeds selection, leading to S_{Sel} , increases the accuracy of the LrFM segmentation compared to S_{All} . The mean global DSC remains stable (0.86 ± 0.03) but the mean local DSC increases from 0.55 ± 0.09 to 0.59 ± 0.14 for S_{Sel} , showing that the seeds selection mainly improves the small organs contours. S_{Sel} leads to a more accurate segmentation with a mean local PR reaching 0.61 ± 0.14 when it was 0.53 ± 0.06 in S_{All} . On the contrary, the local RE decreases slightly from 0.66 ± 0.11 in S_{All} to 0.63 ± 0.16 for S_{Sel} . The detection of right positive regions is hence clearly improved in S_{Sel} , along with the increase of false negatives. Both strategies seems comparatively robust since they obtain similar mean and median distance metrics, however the contours variability is reduced in S_{Sel} . The seed selection led to a slight increase in mean distance metrics (MSD increases from 6.88 ± 3.02 to 7.17 ± 3.30 , HD from 42.12 ± 10.54 to 44.24 ± 6.70 and RMSSD from 10.45 ± 3.80 to 10.77 ± 3.90), which comes from the suppression of some relevant Background seeds useful to constrain segmentation; their removal created spreading with small volume that did not impact dice scores. Figure 6(a) shows an example of S_{Sel} which reduces potential false positive regions. Subsequent manual seeds correction S_{CorReg} enables to significantly refine the segmentation to obtain a satisfactory result (figure 6(b)).

4.2.3.4. Interest of the regularization:

The regularization allows the LrFM pipeline to be more accurate and robust to intensity bridges, mostly for weakly contrasted organs. Compared with $S_{CorNoReg}$, S_{CorReg} presents an increase of both mean

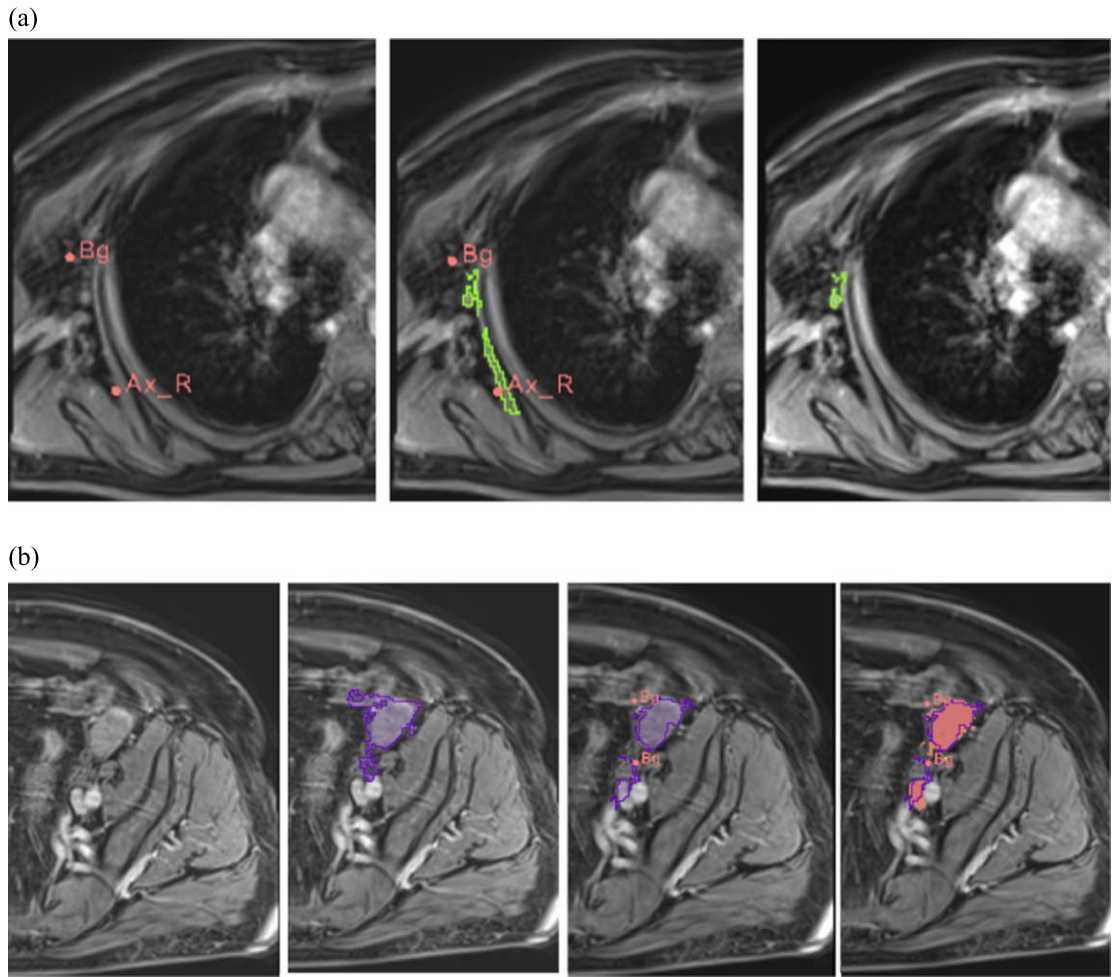


Figure 6. (a) Segmenting with all $M0$ seeds transported on $M1$ MRI (left) leads here to region spreading (middle) from misplaced seed (Ax_R) due to lymph node shrinkage. The automatic seeds selection helps removing outlier seeds (here, seed with intensity lower than the rest of Ax_R seeds). Segmentation result after seed selection (right). (b) The segmentation has spread in regions of similar intensities after seeds selection (second). It is then necessary to manually add two background seeds to constrain the propagation (third) and obtain a similar result as the ground truth superimposed in orange (fourth). Illustration of the interest of automatic seeds selection (a) and manual seeds correction (b).

global and local DSC from 0.90 ± 0.03 and 0.72 ± 0.09 to 0.93 ± 0.001 and 0.81 ± 0.05 , respectively. The local DSC being more affected than the global DSC shows that the smallest organs are the most impacted by the absence of regularization. Indeed, region spreading can quickly double the size of small volumes and reduce local DSC. The absence of regularization may lead to regions leaking through small bridges, and therefore we observe an overall increase of all distance metrics considered for $S_{CorNoReg}$. Hence, with regularization, all distance metrics decreased for S_{CorReg} , reducing segmentation leaks across intensity bridges (MSD decreased from 2.79 ± 1.24 for $S_{CorNoReg}$ to 1.48 ± 0.81 for S_{CorReg} , HD from 39.45 ± 9.79 to 23.03 ± 5.17 , and RMSSD from 5.65 ± 1.79 to 2.92 ± 1.16).

4.2.3.5. Local performances on the organs of interest:

The five segmentation strategies present significantly different local DSC for all the organs of interest ($p < 0.04$, Kruskal-Wallis test) except for the left

cervical node and the retroperitoneum (which were not enough represented in the patients set). The mean local organ DSC are summarized in table 2. Best results are obtained with our LrFM segmentation pipeline (S_{CorReg}) for all considered organs. S_{CorReg} is shown significantly different from all the other strategies ($p < 0.048$, Mann-Whitney test).

For the biggest organs with a good contrast (the liver and spleen), the local DSC results remain quite stable along the different steps of the LrFM pipeline as for S_{Mask} . Indeed, due to their very light deformation from $M0$ to $M1$ exam, very few seeds significantly needs to be eliminated or manually corrected.

Concerning the biggest organs with a weak contrast (e.g. the mediastinum and retroperitoneum), their evolution from $M0$ to $M1$ is better handled with the seed-based LrFM pipeline compared to S_{Mask} . For instance for the mediastinum, a mean local DSC of 0.32 ± 0.25 was obtained with S_{All} , versus 0.20 ± 0.16 with S_{Mask} . In these cases, the seed selection is beneficial to improve the local precision

Table 2. Local dice coefficients for each organ of interest averaged over the 9 patients for all segmentation strategies. The first row presents the maximum p-value obtained with the Mann-Whitney statistical test comparing each segmentation strategy with S_{CorReg} over all the organs of interest. The maximum p-values according to the Kruskal-Wallis statistical test comparing altogether all segmentation techniques considering each organ of interest are given in the last column.

<i>p-value Mann Whitney</i>	S_{Mask} ≤ 0.040	LrFM pipeline with transformed seeds				<i>p-value kruskal-Wallis</i>
		S_{All} ≤ 0.048	S_{Sel} ≤ 0.023	S_{CorReg} —	S_{CorNoReg} ≤ 0.046	
CerR	0.09 ± 0.09	0.25 ± 0.27	0.03	0.76 ± 0.06	0.52 ± 0.13	0.015
CerL	0.08 ± 0.10	0.19 ± 0.26	0.00	0.70 ± 0.10	0.42 ± 0.07	—
AxR	0.26 ± 0.16	0.49 ± 0.32	0.48 ± 0.33	0.78 ± 0.10	0.67 ± 0.12	0.008
AxL	0.24 ± 0.22	0.32 ± 0.24	0.36 ± 0.26	0.78 ± 0.10	0.74 ± 0.11	0.0
IngR	0.40 ± 0.09	0.63 ± 0.08	0.34 ± 0.06	0.82 ± 0.12	0.67 ± 0.17	0.001
IngL	0.32 ± 0.10	0.52 ± 0.16	0.46 ± 0.14	0.82 ± 0.08	0.73 ± 0.19	0.004
Iliar	0.29 ± 0.16	0.43 ± 0.18	0.44 ± 0.19	0.71 ± 0.11	0.59 ± 0.14	0.004
IliarL	0.30 ± 0.15	0.41 ± 0.19	0.39 ± 0.23	0.76 ± 0.07	0.65 ± 0.11	0.001
Med	0.20 ± 0.16	0.32 ± 0.25	0.37 ± 0.30	0.64 ± 0.15	0.47 ± 0.22	0.04
Retrop	0.47 ± 0.07	0.55 ± 0.18	0.62 ± 0.03	0.81 ± 0.06	0.79 ± 0.07	—
Liver	0.85 ± 0.05	0.90 ± 0.03	0.89 ± 0.04	0.93 ± 0.02	0.92 ± 0.02	0.003
Spleen	0.79 ± 0.06	0.89 ± 0.05	0.88 ± 0.06	0.94 ± 0.02	0.92 ± 0.02	0.0

(0.37 ± 0.30 with S_{Sel}) and the manual seeds correction remains necessary to obtain a satisfactory result (0.64 ± 0.15 with S_{CorReg}). Due to their low contrast, the regularization is very relevant for these regions, (0.47 ± 0.22 for S_{CorNoReg}).

Furthermore, the small organs with weak contrast (e.g. cervical lymph nodes), are the most complex to delineate. Their mean local DSC is very low ($< 0.10 \pm 0.10$) with S_{Mask} , because their large deformation is poorly handled. In comparison, S_{All} offers more robust results (0.20 ± 0.27). Due to their local variability and complex registration, the seeds selection is very drastic, and leads to the elimination of almost all seeds for cervical nodes. Therefore, the manual seeds correction is essential to add missing ones and obtain a correct segmentation (0.73 ± 0.08). In these regions, the regularization is essential due to the presence of multiple regions with similar intensities as that of cervical nodes (0.47 ± 0.22).

In comparison, for small regions with a better contrast (e.g. axillary or inguinal lymph nodes), S_{All} presents a higher accuracy than S_{Mask} thanks to the algorithm robustness towards seeds location. The seeds selection did not have much impact on these lymph nodes, with quite stable mean local DSC. The manual seeds corrections enables to achieve good delineation with high mean DSC. The regularization was less significant for these regions thanks to their better contrast.

5. Discussion

We have proposed a longitudinal segmentation framework, that relies on a semi-automatic regularized multi-label fast marching (rFM) segmentation to delineate organs of interest. This proposed rFM segmentation uses a regularization map, which takes advantage of both properties from the Fast Marching

(very fast) and regularized methods (robust to seed displacement). We compared at $M0$ the implemented segmentation algorithm with a software available in clinical routine, Iplan RT⁵, and the semi-automatic delineation outperformed the cumbersome manual task. This semi-automatic rFM approach thus enables to reduce both processing time and user-variability. In addition, the provided approach is sufficiently generic to be adopted for the delineation of any organ.

The seed-based extended longitudinal pipeline offered better longitudinal segmentation results compared to registration-based transformation of baseline rFM masks on $M1$. This can be due to the higher flexibility of our regularization model on seeds propagation, compared to the implicit spatial regularization required by locally rigid deformations. Indeed, the rFM segmentation process was shown to be robust enough to seeds location to handle smooth deformations. Thus, the use of seeds appears more tractable to handle heterogeneous spatial deformations of multiple levels without requiring very accurate longitudinal registration.

Besides, to cope with potential misplacement of organs seeds after longitudinal registration, we introduced a step of automatic seeds selection. This helped reducing the number of seeds to verify to only consider the relevant ones. This step also permitted to reduce regions spreading during FM corresponding to false positive regions. Although a final manual seeds correction remained necessary, it is worth noting that this automatic seeds registration allows for a considerable time gain of image analysis. Indeed, the manual seeds correction remains less time consuming than placing one by one the seeds for another direct semi-automatic rFM segmentation on Mi . As a perspective, the seeds selection can be more robust by identifying

⁵ www.brainlab.com

an organ with important shrinkage and automatically remove all its corresponding seeds.

To improve the robustness of the FM propagation, we introduced a regularized cost to penalize the crossing of narrow bridges during the FM propagation. It would be relevant to investigate this leaking problem more globally, notably by raising the distance metric to powers higher than two, in order to approximate the geodesic distance to the f_{\max} function as proposed in (Miranda and Falcão 2009, Couprie *et al* 2010).

More practically, the introduced rFM segmentation approach helped us draw a medical interpretation of CLL evolution on MR exams in particular in quantifying significant organs volume shrinkage at M1, and more patient-dependent evolution at M12. In this article, we focused our LrFM experiments on both synthetic and clinical data on the evolution of structures from M0 to M1. Further experiments on M12 data will be conducted to evaluate the ability of LrFM to handle regions deformation throughout the whole follow-up process (i.e. for all follow-up exams).

Our approach needs as inputs the initial seeds defined on baseline MR exam for all considered organs of interest. As mentioned earlier, this step can still be time-consuming for clinical experts, although less than a complete manual delineation. But these seeds are defined once for all and can be used for all MR follow-up exams at M_i . Therefore, one perspective would be to prevent the definition of new seeds for each new patient and consider the pre-existing seeds of the cohort. Multiple strategies could for instance be considered. First, we could use the pre-defined seeds of the closest pre-segmented patient of the database, e.g. the one with similar height and weight, or with the closest coarse segmentation of body envelopes. A second strategy would be to build a mean seed-based atlas for each organ of interest, considering all pre-defined seeds of pre-segmented patients and transport this atlas on each new patient. Another strategy would be to benefit from the whole seeds database and integrate it in a machine-learning approach to identify the position of newly defined seeds or to select the pertinent ones for each new patient, considering voxels intensity and anatomical features for instance.

6. Conclusion

Motivated by the quantitative evaluation of Chronic Lymphocytic Leukemia (CLL) response to a new drug treatment over several follow-up time points, we proposed a Longitudinal regularized Fast Marching segmentation pipeline (LrFM) for multiple organs of interest. The LrFM relies on our computationally efficient regularization strategy for the Fast Marching segmentation (rFM (Risser *et al* 2018)). The proposed LrFM pipeline takes advantage of the seeds pre-defined on a baseline image with the rFM technique to segment new images of the same patient at a different

follow-up time, and, thus, limiting user manual intervention. In our pipeline, initial seeds are registered on the new patient exam with a two-step affine registration, the seeds set is reduced from the inaccurate elements potentially transported outside the considered organs after registration; the clean set of seeds is then used to propagate the rFM.

The pipeline has been evaluated on 3D synthetic and clinical MR data with CLL, emphasizing its relevance. Results show a better performance of our technique with a better accuracy and a better robustness to organs deformation. In addition, it also allows to reduce the number of seeds to verify manually, in comparison to the direct segmentation at M1 using rFM, allowing for an important time gain for follow-up patient analysis. Consequently, the proposed segmentation pipeline performs very favorably to the tools usually available for clinicians for the longitudinal segmentation of several organs.

Acknowledgments

This work is part of the CompuTreatCLL project (C15092BS) funded by the French Plan Cancer 2015 call.

Conflict of interest

The authors declare no conflict of interest.

Ethical statement

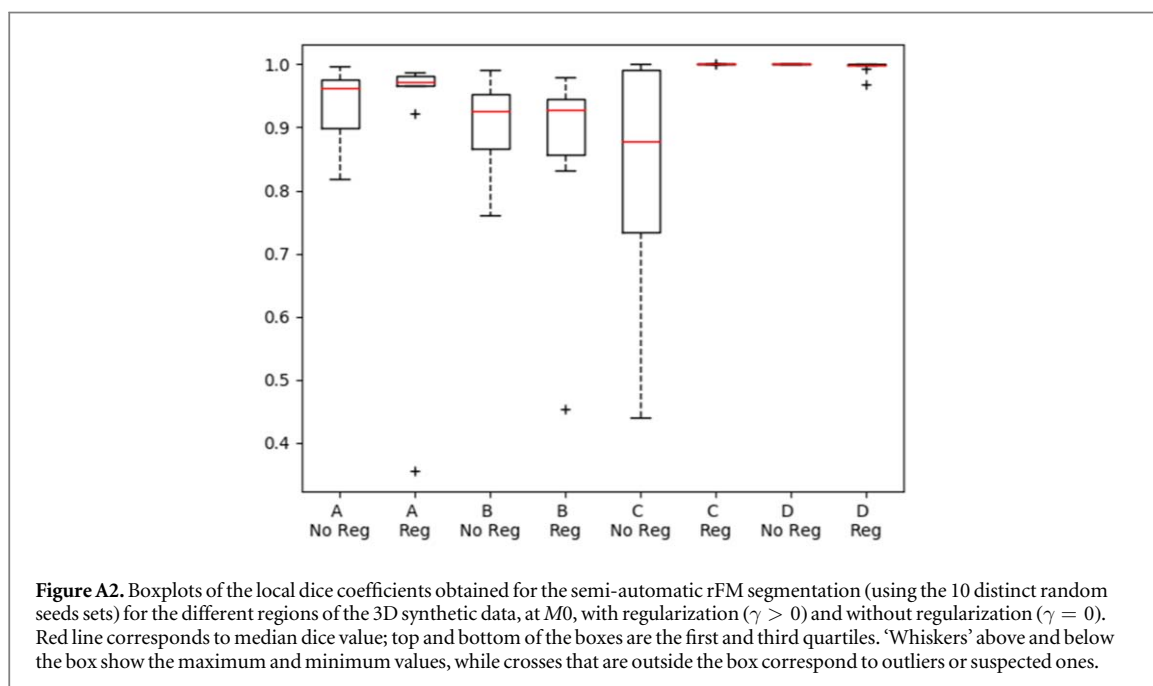
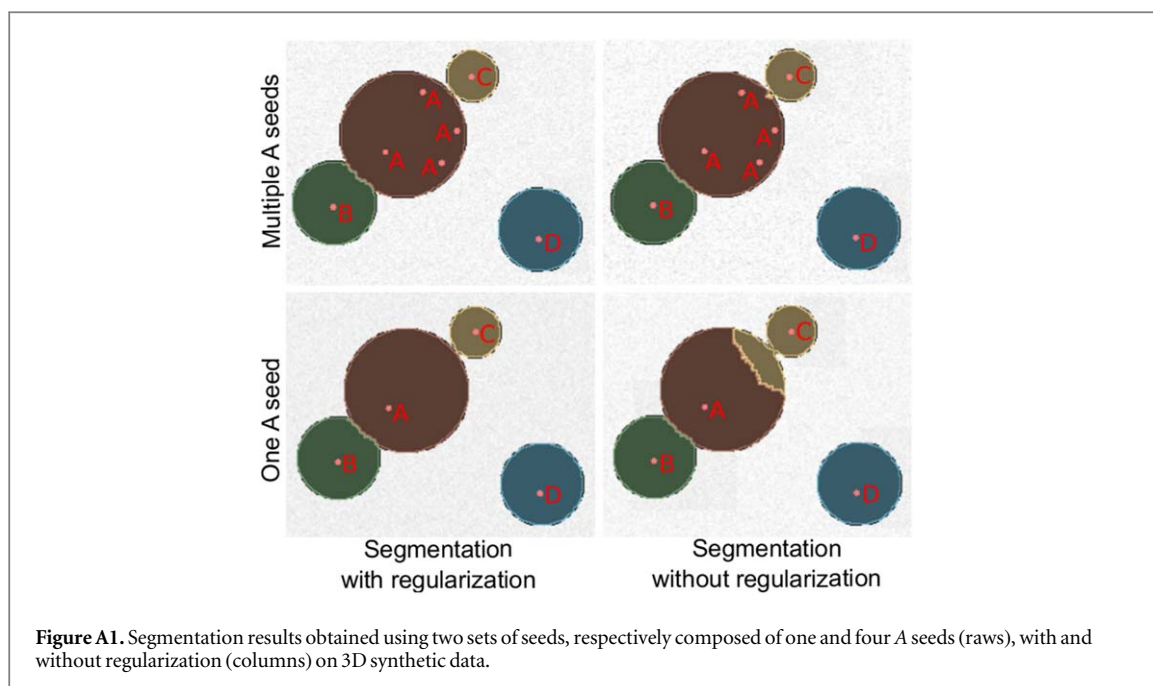
All patients involved in this study are part of a clinical trial and signed an informed consent.

Appendix. Experiments and results

A.1. Results on baseline 3D synthetic data

In this section, we show that the regularization in the rFM strategy allows the segmentation on synthetic data to be more accurate and robust to seeds placement than with non-regularized FM.

A.1.1. Methodology. The 3D synthetic data considered for the validation is described in section 4.1.1. We evaluated the robustness of the semi-automatic rFM segmentation strategy towards seeds position and the interest of the regularization, by running the segmentation process on the 3D synthetic data, considering 10 distinct sets of seeds, with regularization ($\gamma > 0$) and without regularization ($\gamma = 0$). In each of these distinct sets, one seed was randomly put in each region B , C and D , whereas a random amount of one to four seeds were randomly



placed in region *A*. Four of these segmentations are illustrated on figure A1.

A.1.2. Results. The segmentation accuracy was assessed considering the Dice coefficient (DSC), computed locally on each region of interest. Here, the semi-automatic rFM segmentation result is compared to the discrete spheres volume of each region. The local DSC, obtained for each region, are summarized in the boxplots in figure A2.

The segmentation is globally more robust to the seeds location and more accurate when the regularization is considered. We obtain a mean local dice of 0.94 ± 0.10 with regularization and 0.91 ± 0.08 without regularization, and a median local dice of 0.98

and 0.94 respectively, over all four regions averaged on the 10 considered seeds sets. More particularly, it enables to obtain higher median DSC for all regions.

The regularization also leads to a more stable segmentation with smaller variability for regions that are connected by a narrow bridge, such as regions *A* and *C*. Indeed, the spreading between regions *A* and *C* is stopped with the regularization, as observed for instance on figure A1, improving local dice coefficients. For region *C*, the median dice increases from 0.88 (without regularization) to 1.0 (with regularization).

The regularization also slightly improves the accuracy of region *B*, linked to *A* by a large bridge, a case for which the regularization was not specifically

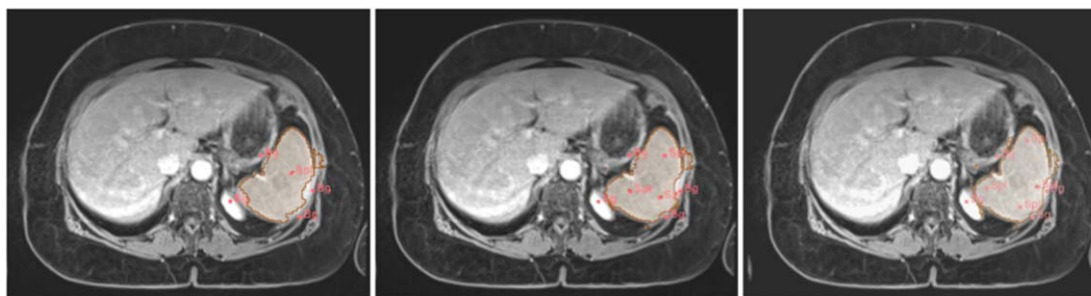


Figure A3. Segmentation results of the spleen organ obtained with different sets of seeds with variable number and positions of seeds (from left to right: 1, 3 and 4 organ seeds were used in this axial plan) using rFM segmentation. Seeds were positioned to obtain coherent and good quality segmentation.

designed. The median dice for B is 0.92 and 0.93, without and with regularization, respectively; and it is 0.96 and 0.97 for A . The variability observed on the boxplot for region Ax_R is due to the required equidistance property of seeds to limit the FM propagation in A , and that is not always fulfilled with the random position of the seeds. It is however limited with regularization. Additionally, region D is very well segmented regardless of seeds placement with a median DSC of 0.999.

A.2. Results on clinical data

The contributions in this section are twofold. In section A.2.1, we propose further results than the ones detailed in (Risser *et al* 2018) to show that the seed-based rFM segmentation of clinical data is more robust, accurate and quicker with the proposed FM regularization, compared to the complete manual delineation of regions by clinical experts. In section A.2.2, we evaluate the two registration approaches considered for the longitudinal pipeline. Details on the MR dataset of CLL patients considered for the experiments are given in section 4.2.1.

A.2.1. Validation of whole-body rFM segmentation at M0. As mentioned earlier, the semi-automatic rFM segmentation was validated at M0 for the delineation of 12 anatomical structures disseminated over the whole-body MR. These regions present a large anatomical variability and their anatomical contours are not necessarily clearly visible and require interactive segmentation.

Two experts with experience compared the time dedicated to the segmentation of these 12 organs in 3 MR images using our semi-automatic rFM approach and the manual segmentation editor of Iplan RT⁶.

The 3 rFM segmentations were performed after a training phase of 30 segmentations using the described algorithm and much more ones using Iplan RT. The physician did not change the default values of γ and $thresh$ for the experiments.

Segmentation was performed until similar accuracy was obtained with both methods, which required about 350 seeds for each patient. An average global DSC of 0.90 ± 0.03 and average local DSC of 0.67 ± 0.1 were indeed measured when comparing the rFM and the manual segmentations for the 3 patients. Mean segmentation time was 95 minutes using iPlan for the 3 segmented images and 50 minutes using our rFM segmentation for the 9 segmented images. The pre-computation of the regularization map R required in average 50 seconds and each seed propagation from 2 to 120 seconds, depending on the number of seeds.

The liver (the biggest region of interest) counted for 50 seeds in average. In comparison, a mean of 26 seeds was used for the spleen (the second biggest region of interest). The parallel segmentation of these two organs required in average 12 propagation runs. For the axillary and inguinal lymph nodes, they required less seeds (a mean of 22 and 12 seeds, respectively) and the propagation was runned 6 and 4 times, thanks to their more contrasted and regular shape. Regarding the iliac and cervical lymph nodes, they were less contrasted and needed more seeds (19 and 11 seeds in average respectively) and also more propagation runs to reach satisfactory segmentation (6 and 8 runs respectively). The propagation runs for the mediastinal and retroperitoneal lymph nodes were in average 3 and 5. The mean number of seeds reached 4 and 12 for those organs. Concerning the background seeds, required to constrain regions contours, the number of seeds were 368 in average.

Figure A3 illustrates three rFM segmentations of the spleen organ performed on the same clinical exam using different sets of seeds with variable amount and position of seeds, positioned by clinical expert with experience. Results show the low sensitivity of the rFM segmentation to the number and position of seeds. This gives a qualitative idea of the user-friendliness and robustness of the rFM tool (i.e. low inter-user variability) to obtain clinically-valid results.

⁶ www.brainlab.com

A.2.2. Evaluation of longitudinal registration in LrFM pipeline. In the LrFM pipeline, baseline seeds are propagated on follow-up images using a displacement field resulting from the two-step registration process of M_0 on M_i , and we consider here the exams acquired after one month of treatment (i.e. M_1).

We recall that the first step is a coarse global alignment of exams while the second step performs region-wise registration. We consider an affine approach for the global alignment of body envelopes between M_0 and M_1 and investigated two classical strategies for the local step. We compared the elastic registration by BSpline with affine registration. We evaluated the segmentation results considering transported seeds, according to these registration strategies, before and after seeds selection. We compared these results with a reference segmentation obtained with the direct rFM segmentation on M_1 , in terms of global and local dice scores.

We obtained equivalent segmentation results using all seeds registered with BSpline or affine techniques (equal global dice score, and similar local dice of 0.58 and 0.56 respectively for BSpline and affine registrations). Similarly, equivalent segmentation performances were obtained considering transported seeds after automatic selection (global dice of 0.82 and local dice of 0.57 for both strategies). In addition, the number of seeds to verify after seeds selection is lower for the affine approach (329) compared to the BSpline one (352). This reduction is significant to limit the time dedicated to manual seeds correction.

Hence, we chose to consider an affine registration for the region-wise step of our process since both strategies achieved similar performances, however the BSpline approach led to longer processing (more computationally expensive processes) and analysis time (higher number of seeds to verify).

ORCID iDs

Soléakhéna Ken  <https://orcid.org/0000-0001-7904-5882>

References

- Avants B B, Epstein C L, Grossman M and Gee J C 2008 Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain *Med. Im. Anal.* **12** 26–41
- Bai W, Shi W, O'Regan D P, Tong T, Wang H, Jamil-Copley S, Peters N S and Rueckert D 2013 A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac mr images *IEEE Trans. Med. Imag.* **32** 1302–15
- Bai X and Sapiro G 2009 Geodesic matting: a framework for fast interactive image and video segmentation and matting *Int. J. Comp. Vis.* **82** 113–32
- Boykov Y and Jolly M-P 2001 Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images *Proc. of ICCV (Vancouver, Canada)* pp 105–12
- Cerrolaza J J, Summers R M and Linguraru M G 2016 Soft multi-organ shape models via generalized PCA: a general framework *Proc. of MICCAI (Athens, Greece)* pp 219–28
- Chen X and Bagci U 2011 3D automatic anatomy segmentation based on iterative graph-cut-ASM *J. Med. Phys.* **38** 4610–22
- Coupric C, Grady L, Najman L and Talbot H 2010 Power watershed: a unifying graph-based optimization framework *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 1384–99
- Falcão A X, Stolfi J and De Alencar Lotufo R 2004 The image foresting transform: Theory, algorithms, and applications *IEEE Trans. Pattern Anal. Mach. Intell.* **26** 19–29
- Fedorov A et al 2012 3d slicer as an image computing platform for the quantitative imaging network *Magn. Reson. Imaging* **30** 1323–41
- Gordon S, Dolgopyat I, Kahn I and Raviv T R 2018 Multidimensional co-segmentation of longitudinal brain mri ensembles in the presence of a neurodegenerative process *NeuroImage* **178** 346–69
- Grady L 2006 Random walks for image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **28** 1768–83
- Johnson H et al 2007 BRAINSFit: mutual information rigid registrations of whole-brain 3d images, using the insight toolkit *Insight J.* **57** 1 (<http://hdl.handle.net/1926/1291>)
- Kolmogorov V and Zabini R 2004 What energy functions can be minimized via graph cuts? *IEEE T. Pattern Anal. Mach. Intell.* **26** 147–59
- Küstner T, Müller S, Fischer M, Wei J, Nikolaou K, Bamberg F, Yang B, Schick F and Gatidis S 2018 Semantic organ segmentation in 3D whole-body MR images *Proc. of ICIP (Athens, Greece)* pp 3498–502
- Lavdas I, Glocker B, Kamnitsas K, Rueckert D, Mair H, Sandhu A, Taylor S A, Aboagye E O and Rockall A G 2017 Fully automatic, multiorgan segmentation in normal whole body magnetic resonance imaging, using classification forests, convolutional neural networks, and a multi-atlas approach *J. Med. Phys.* **44** 5210–20
- Lay N, Birkbeck N, Zhang J and Zhou S K 2013 Rapid multi-organ segmentation using context integration and discriminative models *Proc. of IPMI (Asilomar, CA, USA)* pp 450–62
- Lermé N, Malgouyres F and Rocchisani J 2010 Fast and memory efficient segmentation of lung tumors using graph cuts *Proc. of MICCAI, Workshop on Pulmonary Image Analysis (Beijing, China)* pp 9–20
- Li Y, Sun J, Tang C-K and Shum H-Y 2004 Lazy snapping *ACM Trans. on Graph.* **23** 303–8
- Losnegård A, Hysing L B, Muren L P, Hodneland E and Lundervold A 2010 Semi-automated segmentation of the sigmoid and descending colon for radiotherapy planning using the fast marching method *Physics in Med & Bio* **55** 5569
- Mattes D, Haynor D R, Vesselle H, Lewellen T K and Eubank W 2003 PET-CT image registration in the chest using free-form deformations *IEEE Trans. Med. Imag.* **22** 120–8
- McInerney T and Terzopoulos D 1996 Deformable models in medical image analysis: a survey *Med. Image Anal.* **1** 91–108
- Miranda P A and Falcão A 2009 Links between image segmentation based on optimum-path forest and minimum cut in graph *J. of Math. Imag. and Vis.* **35** 128–42
- Moltz J H, Schwier M and Peitgen H-O 2009 A general framework for automatic detection of matching lesions in follow-up ct *Proc. of ISBI* 843–6
- Noone A et al 2016 Seer cancer statistics review *National Cancer Institute Bethesda, MD 1975-2015* 7 (https://seer.cancer.gov/csr/1975_2015/)
- Pham D and Prince J 2000 Current methods in medical image segmentation *Annual Rev. Biomed. Eng.* **2** 315–37
- Risser L, Ken S, Lebreton S, Grossiord E, Kanoun S and Malgouyres F 2018 Regularized multi-label fast marching and application to whole-body image segmentation *Proc. of ISBI* 1323–6
- Schmidt P et al 2019 Automated segmentation of changes in flair-hyperintense white matter lesions in multiple sclerosis on serial magnetic resonance imaging *NeuroImage: Clinical* **23** 101849
- Sethian J A 1996 A fast marching level set method for monotonically advancing fronts *Proc. National Acad. Sciences* **93** 1591–5

- Sharma N and Aggarwal L M 2010 Automated medical image segmentation techniques *J. Med. Phys.* **35** 3–14
- Unkelbach J, Menze B H, Konukoglu E, Dittmann F, Le M, Ayache N and Shih A 2014 Radiotherapy planning for glioblastoma based on a tumor growth model: improving target volume delineation *Physics in Med & Bio* **59** 747
- Valindria V V, Lavdas I, Cerrolaza J, Aboagye E O, Rockall A G, Rueckert D and Glocker B 2018 Small organ segmentation in whole-body mri using a two-stage fcn and weighting schemes *Proc of MLMI* pp 346–54
- Vivanti R, Szeskin A, Lev-Cohain N, Sosna J and Joskowicz L 2018 Patient-specific and global convolutional neural networks for robust automatic liver tumor delineation in follow-up ct studies *Med & Bio Engin & Comput* **56** 1699–713
- Wachinger C, Toews M, Langs G, Wells W and Golland P 2018 Keypoint transfer for fast whole-body segmentation *IEEE Trans. Med. Imag.* **39** 273–282
- Wang J and Cohen M F 2005 An iterative optimization approach for unified image segmentation and matting *Proc. of ICCV (Beijing, China)* 936–43
- Ysebaert L, Cadot S, Kanoun S, Ken S, Recher C, Malgouyres F, Gadat S, Dupre L and Quillet Mary A 2017 Computational integration to model tumor dynamics in CLL patients treated with the Btk Inhibitor Ibrutinib (CompuTreatCLL): first results of an integrative systems biology approach *Blood* **130** 2990
- Zuluaga M A, Orkisz M, Dong P, Pacureanu A, Gouttenoire P-J and Peyrin F 2014 Bone canalicular network segmentation in 3d nano-ct images through geodesic voting and image tessellation *Physics in Med & Bio* **59** 2155