



HAL
open science

Anatomically plausible models and quality assurance criteria for online mono- and multi-modal medical image registration

Cornel Zachiu, B Denis de Senneville, Chrit T W Moonen, Bas Raaymakers,
Mario Ries

► To cite this version:

Cornel Zachiu, B Denis de Senneville, Chrit T W Moonen, Bas Raaymakers, Mario Ries. Anatomically plausible models and quality assurance criteria for online mono- and multi-modal medical image registration. *Physics in Medicine and Biology*, 2018, 63 (15), pp.155016. hal-01962468

HAL Id: hal-01962468

<https://hal.science/hal-01962468v1>

Submitted on 21 Dec 2018

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.

Anatomically plausible models and quality assurance criteria for online mono- and multi-modal medical image registration

C Zachiu¹, B Denis de Senneville^{2,3}, C T W Moonen², B W Raaymakers¹ and M Ries²

¹ Dept of Radiotherapy, UMC Utrecht, Heidelberglaan 100, 3508 GA, Utrecht, Netherlands

² Imaging Division, UMC Utrecht, Heidelberglaan 100, 3508 GA, Utrecht, Netherlands

³ Institut de Mathématiques de Bordeaux (IMB), UMR 5251 CNRS/University of Bordeaux, F-33400 Talence, France.

E-mail: C.Zachiu@umcutrecht.nl

Abstract. Medical imaging is currently employed in the diagnosis, planning, delivery and response monitoring of cancer treatments. Due to physiological motion and/or treatment response, the shape and location of the pathology and organs-at-risk may change over time. Establishing their location within the acquired images is therefore paramount for an accurate treatment delivery and monitoring.

A feasible solution for tracking anatomical changes during an image-guided cancer treatment is provided by image registration algorithms. Such methods are, however, often built upon elements originating from the computer vision/graphics domain. Since the original design of such elements did not take into consideration the material properties of particular biological tissues, the anatomical plausibility of the estimated deformations may not be guaranteed. In the current work we adapt two existing variational registration algorithms, namely Horn-Schunck and EVolution, to **online** soft tissue tracking. This is achieved by enforcing an incompressibility constraint on the estimated deformations during the registration process. The existing and the modified registration methods were comparatively tested against several quality assurance criteria on abdominal *in-vivo* MR and CT data. These criteria included: the Dice Similarity Coefficient (DSC), the Jaccard index, the target registration error (TRE) and three **additional** criteria evaluating the anatomical plausibility of the estimated deformations.

Results demonstrated that both the original and the modified registration methods have similar registration capabilities in high-contrast areas, with DSC and Jaccard index values predominantly in the 0.8 - 0.9 range and an average TRE of 1.6 - 2.0 mm. In contrast-devoid regions of the liver and kidneys, however, the three **additional** quality assurance criteria have indicated a considerable improvement of the anatomical plausibility of the deformations estimated by the incompressibility-constrained methods. **Moreover**, the proposed registration models maintain the potential of the original methods for online image-based guidance of cancer treatments.

Keywords: Medical image registration, anatomically plausible registration models, registration quality assurance criteria

Submitted to: *Phys. Med. Biol.*

1. Introduction

Advances in medical imaging technology have led to an improvement of cancer treatment, by offering the possibility to non-invasively localize the tumor and analyze its physical and functional characteristics (Fass 2008). With the introduction of imaging devices within the operating room, medical images also began to play an important role for therapy guidance (Fass 2008, Roach 3rd et al. 2011). For treatments of long duration or repetitive interventions, the pathology and the organs-at-risk (OARs) may change their position and shape over time, due to physiological motion and/or the effects of the treatment (Keall et al. 2006, Eisenhauer et al. 2009, Zachiu et al. 2015). Therefore, in order to ensure an accurate treatment delivery and monitoring, tracking anatomical changes within the images is paramount (Keall et al. 2006, Timmerman & Xing 2009, Fanchon et al. 2017). In clinical practice, such tracking is frequently performed manually by experienced clinicians (Eisenhauer et al. 2009, Mundt & Roeske 2011). With the widespread availability of imaging systems, resulting in a growing volume of image data acquired using several modalities, manual tracking becomes increasingly cumbersome, time consuming and error-prone. Moreover, for online therapy guidance manual tracking is hardly feasible, since it requires tracking within time intervals of several minutes down to sub-second processing latencies (Zachiu et al. 2015, Denis de Senneville et al. 2015, Ries et al. 2010). Thus, an automatic solution which is capable of tracking displacements/deformations across the images acquired over the course of an image-guided cancer treatment would be highly beneficial.

A feasible solution for automatic tracking of the pathology and OARs is provided by image registration (Mani & Arivazhagan 2013), initially proposed for tracking objects within optical video sequences (see (Zitová & Flusser 2003) for an overview). A broad range of these methods were also adopted in the medical image registration domain (Maintz & Viergever 1998, Hill et al. 2001, Zitová & Flusser 2003, Brock et al. 2010, Mani & Arivazhagan 2013, Sotiras et al. 2013). A category of registration methods particularly suitable for medical image registration are variational methods (Weickert et al. 2003). This is due to their fast numerical schemes, low number of input parameters and their capability of estimating dense and elastic deformations. As a functioning principle, variational registration methods aim to find the deformation which optimizes a cost function dependent on the similarity between two images (Vese & Le Guyader 2016). The most elementary term of cost function is typically known as a similarity metric and compares the difference of common image features during the registration process. Optimizing the similarity metric alone, however, is usually an ill-posed problem with an infinity of solutions. In order to address this, constraints on the estimated deformations of the tracked structures have been included in the registration model, referred to as regularization (Vese & Le Guyader 2016, Osman et al. 2016). An example of such a cost function is the conservation of image intensity under deformation as the similarity metric combined with the constraint of local smoothness of the deformation as the regularization, proposed in the seminal work of Horn & Schunck

(Horn & Schunck 1981). The smoothness regularization basically hypothesizes that neighboring points within an image usually move together (or at least in a similar fashion) and penalizes motion patterns such as shearing. The Horn-Schunck algorithm has been adopted successfully for medical image registration in several independent studies (Ostergaard et al. 2008, Roujol et al. 2010, Glitzner et al. 2015, Yang et al. 2016). Although the algorithm suggested by Horn & Schunck was originally not developed for medical image registration and thus does not specifically take into account the material properties of biological tissues, it has been shown to be a reliable performer for the mono-modal image registration of soft-tissue structures. However, the Horn-Schunck algorithm suffers from several limitations: The similarity metric of image intensity conservation employed by the algorithm is only directly suitable for images acquired with the same image modality and the same image contrast. Furthermore, even for mono-modal image registration, intensity fluctuations not related to deformation/displacement, such as image artifacts or noise, in-flow artifacts due to arterial pulsations, contrast changes due to the presence/arrival of contrast agents, [etc.], are frequently incorrectly interpreted by the algorithm as genuine displacement (Ostergaard et al. 2008, Martel et al. 2007, Zachiu et al. 2015b). This is due to the hypothesis that any spatial location of an image point does not change its gray-level intensity except if it is displaced or deformed.

From an application point of view, the inability to perform a medical image registration across different imaging modalities and/or contrasts is particularly penalizing for the planning and guidance of image-guided cancer treatments, which frequently require the registration of several different imaging modalities (such as CT, cone-beam CT, MRI, PET) and/or image contrasts to a common frame of reference (Roach 3rd et al. 2011, Dupuy et al. 2013). This shortcoming motivated recently the development of several new algorithms that have this capability, such as MIND (Heinrich et al. 2012), SeSaMI (Rivaz et al. 2014) and EVOlution (Denis de Senneville et al. 2016).

The recently proposed EVOlution algorithm is hereby derived directly from previously suggested variational image registration algorithms based on optical-flow, but uses in the cost function a similarity metric that hypothesizes the conservation of contrast patterns between the images rather than image intensity. The rationale is that while image intensities are generally not conserved across different image modalities/contrasts, the principal contrast patterns of soft-tissue structures, tissue-bone and tissue-air interfaces are generally preserved. It is important to understand that for both the Horn-Schunck and the EVOlution algorithm, high contrast areas (e.g. organ boundaries) provide sufficient structural information for the respective similarity metric, while the deformation constraint present in the regularization is playing only a secondary role. In homogenous iso-intensity areas however, the registration process is entirely driven by the regularization as the assumed constraint of the underlying true anatomical deformation. While the seminal work on both the Horn-Schunck and the Evolution algorithm evidenced that the constraint of a smooth deformation as the regularization is generally a sensible choice, it should be noted that it is neither the only sensible choice, nor does it necessarily represent a generally valid assumption. For example, while smoothness as

a constraint of the estimated motion correctly penalizes shearing motion within organs and muscles, it also does so at organ boundaries, where this type of motion patterns frequently occurs, in particular in-between soft-tissue interfaces subject to cardiac or respiratory motion. In the context of the original Horn-Schunck algorithm, the problem to address motion discontinuities at object boundaries motivated several proposed amendments to the constraints (see for example (Terzopoulos 1986, Nessi 1993, Weickert & Schnör 2001a, Lei et al. 2007, Monzón et al. 2016)) or to employ a spatio-temporal regularization (see (Nagel 1990, Black & Anandan 1991, Weickert & Schnör 2001b)). A rather interesting approach was thereby pursued by Yang *et al* (Yang et al. 2000), in the context of analyzing satellite image observations of ocean currents, with a variational based on the optical-flow principle. In this work, the regularization of the original optical flow implementation of Horn & Schunck, also included a term enforcing incompressibility of the matter under observation. The idea to use incompressibility as a sensible constraint has also been **subsequently integrated within several existing medical image registration frameworks relying on concepts such as B-splines (Rohlfing et al. 2003, Yin et al. 2009, Wilczewska et al. 2017), demons (Mansi et al. 2011, Xing et al. 2017) or variationals (Haber & Modersitzki 2004, Haber & Modersitzki 2006)**. However, while **this generally led to an improved anatomical plausibility of the estimated deformations, in some cases it also implied an increase in the number of algorithm control parameters and/or computational time, hampering their application for online/real-time image-guided cancer therapies.**

In this paper, (I) we propose to evaluate incompressibility as the *sole* constraint to regularize the Horn-Schunck variational for medical image processing. **Compared to previous studies (Yang et al. 2000, Haber & Modersitzki 2004, Haber & Modersitzki 2006), this results in a reduction in the number of control parameters and in turn to an improved suitability for online/real-time therapy guidance.** However, while the original Horn-Schunck algorithm with its preservation of image intensity inherently leads to motion vector fields which conserve the material/structure under observation, the derived EVolution algorithm, which relies on the assumption of image contrast conservation, does not necessarily do so. As shown in the seminal paper (Denis de Senneville et al. 2016), this can be an advantage, since this allows the EVolution algorithm also to address registration problems of structures, which clearly change in volume, for example the bladder. On the other hand, this might potentially be disadvantageous for the registration of soft-tissue structures like the liver or the kidneys, which generally preserve their volume in-between observations. As a consequence, (II) we propose a **new** modified variational algorithm based on EVolution, which uses tissue incompressibility as the constraint for the regularization. All four algorithms, the modified Horn-Schunck variational, the modified EVolution variational and their original counterparts are compared in terms of performance for mono and multi-modal medical image registration **on several varied datasets**. Particular attention was paid to the differences in performance between the smoothness-regularized and the incompressibility-regularized formulations. The implementation of the proposed

algorithms was also optimized such that they are compatible with the short latencies required by online/real-time image-guided cancer therapies.

Furthermore, it has to be emphasized that for a thorough evaluation of these registration algorithms previously established matching criteria, such as the dice similarity coefficient (DSC), Jaccard index or the target registration error (TRE), are certainly mandatory but not necessarily sufficient. The DSC, Jaccard index or the TRE, are very sensitive to evidence image differences in high contrast regions such as organ boundaries or embedded vessel structures. Therefore, in case the clinical application requires the propagation of organ contours and/or labels in between the acquired images, such criteria would indeed be representative and sufficient for evaluating algorithm performance. Their sensitivity in areas which are essentially devoid of contrast is, however, limited. This can become problematic in case the application requires, for example, mapping quantitative data such as radiation dose, for the purpose of inter and intra-fraction dose accumulation in a spatially consistent manner (Brock et al. 2017). Since the three criteria are unable to indicate potential errors within deformations estimated in contrast-devoid regions, faulty mappings of the dose distribution with respect to the estimated deformations may go undetected. This in turn may introduce errors in the delivery plan re-optimization process of adaptive radiotherapy treatments. Similarly, quality assurance of deformations estimated in contrast-devoid regions, for the purpose of mapping other quantitative information such as contrast agent absorption (Hill et al. 2009) or diffusion/perfusion coefficients (Rose et al. 2001), may not be guaranteed by the three aforementioned criteria. As a consequence, we hypothesize that the impact of the modified regularization terms, which are particularly dominant in contrast-devoid regions, not only have the potential to improve the accuracy of the deformations estimated in such areas, but also require additional criteria for benchmarking. Therefore, we (III) propose the mapping of the Jacobian determinant and the curl magnitude of the resulting motion fields and an independent evaluation with spin-tagged MRI images (Fischer et al. 1993, Ibrahim 2011) as additional evaluation criteria, to assess the anatomical plausibility of the resulting image registrations.

2. Method description

2.1. Description of the investigated registration algorithms

2.1.1. *The original Horn-Schunck and EVolution algorithms* The original Horn-Schunck algorithm (referred to as HSO in the scope of this work) estimates the deformation between two images I_1 and I_2 as the minimizer of the following functional (Horn & Schunck 1981):

$$E_{HSO}(\mathbf{u}) = \sum_{\vec{r} \in \Omega} (I_1(\vec{r}) - I_2(\vec{r} + \mathbf{u}(\vec{r})))^2 + \alpha \|\vec{\nabla} \mathbf{u}(\vec{r})\|_2^2 \quad (1)$$

where \mathbf{u} is the 2D/3D displacement, Ω is the image domain, \vec{r} is the image location, $\vec{\nabla}$ is the spatial gradient operator, $\|\cdot\|_2$ is the Euclidean norm and α is an algorithm

input parameter. As previously mentioned, the Horn-Schunck functional is composed of 2 terms: a data fidelity term and a regularization term. The data fidelity term is a measure of similarity between the images, which for this particular algorithm is defined as the sum of squared differences between the image gray-level intensities. From a physical point-of-view, such a data fidelity term is built on the assumption that a voxel, as it moves, it preserves its gray-level intensity. The regularization term on the other hand, consists of the square of the Euclidean norm applied on the gradient of the displacements, constraining the estimated deformations to be spatially smooth. The input/regularization parameter α controls the degree of smoothness: the higher the value of α , the smoother the estimated deformations will be.

The original EVOlution algorithm (referred to as EVO in the scope of this work) replaces the intensity-based data fidelity term of the Horn-Schunck functional, with a term dependent on the spatial gradients of the registered images (Denis de Senneville et al. 2016). Therefore, the deformation between two images I_1 and I_2 is obtained as the minimizer of the following functional:

$$E_{EVO}(\mathbf{u}) = \sum_{\vec{r} \in \Omega} e^{-C(\mathbf{u}(\vec{r}))} + \alpha \|\vec{\nabla} \mathbf{u}(\vec{r})\|_2^2 \quad (2)$$

where:

$$C(\mathbf{u}(\vec{r})) = \frac{\sum_{\vec{s} \in \Gamma} |\vec{\nabla} I_1(\vec{s}) \cdot \vec{\nabla} I_2(\vec{s} + \mathbf{u}(\vec{s}))|}{\sum_{\vec{s} \in \Gamma} \|\vec{\nabla} I_1(\vec{s})\|_2 \|\vec{\nabla} I_2(\vec{s} + \mathbf{u}(\vec{s}))\|_2} \quad (3)$$

with Γ being a symmetric neighborhood around \vec{r} . All the other terms preserve their meaning from Eq. 1. For a better understanding of its principle, the term in Eq. 3 can be re-written under the following shape:

$$C(\mathbf{u}(\vec{r})) = \frac{\sum_{\vec{s} \in \Gamma} w_{\mathbf{u}}(\vec{s}) |\cos(\Delta\phi_{\mathbf{u}}(\vec{s}))|}{\sum_{\vec{s} \in \Gamma} w_{\mathbf{u}}(\vec{s})} \quad (4)$$

where $w_{\mathbf{u}}(\vec{s})$ and $\Delta\phi_{\mathbf{u}}(\vec{s})$ are computed based on the magnitude M and orientation ϕ of the image spatial gradients:

$$\begin{aligned} w_{\mathbf{u}}(\vec{s}) &= M_{I_1}(\vec{s}) M_{I_2}(\vec{s} + \mathbf{u}(\vec{s})) \\ \Delta\phi_{\mathbf{u}}(\vec{s}) &= \phi_{I_1}(\vec{s}) - \phi_{I_2}(\vec{s} + \mathbf{u}(\vec{s})) \end{aligned} \quad (5)$$

Basically speaking, the weight $w_{\mathbf{u}}(\vec{s})$ favors strong gradients that occur in both of the images. As a consequence, the registration process is driven only by structures which are common in the two images. The term $|\cos(\Delta\phi_{\mathbf{u}}(\vec{s}))|$, on the other hand, favors deformations which align both parallel and anti-parallel gradients. In effect, this renders the EVOlution algorithm robust to any possible contrast reversals between the aligned images. Note that the regularization term of the EVOlution functional remains unchanged compared to the Horn-Schunck algorithm. Therefore, the spatial smoothness constraint is maintained on the estimated deformations.

2.1.2. The Horn-Schunck and EVolution algorithms, with modified regularization It is a known fact from continuum mechanics that, when an incompressible material is subjected to an external force, the Jacobian determinant of its deformation is equal to one (Rohlfing et al. 2003, Bistoquet et al. 2008, Stout et al. 2016). Thus, in order to constrain the deformations estimated by the Horn-Schunck and the EVolution algorithms to be incompressible, deviations from unity of the Jacobian determinant of the deformations are penalized within the regularization term of their respective functionals. Registration is hereby performed via the following optimization problems:

$$E_{HSI}(\mathbf{u}) = \sum_{\vec{r} \in \Omega} (I_1(\vec{r}) - I_2(\vec{r} + \mathbf{u}(\vec{r})))^2 + \alpha \|J(\vec{r} + \mathbf{u}(\vec{r})) - 1\|_2^2 \quad (6)$$

for the incompressibility-regularized Horn-Schunck algorithm (abbreviated as HSI in the scope of this work) and:

$$E_{EVI}(\mathbf{u}) = \sum_{\vec{r} \in \Omega} e^{-C(\mathbf{u}(\vec{r}))} + \alpha \|J(\vec{r} + \mathbf{u}(\vec{r})) - 1\|_2^2 \quad (7)$$

for the incompressibility-regularized EVolution algorithm (abbreviated as EVI in the scope of this work). In both Eq. 6 and 7, J is defined as:

$$J(\vec{r} + \mathbf{u}(\vec{r})) = \begin{vmatrix} 1 + \frac{\partial u_1(\vec{r})}{\partial x} & \frac{\partial u_1(\vec{r})}{\partial y} & \frac{\partial u_1(\vec{r})}{\partial z} \\ \frac{\partial u_2(\vec{r})}{\partial x} & 1 + \frac{\partial u_2(\vec{r})}{\partial y} & \frac{\partial u_2(\vec{r})}{\partial z} \\ \frac{\partial u_3(\vec{r})}{\partial x} & \frac{\partial u_3(\vec{r})}{\partial y} & 1 + \frac{\partial u_3(\vec{r})}{\partial z} \end{vmatrix} \quad (8)$$

where $\mathbf{u} = (u_1, u_2, u_3)$ are the 3D displacements.

2.1.3. Optimization of the motion estimation functionals Since the functionals illustrated in Eq. 1, 2, 6 and 7 are differentiable, their optimization was carried-out by solving the associated Euler-Lagrange equations. This resulted in all cases in a non-linear system of equations which was resolved using a fixed-point scheme similar to the one described in (Denis de Senneville et al. 2016).

2.1.4. Coarse-to-fine strategy In order to reduce the risk of the optimization scheme converging towards a local minima, the input images were processed using a coarse-to-fine strategy. This basically iterated the optimization scheme from a 16-fold downsampled version of the images up to the original resolution, with a resolution upsampling factor of 2. The displacements estimated at one resolution level were used as initial values for the next resolution level.

2.1.5. Employed convergence criterion The fixed-point scheme used to solve the Euler-Lagrange equations, is an iterative procedure. Similar to (Denis de Senneville et al. 2016), the iterations were stopped when the average absolute difference between two successive iterates of the motion vector field was smaller than 10^{-3} voxels/pixels.

2.2. Experimental setup

The original and the re-formulated Horn-Schunck and EVolution registration methods were evaluated in two complementary studies: one conducted on a dataset comprised of 3D MR and CT images, and one on spin-tagged MR images. The upcoming sections provide details on these two studies, including the quality assessment criteria used to evaluate the investigated registration methods.

2.2.1. Evaluation of the registration methods on 3D CT and MR data

Description of the selected datasets The Horn-Schunck and the EVolution algorithms, in both their original and modified versions, were initially employed for aligning five batches of 3D medical images. The five batches consisted of the following: (1) Five pairs of phase-cycled gradient-recalled turbo field echo T1-weighted (T1w) MR images, (2) Five pairs of CT images, (3) A dataset comprised of five pairs of T1w and T2-enhanced (T2e) steady-state gradient echo MR images, (4) Five pairs of CT and cone-beam CT (CBCT) images and (5) A dataset which included five pairs of CT and turbo spin echo T2-weighted (T2w) MR images. Datasets (1) and (3) were acquired on the abdomen of a healthy volunteer, while datasets (2), (4) and (5) were obtained from patients suffering from kidney cancer †. Except for dataset (5), whose field-of-view only includes the kidney, all the other images include both the liver and the kidneys. The acquisition parameters for the MR images are reported in Table 1. Note that the T1w MR images from datasets (1) and (3) were acquired using the same MR sequence. The CT and the CBCT images had a matrix and a voxel size of $512 \times 512 \times 123$ and $1.1 \times 1.1 \times 3\text{mm}^3$, respectively. In order to ensure spatial consistency, all images were mapped prior to registration onto common lattice of size $256 \times 256 \times 128$ with an isotropic voxel volume of $1.5 \times 1.5 \times 1.5\text{mm}^3$.

In a complementary experiment, the four algorithms were also used to register datasets #1 – #5 from the DIR-Lab database (Castillo et al. 2010). Each dataset is a pair of CT images acquired on an esophagus cancer patient, at the two extreme phases of the respiratory cycle. For details regarding the image acquisition parameters please consult (Castillo et al. 2010).

Quality assessment criteria for the estimated deformations The five batches of 3D CT and MR images described in the previous paragraph were registered using the four investigated registration methods. Since the data fidelity term of the Horn-Schunck algorithm in both its original and modified version is unsuitable for cross-contrast and cross-modality registration, the two methods were only employed for the first two data batches (i.e. for T1w MR to T1w MR and CT to CT registration).

The performance of the registration methods was evaluated against several criteria. In

† Ethical approval for the acquisition, usage and publication of the data was provided by the Ethics Board of the UMC Utrecht, Utrecht, The Netherlands.

Sequence	T1w-MR [Dataset (1)]	T2e-MR [Dataset (3)]	T2w-MR [Dataset (5)]
TR [ms]	4.3	3	523
TE [ms]	2	1.5	80
FA°	10	40	90
Volume size	192 × 192 × 75	192 × 192 × 75	288 × 288 × 53
Voxel size [mm ³]	2 × 2 × 2	2 × 2 × 4	1.4 × 1.4 × 3
BW _{readout} [Hz]	373	723	478

Table 1: Acquisition parameters for the 3D MR images used in the evaluation of the four investigated registration methods. Distinct MR acquisition sequences were used for the T1w image from dataset (1) (second column), the T2e image from dataset (3) (third column) and the T2w image from dataset (5) (fourth column). The abbreviations for the acquisition parameters are as follows: TR - repetition time, TE - echo time, FA - flip angle and BW - bandwidth.

an initial step, two typically-employed quality assessment metrics were analyzed: the Dice Similarity Coefficient (DSC) (Dice 1945) and the Jaccard index (Jaccard 1912). The two criteria were used to determine the contour overlap for the liver and kidneys in the images, before and after registration. Such an analysis showcases the capability of the investigated registration algorithms for organ contour alignment, which is representative for their performance in high-contrast areas. Mathematically, the DSC and the Jaccard index are defined as follows:

$$DSC(A, B) = \frac{2 * |A \cap B|}{|A| + |B|}; \quad \text{Jaccard}(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (9)$$

where A and B are the two contours whose overlap is determined and $|\cdot|$ corresponds to the number of voxels within a contour. Note that for all images in the five initial data batches, delineations for the liver and kidneys were made available by an experienced radiologist.

The performance of the algorithms in high contrast areas was further evaluated by comparing the estimated and the manually-annotated locations of 1500 landmark features identified by the DIR-Lab database provider in the five datasets analyzed in the scope of this study. The evaluation was performed in terms of the target registration error (TRE) (Maurer et al. 1997):

$$\text{TRE}[\text{mm}] = \sqrt{(x_I + u_1(\vec{r}_I) - x_J)^2 + (y_I + u_2(\vec{r}_I) - y_J)^2 + (z_I + u_3(\vec{r}_I) - z_J)^2} \quad (10)$$

where (x_I, y_I, z_I) and (x_J, y_J, z_J) are the gold standard landmark coordinates in the reference and the moving image, and $(u_1(\vec{r}_I), u_2(\vec{r}_I), u_3(\vec{r}_I))$ is the estimated 3D displacement for that particular landmark.

In order to evaluate the validity of the deformations in contrast-devoid regions, the local Jacobian determinant (see Eq. 8) and the local curl magnitude of the estimated

deformations were analyzed with respect to the underlying anatomy. The Jacobian determinant is used as an indicator whether the estimated deformations are in good correspondence with the incompressibility property of biological soft tissues. For the incompressibility-regularized algorithms however, the analysis of the Jacobian determinant is simply an indicator for algorithm convergence. Therefore, an independent quality assurance criterion is necessary for their evaluation. Since the incompressibility constraint enforces the estimation of a solenoidal deformation vector field, the formation of vortices is not penalized. In effect, **as also suggested by previous studies (Screibmann et al. 2012)**, we propose the analysis of the local vorticity of the estimated deformations, provided by the curl magnitude, as an additional quality assurance criterion which is not explicitly enforced by any of the investigated motion estimation methods. Mathematically, the curl of a 3D vector field is computed as:

$$\text{curl}(\vec{\mathbf{u}}) = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ u_1 & u_2 & u_3 \end{vmatrix} \quad (11)$$

where $\vec{\mathbf{u}} = (u_1, u_2, u_3)$ is the 3D vector field and $(\vec{i}, \vec{j}, \vec{k})$ are the versors of the 3D Euclidean space.

2.2.2. Evaluation of the investigated registration methods on spin-tagged MR images

MR-tagging is a technique, which has been historically introduced to investigate the dynamics of the myocardium for diagnostic purposes (Fischer et al. 1993, Ibrahim 2011). RF-saturation pulses are used to place low-signal stripes or grids on the images, which move together with the underlying tissue. In the current work, this technique was used for validation purposes, since it provides a non-invasive way to acquire an *in-vivo* gold standard for organ/tissue displacement, in particular in image regions devoid of anatomical contrast.

Two pairs of MR images were acquired during this experiment: one without and one with spin-tagging. The tagging consisted in a horizontal stripe pattern with a 5 mm distance between the stripes. Each pair contained two sagittal 2D MR slices acquired on the abdomen of a healthy volunteer at the two extreme phases of the respiratory cycle, via respiratory gating.

The four investigated algorithms were initially used to register the pair of un-tagged MR images. Similar to the analysis described in section 2.2.1, the pre- and post-registration DSC and Jaccard index were analyzed for the liver, kidney and, for this particular experiment, also the thoraco-abdominal wall. The Jacobian determinant and the curl magnitude of the deformations were also comparatively analyzed between the original and the modified versions of the Horn-Schunck and the EVolution algorithms. The deformations resulting from the registration of the un-tagged pair of images were then used to align the spin-tagged images. In the ideal case, the post-alignment tagging patterns in the two images should match. For a quantitative evaluation of

the registration errors, the distance between the tag lines pre- and post-alignment was also calculated.

The MR-acquisition sequence, for both the un-tagged and tagged images, employed the following parameters: TR = 90 ms, TE = 8.8 ms, FA = 35°, image size = 224 × 224, voxel size = 1.8 × 1.8 × 5mm³ and BW_{readout} = 996 Hz.

2.3. Algorithm parameter calibration and implementation

The original and the modified Horn-Shunck algorithm require the calibration of the regularization parameter α (see Eq. 1 and 6). Besides this, the EVolution algorithm in both its formulations also requires configuring the size of the additional parameter Γ (see Eq. 2, 3 and 7). For the experiment described in 2.2.1 an exhaustive search was conducted for these parameters and the values which provided the highest average post-registration DSC for all delineated structures were then chosen for use. The rationale behind such an approach is that, in a clinical setting, the output of registration algorithms is often evaluated based on a visual inspection of the post-registration organ boundary alignment between the images. Thus, maximizing the DSC is representative for such an evaluation criterion. Note that only the first image pair in each of the five initial data batches were used for the calibration procedure. This allowed investigating the re-usability of the parameters α and Γ for the remainder of the 3D images.

On the other hand, for the spin-tagged datasets, the values of the parameters which provided the best match between the tag patterns after registration, were selected for use.

The numerical schemes of all four registration algorithms were implemented and executed on an nVidia Tesla K20 graphics card, using the compute unified device architecture (CUDA).

3. Results

3.1. Performance of the investigated registration methods in high-contrast areas

Table 2 reports the DSC and the Jaccard coefficients for the liver and the kidneys before and after registration of each of the five initial data batches described in section 2.2.1. The reported DSC and Jaccard index values were obtained following an averaging over the five image pairs contained by each batch. As a reminder, in the scope of this study the four algorithms were abbreviated as follows: HSO - the original Horn-Schunck algorithm with a constraint on smoothness, HSI - the modified Horn-Schunck algorithm with a constraint on the Jacobian determinant, EVO - the original EVolution algorithm with a constraint on smoothness and EVI - the modified EVolution algorithm with a constraint on the Jacobian determinant. All the algorithms provide high values of the post-registration DSC and the Jaccard index, predominantly in the 0.8 - 0.9 range. It can also be observed that HSI and the EVI provide similar results compared to their smoothness-regularized counterparts.

Table 3 showcases the average TRE obtained on the five DIR-Lab datasets (see section

Criterion	Data batch	Organ	Algorithm				
			None	HSO	HSI	EVO	EVI
DSC	T1w MR - T1w MR	Liver	0.91	0.96	0.96	0.96	0.96
		Kidneys	0.85	0.94	0.94	0.94	0.94
	CT - CT	Liver	0.84	0.90	0.90	0.91	0.91
		Kidneys	0.83	0.90	0.90	0.92	0.92
	T1w MR - T2e MR	Liver	0.89	-	-	0.94	0.94
		Kidneys	0.75	-	-	0.86	0.86
	CT - CBCT	Liver	0.84	-	-	0.90	0.90
		Kidneys	0.83	-	-	0.90	0.90
	CT - MR	Liver	-	-	-	-	-
		Kidneys	0.66	-	-	0.87	0.86
Jaccard Index	T1w MR - T1w MR	Liver	0.84	0.92	0.92	0.93	0.93
		Kidneys	0.76	0.88	0.88	0.89	0.89
	CT - CT	Liver	0.74	0.83	0.82	0.83	0.84
		Kidneys	0.72	0.82	0.81	0.85	0.85
	T1w MR - T2e MR	Liver	0.83	-	-	0.88	0.89
		Kidneys	0.62	-	-	0.76	0.76
	CT - CBCT	Liver	0.74	-	-	0.81	0.81
		Kidneys	0.72	-	-	0.81	0.82
	CT - MR	Liver	-	-	-	-	-
		Kidneys	0.52	-	-	0.77	0.77

Table 2: Evaluation of the performance of the four investigated algorithms in high contrast areas. The table reports the pre- and post-registration DSC and the Jaccard index for the liver and the kidneys following the registration of the five data batches containing 3D MR and CT images. The values for the DSC and Jaccard indices were averaged over the five pairs of images included by each data batch. The missing entries either indicate the absence of a delineation for a particular anatomical structure or the fact that a method was unsuitable for a particular data set. The abbreviations for the algorithms are provided in the manuscript text.

2.2.1 for details). It can be observed that improvements were achieved in all cases, with a registration accuracy and precision smaller than 2 mm. Also, only marginal differences can be noticed between the TREs calculated for the original Horn-Schunck and EVOlution methods compared to their versions with modified regularization.

Method	TRE [mm]
None	6.39 ± 4.24
Horn-Schunck Original	1.85 ± 1.68
Horn-Schunck Incompressible	1.94 ± 1.58
EVolution Original	1.66 ± 1.53
EVolution Incompressible	1.75 ± 1.74

Table 3: Accuracy and precision of the four investigated algorithms with respect to manually tracked landmarks. The table reports the target registration errors associated to the methods, averaged over the five DIR-Lab datasets. The errors are reported under the format *mean \pm standard deviation*.

3.2. Compression/Expansion and vorticity of the deformations provided by the investigated registration methods

Table 4 reports several statistical parameters of the Jacobian determinant and the curl magnitude of the deformation vector fields estimated by the four investigated registration methods on the five data batches. Analysis of the two criteria was performed jointly for the liver and kidneys, with the resulting statistical parameters being averaged over all five image pairs contained by each data batch. It can be observed that, while all methods have a mean Jacobian determinant close to one, the incompressibility-regularized algorithms showcase a considerably lower standard deviation compared to their smoothness-constrained counterparts. In addition, the range of the Jacobian determinant for HSI and EVI is overall notably smaller compared to HSO and EVO. On the other hand, no obvious tendencies can be observed when comparing HSO and HSI in terms of the curl magnitude. In this regard however, the original EVolution algorithm leads to marginally larger values relative to the proposed incompressibility-regularized version.

The statistical distribution of the Jacobian determinant of the joint liver and kidney deformations is also displayed in Figure 1. The illustrated boxplots contain the Jacobian determinants of the liver and kidney deformations, pooled from all pairs of images included in each data batch. This particular analysis re-confirms the results reported in Table 4, with HSI and EVI showcasing overall smaller fluctuations of the Jacobian determinant from unity compared to HSO and EVO. A one-tailed Mann-Whitney-Wilcoxon test applied for each data batch, in between the original and the modified versions of the algorithms, revealed a statistically significant difference in all cases at 0.05 significance level.

For a more refined analysis of the Jacobian determinant relative to the observed anatomy, figure 2 displays the spatial distribution of the Jacobian determinant of the deformations provided by the four registration methods. The illustration is performed for a sagittal slice selected from the first image pair of each of the five 3D data batches.

Data batch	Stat.	Jacobian				Curl Magnitude			
		HSO	HSI	EVO	EVI	HSO	HSI	EVO	EVI
T1w MR - T1w MR	Mean	1.06	1.03	1.00	1.00	0.13	0.11	0.17	0.12
	Std	0.30	0.07	0.17	0.05	0.09	0.08	0.11	0.06
	Min	-0.27	0.51	0.36	0.78	0	0	0	0
	Max	8.72	1.85	2.44	1.43	1.73	1.22	1.16	0.54
CT - CT	Mean	1.00	1.00	1.00	1.00	0.04	0.07	0.20	0.14
	Std	0.05	0.01	0.27	0.08	0.02	0.05	0.13	0.08
	Min	0.51	0.52	-0.10	0.60	0	0	0	0
	Max	2.04	1.62	3.04	1.63	0.94	1.01	1.40	0.71
T1w MR - T2e MR	Mean	-	-	1.02	1.01	-	-	0.22	0.14
	Std	-	-	0.20	0.06	-	-	0.12	0.06
	Min	-	-	0.36	0.78	-	-	0	0
	Max	-	-	2.35	1.28	-	-	0.88	0.43
CT - CBCT	Mean	-	-	1.03	1.02	-	-	0.15	0.11
	Std	-	-	0.18	0.05	-	-	0.09	0.06
	Min	-	-	0.41	0.82	-	-	0	0
	Max	-	-	2.00	1.30	-	-	0.77	0.40
CT - MR	Mean	-	-	0.90	0.93	-	-	0.30	0.18
	Std	-	-	0.31	0.07	-	-	0.14	0.07
	Min	-	-	0.06	0.72	-	-	0	0
	Max	-	-	2.62	1.26	-	-	1.06	0.48

Table 4: The Jacobian determinant and curl magnitude of the deformations estimated by the four investigated registration methods on the five data batches. The table reports the mean, standard deviation, the minimum and the maximum of the two criteria, averaged over all five pairs of images in each data batch. The Jacobian and the curl values were analyzed jointly for the liver and kidneys.

Following a visual inspection, strong fluctuations from unity of the Jacobian determinant can be observed for the HSO and for EVO, in particular within homogeneous regions of the liver and kidneys. These fluctuations are considerably reduced by HSI and EVI, for which the Jacobian determinant is notably closer to one, with only slight deviations. Figure 3 showcases the vorticity of the deformations estimated by the four registration algorithms. The magnitude of the curl of the motion vector fields is displayed for the same sagittal slices selected from the five 3D image pairs, as in figure 2. Within the boundaries of the liver and kidney, except for a few problematic areas, the vorticity of the estimated deformations remains predominantly close to zero. A tendency to increase of the deformation vorticity can be observed, however, in the proximity of the thoracic, peritoneal and back wall, where sliding motion is typically expected.

Figure 4 depicts the Jacobian determinant, the motion vectors and the curl magnitude

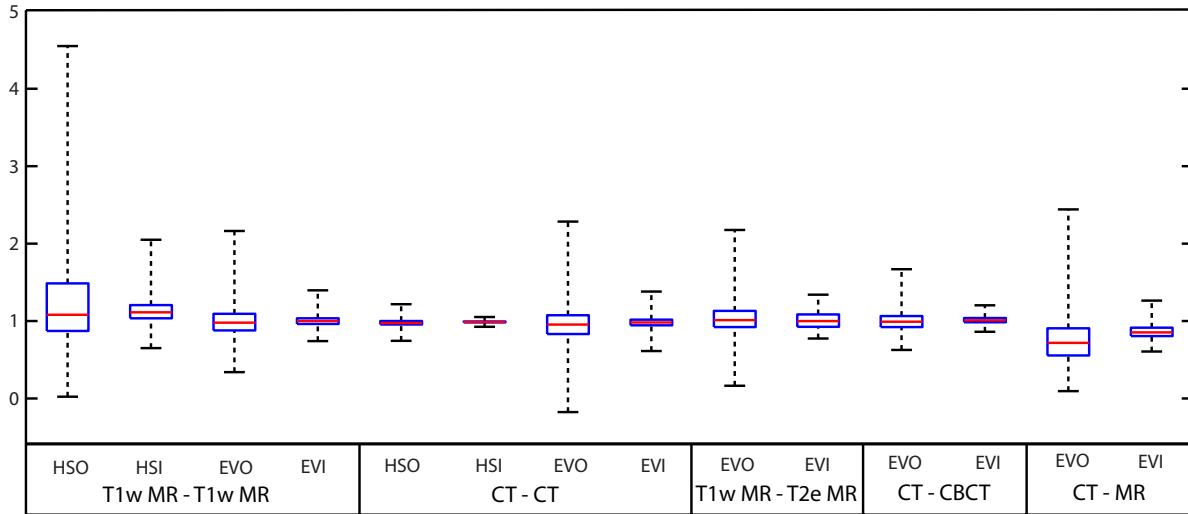


Figure 1: Statistical distribution of the Jacobian determinant of the liver and kidney deformations estimated by the four registration methods. Evaluation was performed for each of the five registered 3D data batches. The illustrated boxplots are constructed as follows: the box limits are the 25th and the 75th percentiles, the whiskers correspond to the 5th and the 95th percentiles and the red line is the set median. The differences between the original and modified versions of the Horn-Schunck and EVolution algorithms, respectively, are statistically significant at 0.05 significance level.

of the deformations estimated by HSO and HSI inside in ROI 1 from figures 2 and 3. Again, it can be observed that there is a link between the values in the Jacobian and the curl maps and the motion vector field estimated by HSO, which showcases a regional tendency to diverge, with a strong vortex in the ventral part of the kidney. This indicates that the HSO estimated both stretches and swirls of the liver and kidney tissues within the selected ROI. Both of these effects are dampened by the HSI algorithm, with the estimated motion fields displaying a higher degree of uniformity.

Figure 5 illustrates the Jacobian determinant, the motion vectors and the curl magnitude of the deformations estimated by EVO and EVI inside ROI 2 from figures 2 and 3. Unnatural divergences and convergences can be observed within the motion vector field, in areas corresponding to deviations of the Jacobian determinant from unity. These, in turn, also introduce a particular degree of vorticity within the deformations, which are the cause of the peak values observed in the illustrated curl map. The deformations estimated by EVI, on the other hand, showcase a higher degree of uniformity, with no notable peaks in either the Jacobian determinant nor the curl magnitude map.

3.3. Absolute registration errors of the investigated algorithms

The images in figure 6 illustrate the two pairs of images described in section 2.2.2, used to determine the accuracy of the four registration algorithms. Figures 6(a) and 6(b)

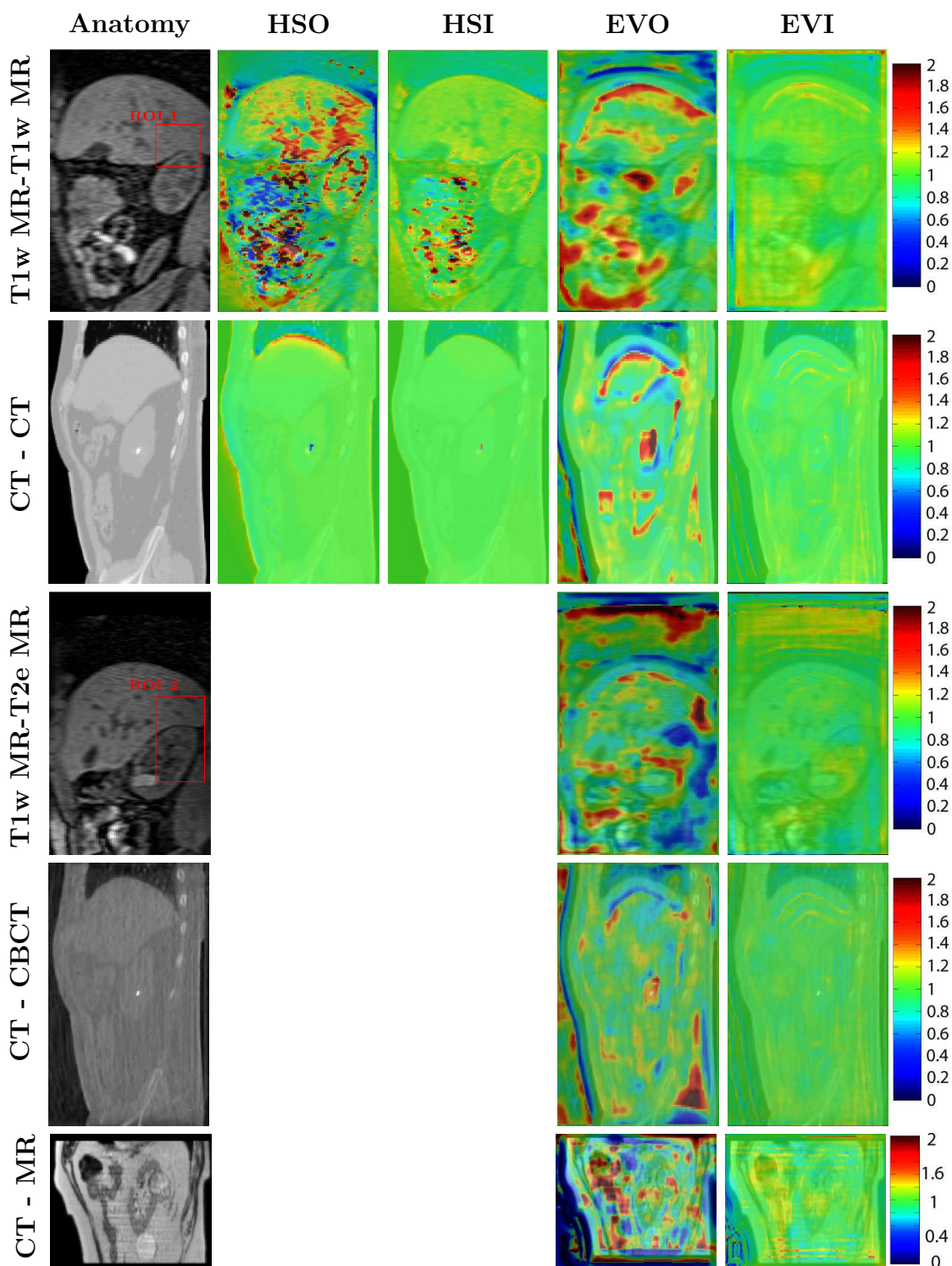


Figure 2: Spatial distribution of the Jacobian determinant. First column: A sagittal slice selected from the first pair of images included by each of the registered 3D data batches. Second to fifth column: Spatial distribution of the Jacobian determinant of the deformations provided by each of the four investigated methods overlapped with the slice from the first column. ROI 1 and 2 are two selected problematic areas, for which a more detailed analysis will be conducted.

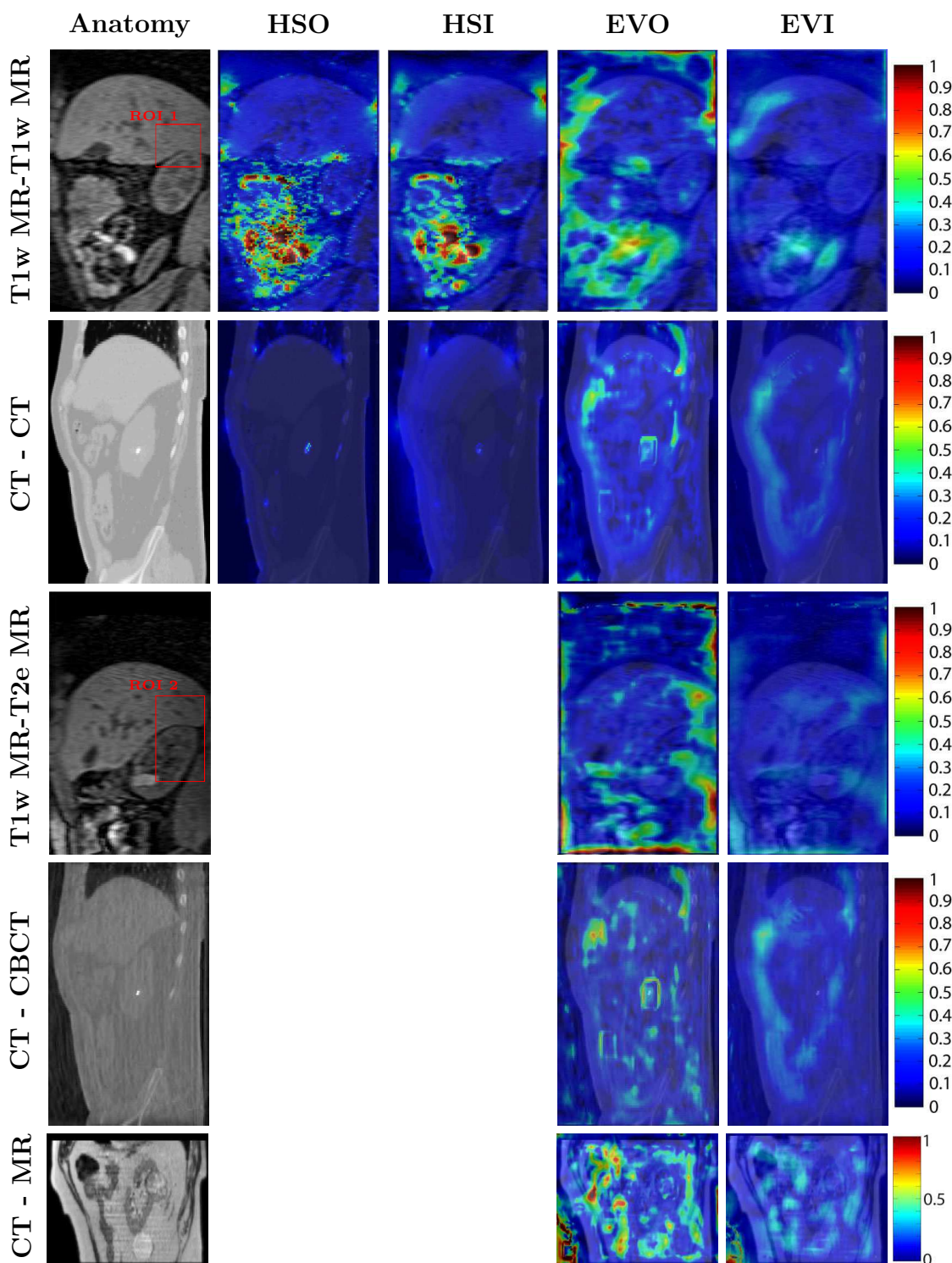


Figure 3: Vorticity of the deformations estimated by the four registration algorithms. The first column depicts a sagittal slice selected from the first pair of images included by each of the registered 3D data batches. Columns two to four showcase the curl magnitude of the motion vector fields estimated by the four methods overlaid on the sagittal slice displayed in the first column. ROI 1 and 2 are the same selected problematic areas indicated in Figure 2.

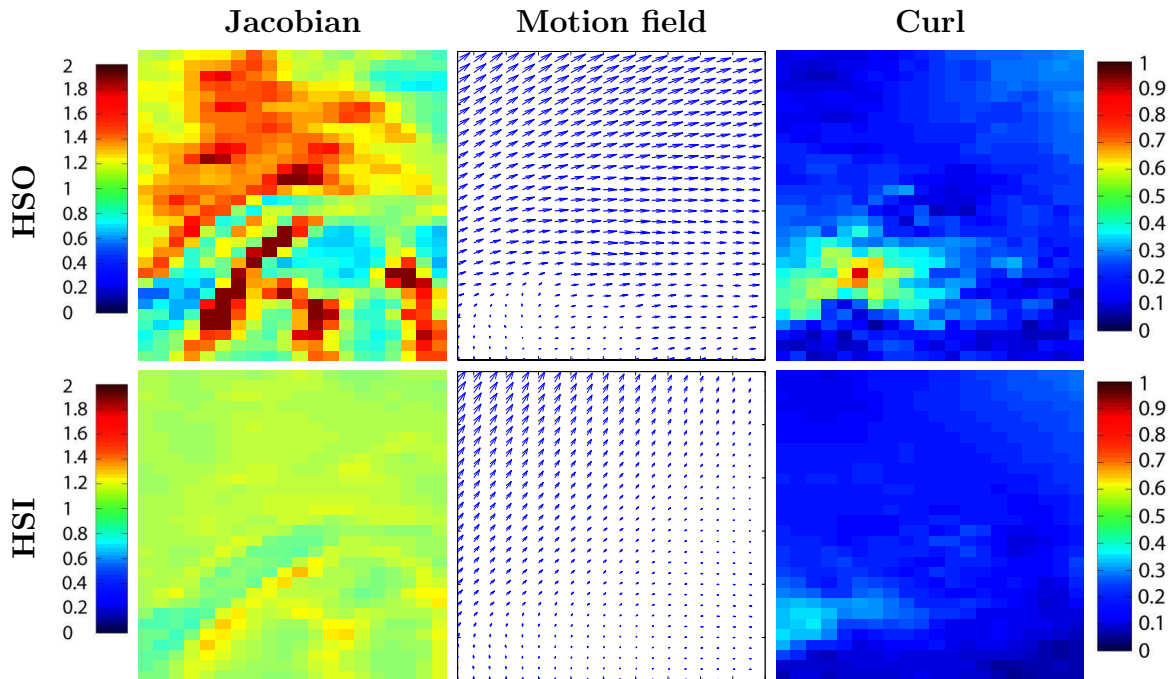


Figure 4: The Jacobian determinant (first column), motion vector field (second column) and curl magnitude map (third column) of the deformations estimated by HSO (first row) and HSI (second row) in ROI 1 from figures 2 and 3. The selection of this region for a more refined analysis was motivated by the large Jacobian determinants of the deformations estimated by the HSO in ROI 1, coupled with the formation of a strong vortex in the ventral part of the kidney.

showcase the first pair of images acquired on the healthy volunteer, without applying spin-tagging. Figure 6(c) and 6(d) illustrate the other pair of 2D scans acquired at the same respiratory phases as the images in 6(a) and 6(b), in the presence of tagging. The tag pattern allows to visually determine the trajectory followed by the spin-tagged tissue, as the volunteer went from inhalation to exhalation.

As described in section 2.2.2, all of the four registration algorithms were used to estimate the displacements between the images in figures 6(a) and 6(b). Table 5 reports the pre- and post-registration DSC and Jaccard index for three ROIs: liver, kidney and the thoraco-abdominal wall. Overall high values (≥ 0.9) were obtained for all three structures by all of the registration methods. Only marginal differences can be observed between the original Horn-Schunck and EVolution algorithms and their respective incompressibility-constrained counterparts.

The resulting deformations were used to align/map the tagged image in figure 6(d) to the image in figure 6(c). In the ideal case, the tag pattern of this mapped image should match the one in figure 6(c). Figure 7 illustrates the following: the mapped tagged images (first row), the Jacobian determinant and the curl magnitude maps of the estimated deformations (second and third row) and spatial distribution of the absolute registration errors (fourth row). Details related to the manner in which these errors were

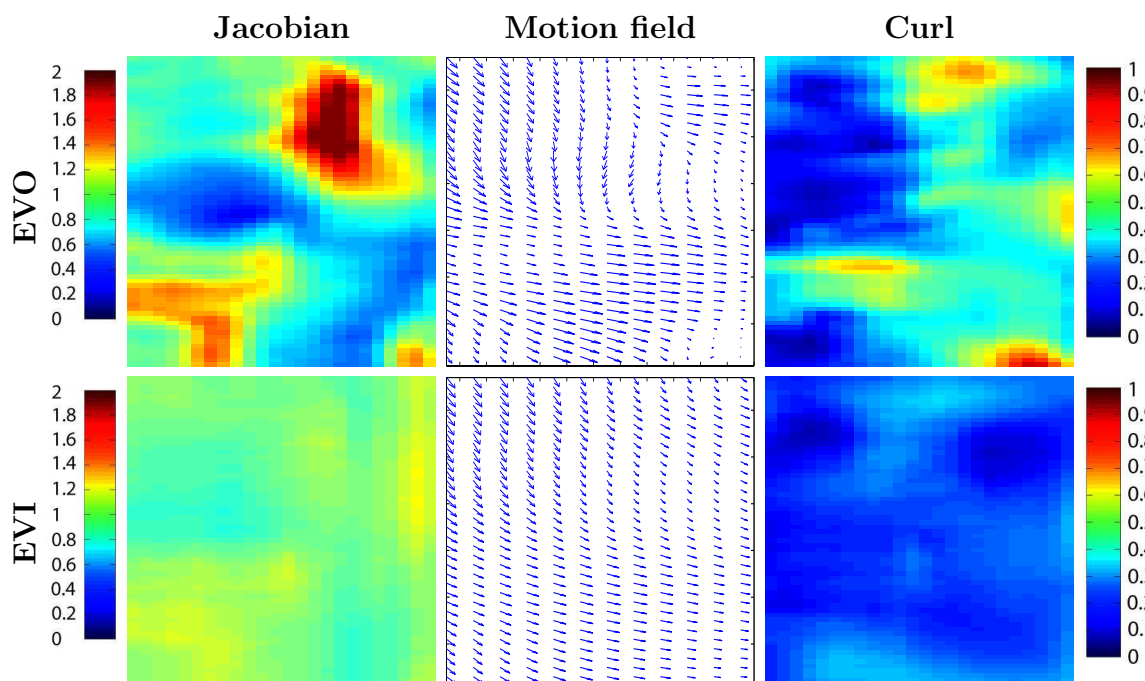


Figure 5: The Jacobian determinant (first column), motion vector field (second column) and curl magnitude map (third column) of the deformations estimated by EVO (first row) and EVI (second row) in ROI 2 from figures 2 and 3. The selection of this region for a more refined analysis was motivated by the large fluctuations in the Jacobian determinant and the curl magnitude maps of the deformations estimated by EVO.

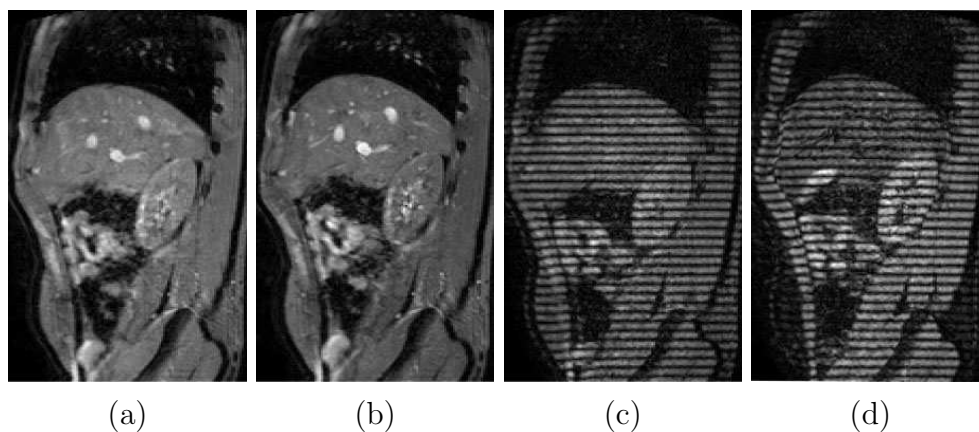


Figure 6: The two pairs of MR images acquired on the abdomen of a healthy volunteer during the study described in section 2.2.2. (a), (b): In the absence of spin-tagging. (c), (d): With spin-tagging applied. The images in each pair were acquired at the two extreme phases of the respiratory cycle.

calculated can be found in section 2.2.2. Each column corresponds to one of the four algorithms. It can be observed that for all of the evaluation criteria showcased in figure 7, the HSO and HSI provide similar results, with moderate deviations of the Jacobian

Criterion	ROI	Registration method				
		None	HSO	HSI	EVO	EVI
DSC	Liver	0.90	0.98	0.98	0.97	0.97
	Kidney	0.91	0.98	0.99	0.97	0.98
	Wall	0.82	0.95	0.94	0.94	0.94
Jaccard index	Liver	0.82	0.96	0.96	0.95	0.95
	Kidney	0.83	0.96	0.97	0.95	0.96
	Wall	0.70	0.90	0.89	0.89	0.89

Table 5: Pre- and post-registration contour overlap for three regions-of-interest (ROIs) in the images from figure 6. The DSC and the Jaccard index is reported for the liver, kidney and thoraco-abdominal wall, for all of the four investigated registration methods (HSO, HSI, EVO and EVI). The pre-registration DSC and Jaccard index is also included.

determinant from unity and moderate values of the curl magnitude. On the other hand, more prominent differences can be distinguished between EVO and EVI, in particular in terms of the Jacobian determinant of the estimated deformations, with EVI showcasing Jacobian determinants considerably closer to one than EVO. Both EVO and EVI have curl magnitude maps containing overall larger values compared to HSO and HSI, manifesting a tendency to increase in the proximity of organ-thoracic/abdominal/back wall interfaces. A peak in the curl magnitude maps can also be observed in the cranio-ventral part of the liver, peak which is coupled with increased registration errors.

In terms of the absolute registration errors, except for EVO, all of the algorithms showcase a good accuracy of 1 - 2 mm within the liver and kidney boundaries. The EVO algorithm, while showcasing predominately low errors within the organ boundaries, it has difficulties in estimating displacements at the interface between the liver and the thoracic/peritoneal wall, with the errors exceeding 5 mm. It can also be noted that all of the methods showcase, to different extents, difficulties in estimating the displacements of the thoracic and peritoneal wall, indicated both by the high registration errors and the mapped/registered images in the first row (the tag lines should be horizontal). Out of the four algorithms, it is EVO which has the best accuracy in this area, with EVI coming-in second best.

3.4. Calibration and computational performance of the investigated algorithms

The parameters α and Γ of the four investigated registration algorithms were configured according to the procedure described in section 2.3. Figure 8 showcases the calibration curves which were used to optimize the algorithm input parameters for the 3D data registered in the scope of this work. More precisely, figures 8(a) - (d) illustrate the dependency of the average post-registration DSC on the values of α and Γ . It can be observed that all curves showcase a global maximum, which was used as an indicator to determine the optimal values for α and Γ . Also, note that the DSC remains close to its

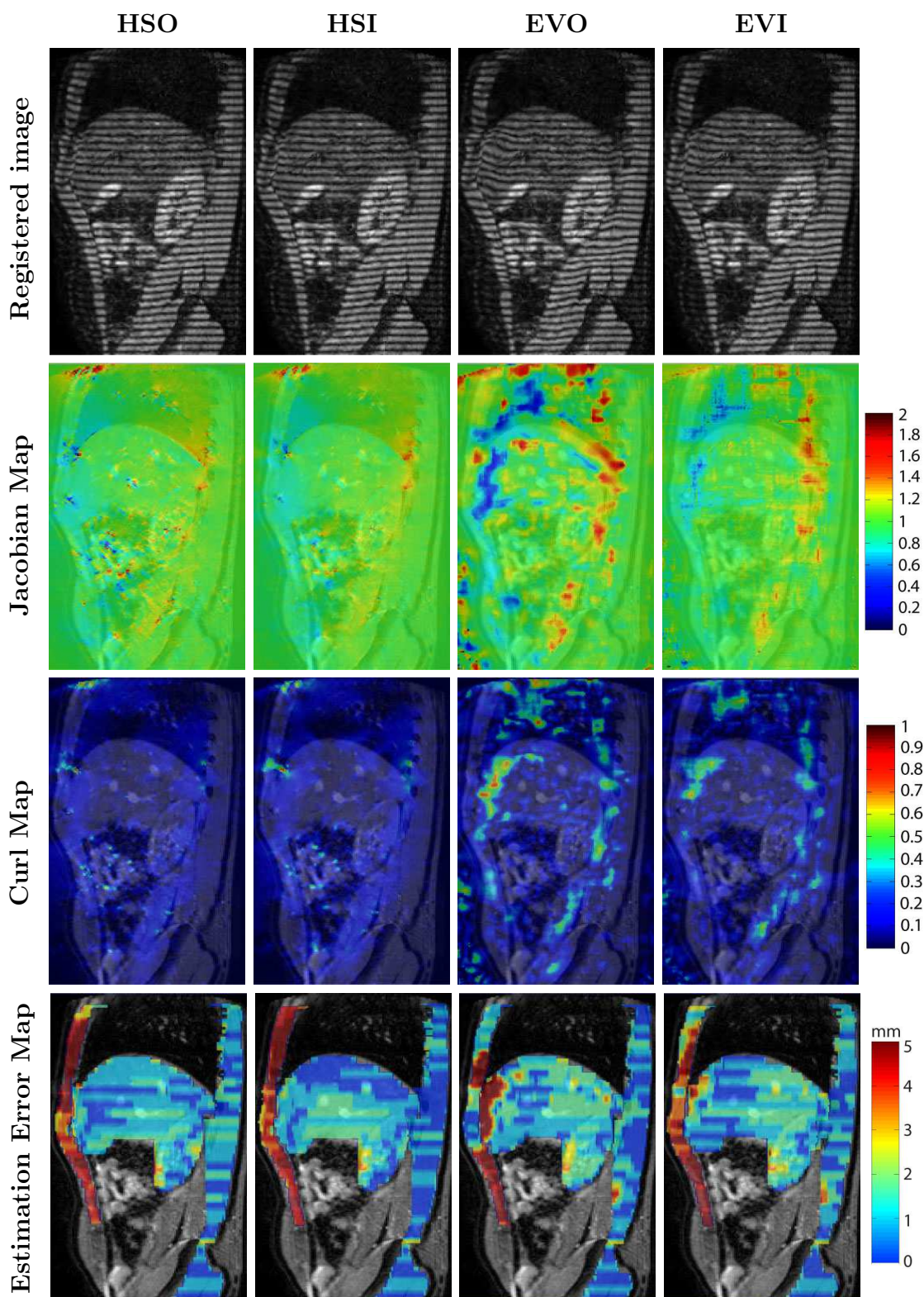


Figure 7: Outcome of the study conducted on the spin-tagged MR images. The four rows of images illustrate: the mapped/registered spin-tagged images, the Jacobian determinant maps, the curl magnitude maps and the spatial distribution estimation errors for each of the four registration methods.

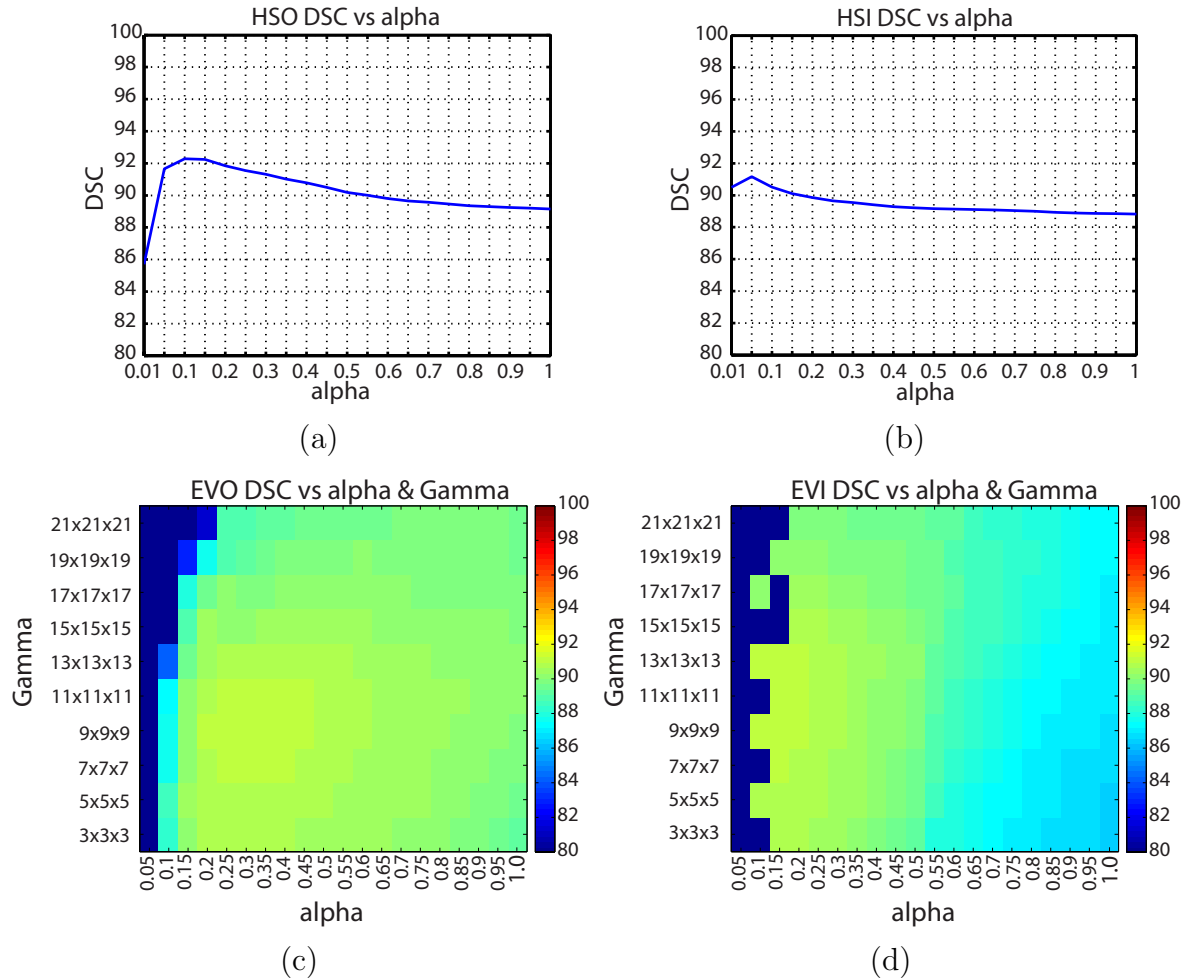


Figure 8: Calibration of the four algorithms for registering the 3D images analyzed in the scope of this work. The four figures showcase the dependency of the average post-registration DSC on the input parameters α and Γ for each of the algorithms. The DSC was averaged over all volumes of interest and over the first image pairs of each of the 3D data batches described in section 2.2.1.

maximum value for a rather wide range of values for α and Γ . On the other hand, for small values of the parameters the algorithms showcase a rapid decrease in performance. A similar analysis was also conducted in order to calibrate the algorithms for the image registration performed in the scope of the spin-tagging experiment. However, instead of the DSC, the average absolute error illustrated in the fourth row of figure 7 was used as an optimization criterion. For brevity, the resulting calibration curves are not included in the manuscript, since the curves showcase similar tendencies as the ones illustrated in figure 8. The optimal values for α and Γ , together with the resulting average DSC or average registration error (depending on the dataset), are reported in Table 6. Table 7 reports the average computational performance of the four registration methods for both 2D and 3D images. Reporting is performed for images of size 224×224 and

Method	$(\alpha, \Gamma, \overline{\text{Absolute Error}} [\text{mm}])$	$(\alpha, \Gamma, \overline{\text{DSC}})$
Horn-Schunck Original	(0.3, −, 1.74)	(0.1, −, 92)
Horn-Schunck Incompressible	(0.25, −, 1.67)	(0.05, −, 91)
EVolution Original	(0.25, 15×15 , 1.71)	(0.3, $11 \times 11 \times 11$, 91)
EVolution Incompressible	(0.15, 13×13 , 1.67)	(0.2, $11 \times 11 \times 11$, 91)

Table 6: Configuration of the input parameters for the four algorithms. The table reports the optimal α and Γ values, together with the resulting average absolute error for the 2D spin-tagged data and together with the average DSC for the 3D data.

$256 \times 256 \times 128$, which largely correspond to the dimensions of the 2D and 3D images registered in the scope of this work.

Method	2D	3D
	$[224 \times 224]$	$[256 \times 256 \times 128]$
Horn-Schunck Original	27 ms	2.7 s
Horn-Schunck Incompressible	105 ms	10 s
EVolution Original	150 ms	67 s
EVolution Incompressible	250 ms	65 s

Table 7: Computational time required by the four investigated registration methods.

4. Discussion

Evaluation criteria for the image registration algorithms

One fundamental problem that complicates the development as well as the evaluation and validation of different medical image registration algorithms is the ground truth dilemma: The true anatomical deformation of soft tissues can frequently only be assessed in contrast rich regions such as organ boundaries or in the vicinity of organ features, while the true physiological deformation in contrast devoid regions is not known. This poses a problem, since in a historical context, medical image registration developed as a side-branch of the more general field of computer vision. As a consequence, many established algorithms and evaluation criteria in the medical image registration domain as well as the associated benchmarking criteria are rooted/derived from methods developed for optical object tracking.

Three criteria that are often used for evaluating the performance of medical image registration algorithms are the Dice Similarity Coefficient (DSC), the Jaccard index and the target registration error (TRE). Note that the DSC and the TRE in particular have also been recommended as image registration validation criteria in the recently published AAPM report of Task Group 132 on the usage and quality assurance of image

registration methods in radiotherapy (Brock et al. 2017). Given the high 0.8 - 0.9 values obtained for the DSC and the Jaccard index on the five 3D data batches (see Table 2) and the 1.6 - 2.0 mm average TRE obtained on the DIR-Lab CT images (see Table 3), if a contour-based validation criterion would be used (either quantitative or qualitative such as visual inspection (Brock et al. 2017)), one would conclude that all four registration methods provide satisfactory deformations. Note that the obtained DSC, Jaccard index and TRE values are also in good correspondence with previous reportings (Kumarasiri et al. 2014, Rigaud et al. 2015, Li et al. 2017, Zachiu et al. 2015, Denis de Senneville et al. 2016, Zachiu et al. 2017c) and to a large extent within the tolerance limits recommended by the guidelines of AAPM report of Task Group 132 (Brock et al. 2017).

The most obvious way to evaluate whether the estimation of the tissue deformation in contrast-void regions corresponds to the true physiological deformation, is to repeat the observation of the deformation process and to introduce non-invasively contrast patterns across the entire image. MR-tagging is essentially an old and well established method, which has been used for many years in the field of cardiac MRI for this purpose. MR-tagging methods, such as C-SPAMM allow to saturate the MR magnetization of a tissue under observation prior to the motion event with a contrast pattern, such as lines or a grid, and to track the subsequent elastic deformation of this pattern. The principal limitations of this MR-tagging approach are two-fold: The tagging-pattern disappears with the longitudinal relaxation time of tissue under observation. This means the tag must be applied immediately before the motion event and - considering muscle tissue - can only be observed over a duration of approx 700 - 1000 ms. The second limitation arises from the fact that tagged liquids, such as for example arterial blood, exits the “tagged stripes” during the observation period and mixes with untagged tissue. This additionally reduces the contrast of the tag-pattern, in particular in high perfused organs such as liver and kidney. In the presented experiments, this effect is main the reason why the presented data was only MR-tagged in one dimension (i.e. a line pattern in the direction of the principal deformation).

Here, respiratory gated MR tagging has been used to evaluate the precision and accuracy of the four investigated registration methods in the abdomen under the assumption that the respiratory cycle is periodic. After registering the un-tagged pair of images, high post-registration values of the DSC and the Jaccard index were obtained for the liver, kidney and the thoraco-abdominal wall for all four of the algorithms. However, the error maps illustrated in the fourth row of figure 7 showcases precisely the areas in which registration was successful and areas which are problematic. While the tag lines deep inside the liver and the kidney do not evidence any obvious signs of misregistration, the thoracic wall and its interface with the liver appears to be particularly problematic for all four algorithms. This finding validates the hypothesis, that while the DSC and the Jaccard index are certainly part of a comprehensive assessment of the image registration, neither is sufficient to evaluate if the estimated deformation really corresponds to the physiological tissue deformation. However, while the tagged MR-images as the “gold-standard” permit a differential analysis of complex respiratory kinetics as shown

in figure 6, the untagged images neither display sufficient anatomical contrast between the different tissue types, nor sufficient contrast within the tissue layers to provide the required information to analyze the complex situation correctly. Similar to the employed motion estimation algorithms, even a careful visual inspection of the entire respiratory cycle as a movie does not allow human observers to analyze the underlying tissue deformations correctly, while C-SPAMM depicts the motion pattern clearly.

Unfortunately, not all physiological forms of motion and not all medical imaging technologies lend themselves to tagging techniques like (or similar to) C-SPAMM and therefore, for many applications, the ground truth dilemma essentially remains. As stated in the introduction, we propose as the next best measure to evaluate the quality of the registration results beyond the established indices such as DSC and/or Jaccard, to investigate the properties of the estimated motion vector fields for anatomical plausibility. Figure 2, 4, 5 and 7 display examples of the spatial distribution of the Jacobian determinant. While several parts of the human anatomy such as lung volume, heart volume, bladder volume, large parts of the digestive tract display volume changes during physiological motion processes/events, most of the inner organs, such as liver, spleen, kidney, muscle and fat layers are liquid filled tissue structures and as a consequence incompressible. Both HSO and EVO preserve to a large extent the overall volume of the liver and kidneys. However, within the organ interiors, the high deviations from unity of the Jacobian determinant are physically implausible, since the composing tissues are incompressible. Also, as shown in figures 4 and 5 such fluctuations correspond to sinks and/or sources in the estimated deformations, which basically indicates the creation and disappearance of tissue. This is aberrant given the material properties of the observed organs and the short time interval between the acquired images.

The counter-part of the Jacobian determinant is an analysis of the vorticity of the estimated deformations under consideration of the underlying local anatomy. Although not quite as intuitive to interpret as the former, variations in vorticity are generally an indicator for shearing motion, local rotations and - similar to the Jacobian determinants - divergence of the motion fields. As a quantitative measure for vorticity, the examples in figure 3 showcase the spatial distribution of the curl magnitude of the deformations estimated on the abdominal data sets by the four investigated registration methods. Moderate to high values of the curl magnitude can be predominately observed in the proximity of shearing interfaces (e.g. abdominal and back wall), as expected, but also within regions in which a high vorticity is unlikely from an anatomical point-of-view. Such a problematic area is highlighted in Fig. 5. Local peaks in the curl map of the EVO algorithm can be observed inside both the cranial half of the kidney and in the dorsal part of the liver. While rotations of the liver and kidney as a whole are expected, they are highly improbable at such fine scales. A more refined interpretation of the curl maps is provided in the associated motion fields, which showcase strong local swirls in areas of high curl magnitude. At the same time note that the peaks in the curl map are coupled to a large extent to deviations of the Jacobian determinant from unity, confirming the fact that local vorticity can also serve as a measure of motion field divergence.

While clearly inferior to local error maps derived from a “true gold standard” as shown in figure 7, both the local Jacobian determinant as well as the vorticity are useful indicators with respect to areas where the registration process has potentially failed. For example in the registration with EVO, the low Jacobian determinant as well as the increased vorticity in the ventral part of the liver corresponds well with the independently assessed area of misregistration depicted in the bottom row. For the incompressibility-constrained methods, however, the map of the local Jacobian determinant is only useful as an indicator if convergence of the functional has been achieved. Thus, the local vorticity maps become an increasingly important indicator of local misregistrations. For example, as shown in figure 7, the elevated local curl magnitude of 0.5 - 0.6 for EVI clearly indicates a small area in the cranio-ventral part of the liver, which as depicted by the independent error estimation, is not correctly registered, with registration errors of 3.5 - 5 mm.

An interesting alternative solution suggested by previous studies and also the report of AAPM Task Group 132, for offline validation of the deformations estimated by registration algorithms, is the usage of so-called “software phantoms” (Brock et al. 2017). This consists in inducing a set of known deformations to an image, followed by its registration to its non-deformed version via the algorithm under evaluation. The achievable accuracy and precision of the algorithm is then determined by comparing the estimated deformations to the ground truth deformations induced to the phantom. A challenge for using such an approach, however, is to simulate a set of physically accurate deformations, specific to the anatomy under observation. In order to mimic such deformations, one would have to have knowledge of the ensemble of forces which typically act on the anatomy, together with an extended set of material parameters specific to the composing tissue(s). This results in a complex simulation, which was circumvented in the current study by the proposed experiment on spin-tagged MR images, allowing direct observation of the liver and kidney anatomical deformations. As previously discussed, however, this approach has its own shortcomings and therefore future studies should also consider as a validation tool, the usage of anatomically accurate simulations conducted on software phantoms.

Performance of the original Horn-Schunck algorithm

While the HSO has been shown to be a reliable performer by several independent studies (Ostergaard et al. 2008, Roujol et al. 2010, Ries et al. 2010, Zachiu et al. 2015b, Glitzner et al. 2015, Yang et al. 2016), the assumption that a material point conserves its intensity as it moves, however, can also be a source of misregistrations (Ostergaard et al. 2008, Martel et al. 2007, Zachiu et al. 2015b). In the current work, this is best illustrated for the MR - MR mono-modal registration in figure 2 and 4. Several superimposed effects and artifacts including arterial pulsations, noise and transient fold-over artifacts of the bowels, have led to violations of the voxel intensity conservation hypothesis in the dorsal part of the liver and the kidney cortex. In its attempt to redistribute image intensities such that the data fidelity term of the HSO is minimized, the algo-

rithm ended-up estimating non-anatomical deformations which translate to local tissue expansion/creation. Similarly, for the CT-CT registration, the superior part of the lung is expanded by the algorithm due to a CT reconstruction artifact at the lung-liver interface, which varies due to a changing curvature during respiration. On the other hand, the smoothness constraint imposed by the regularization term has the tendency to dampen vortices, reason for which the curl magnitudes remain overall low to moderate. With respect to practical considerations, the HSO algorithm allows a very efficient implementation. This is due both to the point-wise nature of the numerical scheme employed for the minimization of the cost function and the overall mathematical simplicity of the optimization process. This resulted in the HSO having the best convergence time among all four investigated registration methods.

Performance of the Horn-Schunck algorithm with modified regularization

In terms of the DSC and the Jaccard index, HSI provides comparable values to the HSO. However, the HSI significantly dampens the effect of intensity variations not attributed to motion on the quality of the estimated deformations. As illustrated in figure 4, the Jacobian determinant in an area which was problematic for the HSO, remains close to unity in the entire region-of-interest, with the associated motion fields uniformly pointing in the general direction of motion of the observed organs. From a physical point-of-view, the incompressibility constraint imposes a lower penalty on shearing motion, which leads to an improved performance in the vicinity of organ/tissue boundaries. This effect can be best observed in figure 3, for CT - CT registration using HSI, where the curl magnitude showcases a tendency to increase in the proximity of the thoracic and abdominal wall. Note that this effect is less pronounced for HSO. On the other hand, during the experiment conducted on the tagged MR images, the differences between the HSO and the HSI methods are less noticeable. A reason for this may be that the registered images did not include violations of the intensity conservation hypothesis to an extent that would induce a differentiated behavior between the two algorithms (like, for example, in figure 2).

In terms of computational time, the more complex numerical scheme of the HSI implied a threefold penalty compared to the HSO. Nevertheless, the resulting computational latency is still acceptable in the context of online therapy guidance.

Performance of the original Evolution algorithm

The principal advantage of the original Evolution (EVO) algorithm compared to HSO and HSI lies in its capability of performing both mono and multi-modal registration. The algorithm demonstrated good contour matching capabilities, with DSC and Jaccard indices predominantly in the 0.8 - 0.9 range. This indicates that the original Evolution algorithm is well suitable for estimating deformations of volume-changing anatomical areas, such as the lung or the bladder (as demonstrated in (Denis de Senneville et al. 2016)). For incompressible organs, on the other hand, the EVO algorithm may be less favorable. In both CT and MR images, image intensity is the density

weighted signal average over a given voxel. Thus, from a physical point-of-view, image intensity in such images can be seen equivalent to mass. The HSO and HSI algorithms imposed the conservation of gray-level intensity and therefore the conservation of mass in between the registered images. The EVO data fidelity term loses this property, which leads to the algorithm resolving “boundary matching” by creating sinks or sources (and therefore to the disappearance and creation of mass) in contrast devoid regions, which is anatomically implausible in incompressible organs. This effect can be observed to different extents in all of the registered data sets, where strong deviations from unity of the Jacobian determinant are present in the homogeneous image regions (see for example the cranial part of the liver in figure 2). A more refined illustration of this tendency of aligning structure boundaries at the expense of creating sinks and sources in the estimated deformations is provided in figure 5.

In terms of the absolute registration error illustrated in figure 7, the EVO algorithm provided the best result among all four methods in registering the abdominal and the thoracic wall. This can be mostly attributed to the algorithm prioritizing the alignment of high contrast areas. However, the transition towards the contrast-devoid area is resolved with high divergences in the estimated deformations, leading to strong misregistrations in the ventral part of the liver. Nevertheless, if sufficient contrast is available within the images, the EVolution algorithm was demonstrated to be capable of high accuracy and precision in several independent studies (Denis de Senneville et al. 2016, Zachiu et al. 2017a, Zachiu et al. 2017b).

Concerning the required computational time, the EVO algorithm introduces a rather large relative penalty, compared to the HSO and HSI. This is mainly due to the time-consuming calculation of the EVO data fidelity term.

Performance of the EVolution algorithm with modified regularization

The EVI algorithm preserves the capability for performing cross-modality and cross-contrast co-registration, while restoring the volume conservation property lost by EVO. Since the data fidelity term is still aimed at aligning similar contrast patterns, the excellent edge alignment capabilities of EVO are maintained by EVI, providing mostly identical DSC and Jaccard indices. On the other hand, maximizing edge-alignment no longer leads to strong compressions and expansions of the tissues in homogeneous areas (see for example figure 5). At the same time, as showcased in both figure 3 and 5, the strong peaks in the curl magnitude maps are also attenuated, with a predominant tendency to increase in the proximity of sliding interfaces. There are, however, a few instances in which the EVI estimates deformations with a moderate magnitude of the curl within the organ boundaries. An example can be seen in the ventral part of the liver in the MR - MR mono-modal dataset (figure 3) and in the cranio-ventral part of the liver in figure 7. Noteworthy for the latter is the fact that there is also good spatial correspondence between an increase in the curl magnitude and of the absolute registration error in the cranio-ventral part of the liver in figure 7.

Overall, for cross-contrast or cross-image modality registration of soft-tissue structures,

EVI provided superior results compared to EVO, with no penalty in terms of computational latency.

Dependency of algorithm performance on the input parameters

All four of the investigated registration algorithms required the optimization of the parameter α , weighting the contributions of the data fidelity and the regularization terms within the cost functions. Additionally, EVO and EVI also required establishing the neighborhood size Γ . For real-time and online applications, it is of particular interest that the values of the algorithm parameters are selected in advance, in order to avoid latencies associated to a potential parameter re-optimization. In this sense, two aspects are important: the sensitivity of the algorithm performance to deviations of the input parameters from their optimal values and the re-usability of the algorithm parameters across different image pairs. Concerning the first aspect, it can be observed from figure 8 that all algorithms showcase a tendency to perform rather poorly at low values of α and Γ , followed by a global optimum at intermediate values and a steady tendency to deteriorate towards high values of the parameters. Note that the performance of the algorithms remains close to the optimum for a rather wide range of α and Γ values. However, given the rapid drop in performance of the algorithms for low α and Γ values, it is preferable to slightly over-estimate them in practice.

In terms of re-usability of the values of α and Γ across different datasets, the parameter optimization process for the 3D data only made use of the first image pair of each of the five data batches. For all the remaining 3D image pairs, the same values were maintained for α and Γ , with consistent algorithm performance. A difference in the optimal value can be observed, however, between the 3D and the spin-tagged 2D datasets (see table 6). Therefore the results indicate that, in practice, as long as the acquisition sequence parameters do not change considerably, the algorithm parameters do not require re-calibration. This observation is also in good correspondence with previous studies (Zachiu et al. 2015, Zachiu et al. 2015b, Denis de Senneville et al. 2016, Zachiu et al. 2017a, Zachiu et al. 2017b).

Selection of a motion estimation model for image-based therapy guidance in moving organs

The selection of a motion estimation model or registration algorithm should primarily depend on the particularities of the application and the intended usage. For example, if mapping of organ labels/segmentations between images is of interest, all four of the algorithms discussed in the current work would be suited for such a task (since all four demonstrated good contour matching capabilities). On the other hand, if the aim is to map quantitative data such as radiation dose, contrast agent absorption, diffusion/perfusion coefficients, etc., the precise displacement of the tissues within the organ boundaries becomes particularly important. In such applications, the occurrence of non-anatomical sinks and sources within the estimated deformations, would lead to a set of inaccurate post-registration quantitative measurements.

The current study showcases how adapting a branch of registration methods, such as variational algorithms, for a particular image modality/contrast and/or a particular anatomical location, may work towards the detriment of a different registration scenario. For example, in order to address the inability of the Horn-Schunck optical flow formulation for cross-contrast and cross-modality registration, the gray-level intensity conservation constraint was replaced with a more relaxed constraint on the conservation of contrast, imposed by the original EVOlution algorithm. This, however, leads to the loss of the mass conservation property, imposed by the data fidelity term of the Horn-Schunck algorithm, thus decreasing the penalty on the formation of sinks and sources in the estimated motion field. This property is restored by the EVI by replacing the smoothness constraint imposed by the regularization term, with a penalty on deviations of the deformation Jacobian determinant from unity. At a first glance, after addressing these shortcomings, it may seem that the EVI method is best among all of the four investigated algorithms. One has to take into consideration, however, that the incompressibility constraint is optimal only for anatomical structures which possess this particular physical property. In case the tracked anatomy changes volume in between the images to be registered, EVO would be the better option. Therefore, it would be overall preferable that such variational registration methods are selected depending on the specific application, acquisition modalities, acquisition parameters, image contrasts and the physical properties of the tracked anatomy.

5. Conclusion

Current validation procedures for medical image registration algorithms in a clinical setting, often rely on visual inspection and/or contour-based criteria. The current work demonstrates that this is likely insufficient to guarantee the anatomical plausibility of the estimated deformations. Here, three additional criteria were proposed, which can assist such evaluations on a voxel-by-voxel basis: the local Jacobian determinant of the deformations, the local vorticity of the deformations and a criterion relying on MR-tagging. Four image registration algorithms were evaluated against these criteria: two pre-existing and two proposed in the scope of this study. The latter two replaced the smoothness constraint on the estimated deformations imposed by the former two with a constraint on compressibility. It was demonstrated that while both the original and incompressible methods provide high values of the DSC and the Jaccard index (0.8 - 0.9) and low average TRE (1.6 - 2.0 mm), the pre-existing methods may estimate physiologically inaccurate compressions and expansions inside the liver and kidney. The proposed methods significantly reduced such effects, without a negative impact on deformation vorticity, estimation accuracy determined from MR-tagged images or increase in computational latencies. While improvements were obtained for incompressible tissues, the current study also demonstrates that each optimization of variational registration algorithms represents a compromise, which should be chosen consciously while considering the particular anatomical and physiological properties of

the tissue under observation.

Acknowledgments

This work was supported by ITEA 3, project no. 16016 (STARLIT).

References

- Bistoquet, A., Oshinski, J. & Škrinjar, O. (2008). Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model, *Med Image Anal* **12**(1): 69 – 85.
- Black, M. J. & Anandan, P. (1991). Robust dynamic motion estimation over time, *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 296–302.
- Brock et al., K. K. (2010). Results of a multi-institution deformable registration accuracy study (MIDRAS), *Int J Radiat Oncol Biol Phys* **76**(2): 583 – 596.
- Brock, K. K., Mutic, S., McNutt, T. R., Li, H. & Kessler, M. L. (2017). Use of Image Registration and Fusion Algorithms and Techniques in Radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132, *Med Phys* **44**(7): e43 – e76.
- Castillo, E., Castillo, R., Martinez, J., Shanoy, M. & Guerrero, T. (2010). Four-dimensional deformable image registration using trajectory modeling, *Phys Med Biol* **55**(1): 305 – 327.
- Denis de Senneville, B., Moonen, C. & Ries, M. (2015). MRI-Guided HIFU Methods for the Ablation of Liver and Renal Cancers, in Escoffre, J. - M. and Bouakaz, A. (ed.), *Therapeutic Ultrasound*, Springer, pp. 43 – 63.
- Denis de Senneville, B., Zachiu, C., Ries, M. & Moonen, C. (2016). EVolution: An Edge - Based Variational Method for Non - Rigid Multi - Modal Image Registration, *Phys Med Biol* **61**(20): 7377 – 7396.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species, *Ecology* **26**(3): 297 – 302.
- Dupuy, D. E., McCabe, M. & McMullen, W. N. (2013). *Image - Guided Cancer Therapy: A Multidisciplinary Approach*, Springer.
- Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., Shankar, L., Dodd, L., Kaplan, R., Lacombe, D. & Verweij, J. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), *Eur J Cancer* **45**(2): 228 – 247.
- Fanchon, L. M., Apte, A., Schmidlein, C. R., Yorke, E., Hu, Y. C., Dogan, S., Hatt, M., Visvikis, D., Humm, J. L., Solomon, S. B. & Kirov, A. S. (2017). Evaluation of the tumor registration error in biopsy procedures performed under real-time PET/CT guidance, *Med Phys* **44**(10): 5089 – 5095.
- Fass, L. (2008). Imaging and cancer: A review, *Mol Oncol* **2**(2): 115 – 152.
- Fischer, S. E., McKinnon, G. C., Maier, S. E. & Boesiger, P. (1993). Improved myocardial tagging contrast, *Magn Res Med* **30**(2): 191 – 200.
- Glitzner, M., Denis de Senneville, B., Lagendijk, J. J., Raaymakers, B. W. & Crijs, S. P. (2015). On-line 3D Motion Estimation Using Low Resolution MRI, *Phys Med Biol* **60**(16): N301 – N310.
- Haber, E. & Modersitzki, J. (2004). Numerical methods for volume preserving image registration, *Inverse Problems* **20**(5): 1261.
- Haber, E. & Modersitzki, J. (2006). Image registration with guaranteed displacement regularity, *Int J Radiat Oncol Biol Phys* **71**(3): 361 – 372.
- Heinrich, M., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F., Brady, S. & Schnabel, J. (2012). MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration, *Med Image Anal* **16**(7): 1423–1435.

- Hill, A., Mehnert, A., Crozier, S. & McMahon, K. (2009). Evaluating the accuracy and impact of registration in dynamic contrast-enhanced breast MRI, *Concepts in Magnetic Resonance Part B (Magnetic Resonance Engineering)* **35B**(2): 106 – 120.
- Hill, D., Batchelor, P., Holden, M. & Hawkes, D. (2001). Medical image registration, *Phys Med Biol* **46**(3): R1–R45.
- Horn, B. P. & Schunck, B. G. (1981). Determining the Optical Flow, *Artificial Intelligence* **17**(1 - 3): 185 – 203.
- Ibrahim, E.-S. H. (2011). Myocardial tagging by cardiovascular magnetic resonance: Evolution of techniques—pulse sequences, analysis algorithms, and applications, *J Cardiovasc Magn Reson* **13**(1): 36.
- Jaccard, P. (1912). The distribution of the flora in the alpine zone, *New Pathologist* **11**(2): 37 – 50.
- Keall, P. J., Mageras, G. S., Balter, J. M., Emery, R. S., Forster, K. M., Jiang, S. B., Kapatoes, J. M., Low, D. A., Murphy, M. J., Murray, B. R., Ramsey, C. R., van Herk, M. B., Vedam, S. S., Wong, J. W. & Yorke, E. (2006). The Management of Respiratory Motion in Radiation Oncology Report of AAPM Task Group 76, *Med. Phys* **33**(10): 3874 – 3900.
- Kumarasiri, A., Siddiqui, F., Liu, C., Yechieli, R., Shah, M., Pradhan, D., Zhong, H., Chetty, I. J. & Kim, J. (2014). Deformable image registration vased automatic CT-to-CT contour propagation for head and neck adaptive radiotherapy in the routine clinical setting, *Med Phys* **41**(12): 121712.
- Lei, Y., Jinzong, L. & Dongdong, L. (2007). Discontinuity - preserving optical flow algorithm, *J Sys Eng Electronics* **18**(2): 347 – 354.
- Li, X., Zhang, Y., Shi, Y., Wu, S., Xiao, Y., Gu, X., Zhen, X. & Zhou, L. (2017). Comprehensive evaluation of ten deformable image registration algorithms for controu propagation between ct and cone-beam ct images in adaptive hean & neck radiotherapy, *PLoS One* **12**(4): e0175906.
- Maintz, J. B. & Viergever, M. A. (1998). A Survery of Medical Image Registration, *Med Image Anal* **2**(1): 1 – 36.
- Mani, V. & Arivazhagan, S. (2013). Survey of Medical Image Registration, *J Biomed Eng Technol* **1**: 8 – 25.
- Mansi, T., Pennec, X., Sermesant, M., Delingette, H. & Ayache, N. (2011). iLogDemons: A Demons - Based Registration Algorithm for Tracking Incompressible Elastic Biological Tissues, *Int J Comp Vis* **92**(1): 92 – 111.
- Martel, A. L., Froh, M. S., Brock, K. K., Plewes, D. B. & Barber, D. C. (2007). Evaluating an optical-flow-based registration algorithm for contrast-enhanced magnetic resonance imaging of the breast, *Phys Med Biol* **52**(13): 3803 – 3816.
- Maurer, C., Fitzpatrick, J., Wang, M., Galloway, R., Maciunas, R. & Allen, G. (1997). Registration of head volume images using implantable fiducial markers, *IEEE Transactions on Medical Imaging* **16**: 447–462.
- Monzón, N., Salgado, A. & Sánchez, J. (2016). Robust discontinuity - preserving optical flow methods, *IPOLE Journal* **6**: 165 – 182.
- Mundt, A. J. & Roeske, J. C. (2011). *Image-Guided Radiation Therapy: A Clinical Perspective*, PMPH - USA.
- Nagel, H. H. (1990). Extending the ‘oriented smoothness constraint’ into the temporal domain and the estimation of derivatives of optical flow, in O. Faugeras (ed.), *Computer Vision — ECCV 90: First European Conference on Computer Vision Antibes, France, April 23–27, 1990 Proceedings*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 139–148.
- Nessi, P. (1993). Variational approach to optical flow estimation managing discontinuities, *Imag Comp Vis* **11**(7): 419 – 439.
- Osman, A. B., Ovinis, M., Osei, H. & Hashim, F. M. (2016). A Review of Optical Flow Models Applied for Fluid Motion Estimation, *ARPN J Eng App Sci* **11**(20): 12181 – 12187.
- Ostergaard, N. K., Denis de Senneville, B., Elstrom, U. V., Tanderup, K. & Sorensen, T. S. (2008). Acceleration and validation of optical flow based deformable registration for image-guided radiotherapy, *Acta Oncol* **47**(7): 1286–1293.

- Ries, M., Denis de Senneville, B., Roujol, S., Berber, Y., Quesson, B. & Moonen, C. (2010). Real-time 3D target tracking in MRI guided focused ultrasound ablations in moving tissues, *Magnetic Resonance in Medicine* **64**: 1704 – 1712.
- Rigaud, B., Simon, A., Castelli, J., Gobeli, M., Ospina Arango, J.-D., Cazoulat, G., Henry, O., Haigron, P. & de Crevoisier, R. (2015). Evaluation of deformable image registration methods for dose monitoring in the head and neck radiotherapy, *Biomed Res Int* **2015**: 726268.
- Rivaz, H., Karimaghadoo, Z. & Collins, D. L. (2014). Self - similarity weighted mutual information: A new nonrigid image similarity metric, *Med Image Anal* **18**(2): 343 – 358.
- Roach 3rd, M., Alberini, J. L., Pecking, A. P., Testori, A., Verrecchia, F., Soteldo, J., Ganswindt, U., Joyal, J. L., Babich, J. W., Witte, R. S., Unger, E. & Gottlieb, R. (2011). Diagnostic and therapeutic imaging for cancer: Therapeutic considerations and future directions, *J Surg Oncol* **103**(6): 578 – 601.
- Rohlfing, T., Maurer, C. R. J., Bluemke, D. A. & Jacobs, M. A. (2003). Volume - Preserving Nonrigid Registration of MR Breast Imaged Using Free - Form Deformation with an Incompressibility Constraint, *IEEE Trans Med Imaging* **22**(6): 730 – 741.
- Rose et al., S. E. (2001). MRI based diffusion and perfusion predictive model to estimate stroke evolution, *Magn Reson Imaging* **19**(8): 1043 – 1053.
- Roujol, S., Ries, M., Quesson, B., Moonen, C. & Denis de Senneville, B. (2010). Real - Time MR - Thermometry and Dosimetry for Interventional Guidance on Abdominal Organs, *Magn Reson Med* **63**(4): 1080 – 1087.
- Screibmann, E., Pantalone, P., Waller, A. & Fox, T. (2012). A measure to evaluate deformable registration fields in clinical settings, *J Appl Clin Med Phys* **13**(5): 126 – 139.
- Sotiras, A., Davatzikos, C. & Paragios, N. (2013). Deformable medical image registration: A survey, *IEEE TMI* **32**(7): 1053 – 1090.
- Stout, D. A., Bar-Kochba, E., Estrada, J. B., Toyjanova, J., Kesari, H., Reichner, J. S. & Franck, C. (2016). Mean deformation metrics for quantifying 3D cell-matrix interactions without requiring information about matrix material properties, *PNAS* **113**(11): 2898 – 2903.
- Terzopoulos, D. (1986). Image analysis using multigrid relaxation methods, *IEEE Trans Pattern Analysis and Machine Learning PAMI - 8*(2): 129 – 139.
- Timmerman, R. D. & Xing, L. (2009). *Image-Guided and Adaptive Radiation Therapy*, Wolters Kluwer — Lippincott Williams & Wilkins.
- Vese, L. A. & Le Guyader, C. (2016). *Variational Methods on Image Processing*, CRC Press.
- Weickert, J., Bruhn, A., Papenberg, N. & Brox, T. (2003). Variational optic flow computation: From continuous models to algorithms, *International Workshop on Computer Vision and Image Analysis (ed. L. Alvarez), IWCVIA03, Las Palmas de Gran Canaria*.
- Weickert, J. & Schnör, C. (2001a). A theoretical framework for convex regularizers in PDE-based computation of image motion, *Int J Comp Vis* **45**(3): 245 – 264.
- Weickert, J. & Schnör, C. (2001b). Variational optic flow computation with a spatio - temporal smoothness constraint, *J Math Imag Vis* **14**(3): 245 – 255.
- Wilczewska, A., Cygan, S. & Żmigrodzki, J. (2017). Displacement field estimation for echocardiography strain imaging using b-spline based elastic image registration: Synthetic data study, in R. Jabłoński & R. Szewczyk (eds), *Recent Global Research and Education: Technological Challenges: Proceedings of the 15th International Conference on Global Research and Education Inter-Academia 2016*, Springer International Publishing, pp. 309 – 315.
- Xing, F., Woo, J., Gomez, A. D., Pham, D. L., Bayly, P. V., Stone, M. & Prince, J. L. (2017). Phase Vector Incompressible Registration Algorithm for Motion Estimation From Tagged Magnetic Resonance Images, *IEEE Trans Med Imaging* **36**(10): 2116 – 2128.
- Yang, D., Li, H., Low, D. A., Deasy, J., O. & El Naqa, I. (2016). A Fast Inverse Consistent Deformable Image Registration Method Based on Symmetric Optical Flow Computation, *Phys Med Biol* **53**(21): 6143 – 6165.
- Yang, Q., Parvin, B. & Mariano, A. (2000). Singular features in sea surface temperature data,

- Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 1, pp. 516 – 520.
- Yin, Y., Hoffman, E. A. & Lin, C. L. (2009). Mass Preserving Nonrigid Registration of CT Lung Images Using Cubic B - Spline, *Med Phys* **36**(9): 4213 – 4222.
- Zachiu, C., Denis de Senneville, B., Dmitriev, I. D., Moonen, C. T. W. & Ries, M. (2017c). A framework for continuous target tracking during MR-guided high intensity focused ultrasound thermal ablations in the abdomen, *J Ther Ultrasound* **5**: 27 – 45.
- Zachiu, C., Denis de Senneville, B., Moonen, C. & Ries, M. (2015). A framework for the correction of slow physiological drifts during mr-guided hifu therapies: Proof of concept, *Med Phys* **42**(7): 4137 – 4148.
- Zachiu, C., Denis de Senneville, B., Tijssen, R. H. N., Kotte, A. N. T. J., Houweling, A. C., Kerkmeijer, L. G. W., Lagendijk, J. J. W., Moonen, C. T. W. & Ries, M. G. (2017b). Non - rigid ct/cbct to cbct registration for online external beam radiotherapy guidance, *Phys Med Biol*, [Epub ahead of print] .
- Zachiu, C., Papadakis, N., Ries, M., Moonen, C. & Denis De Senenville, B. (2015b). An improved optical flow tracking technique for real - time mr - guided beam therapies in moving organs, *Phys Med Biol* **60**(23): 9003 – 9029.
- Zachiu, C., Ries, M., Ramaekers, P., Guey, J.-L., Moonen, C. T. W. & Denis de Senneville, B. (2017a). Real-time non-rigid target tracking for ultrasound - guided clinical interventions, *Phys Med Biol* **62**(20): 8154 – 8177.
- Zitová, B. & Flusser, J. (2003). Image Registration Methods: A Survey, *Image and Vision Computing* **21**(11): 977 – 1000.