

PARALLEL AND MEMORY EFFICIENT MULTIMODAL IMAGE REGISTRATION FOR RADIOTHERAPY USING NORMALIZED GRADIENT FIELDS

Lars König, Alexander Derksen, Marc Hallmann, Nils Papenberg

Fraunhofer MEVIS, Project Group Image Registration, Lübeck, Germany

ABSTRACT

We introduce a new highly parallel and memory efficient deformable image registration algorithm to handle challenging clinical applications. The algorithm is based on the normalized gradient fields (NGF) distance measure and Gauss-Newton numerical optimization. By carefully analyzing the mathematical structure of the problem, a matrix-free Hessian-vector multiplication for NGF is derived, giving a highly integrated formulation. Embedding the new scheme in a full, non-linear image registration algorithm enables fast calculations on high resolutions with dramatically reduced memory consumption. The new approach provides linear scalability compared with a traditional sparse-matrix-based scheme. The algorithm is evaluated on a challenging problem from radiotherapy, where pelvis cone-beam CT and planning CT images are registered.

Speedups up to a factor of 149.3 for a single Hessian-vector multiplication and of 20.3 for a complete non-linear registration are achieved.

Index Terms— image registration, computational efficiency, parallel algorithms, radiation therapy

1. INTRODUCTION

In radiotherapy, efficient and fast deformable image registration is a key requirement in many areas of application, e.g. for exact contour propagation [1] or dose reconstruction [2]. As often different image sources are involved, multimodal registration schemes are necessary. Additionally, high accuracy is often needed. For example in prostate cancer treatment, strong local variations in pelvic images require high deformation resolutions and accurate numerical schemes. However, at these resolutions especially complex multimodal registration schemes require high amounts of memory and parallelizability is limited.

In this work, we present a novel formulation for a matrix-free optimization scheme using the normalized gradient fields distance measure (NGF) [3] which is well suited for multimodal registration problems. The new algorithm uses a Gauss-Newton optimization scheme that is especially suited

for least-squares problems [4]. This scheme features a quadratic approximation to the Hessian at every optimization step, allowing for fast convergence and high accuracy. By performing a careful analysis of the Hessian matrix structure, problem-adapted formulations for NGF Hessian-vector multiplications are derived, exploiting known problem structure. These formulations allow direct computations without intermediate storage of sparse matrices or interdependencies, allowing for an extremely memory efficient and fully parallel computation. Thereby, the presented approach enables the calculation of fine deformation resolutions in radiotherapy settings where high accuracy is crucial.

2. RELATED WORK

Many efforts have been made to obtain fast and efficient image registration algorithms and the presented approaches vary in an extreme bandwidth. A general overview is given in [5]. While some methods focus on the application of specialized hardware like DSPs [6, 7] or GPUs (see e.g. [8, 9, 10, 11, 12]), others are using shared-memory [13] or multi-core architectures [14] to decrease algorithm runtime. Methods that involve algebraic analysis of the problem structure to obtain fast and efficient mathematical schemes have been presented in [15, 16, 17]. While image registration with gradient based optimization schemes has been researched in [16, 17], none of those works has extended the derived schemes to involve Hessian matrix multiplications for deformable registration. The increasing complexity of those computations challenges the presented concepts, although benefits in runtime, memory consumption and numerical convergence can be expected [4].

3. METHOD

First the general registration framework is described. Afterwards, the problem specific, parallel and efficient computation for the NGF Hessian-vector multiplication will be derived in detail.

3.1. Image registration

In image registration, the goal is to find a transformation Y , that aligns a moving template image $T(Y)$ to a fixed ref-

This work was partly funded by the European Regional Development Fund (EFRE).

erence image R . In the case of deformable image registration, the transformation Y is in general considered as a three dimensional vector field. With the framework described in [18], using a distance measure $D(T(Y), R)$ and a regularizer $S(Y)$, the registration problem is modeled as an optimization problem $\{J(Y) := D(T(Y), R) + \alpha S(Y)\} \rightarrow \min$, where both images are assumed to be vectorized representations of $M \times N \times O$ sized three dimensional images, yielding $Y \in \mathbb{R}^{3MNO}$.

Here, D measures the distance of the deformed template image $T(Y) \in \mathbb{R}^{MNO}$ and reference image $R \in \mathbb{R}^{MNO}$. As Y contains a vector valued deformation for each point, the optimization problem is highly ill posed and thus needs to be regularized, which is achieved using the regularizer $S(Y)$ penalizing physically implausible deformations. The parameter $\alpha > 0$ is a weighting factor that balances between image distance and deformation regularity. The evaluation of $T(Y)$ is realized with trilinear interpolation.

The optimization problem is solved using derivative based numerical optimization, where iteratively in each step k a descent search direction u_k is computed solving

$$-\nabla^2 J_k u_k = \nabla J_k, \quad (1)$$

where ∇J_k is the gradient and $\nabla^2 J_k$ is an approximated Hessian of the objective function J at step k [4].

As described later on, the specific choices for distance measure and regularizer lead to a registration approach that is a least squares problem $J = \frac{1}{2} \hat{r}^\top \hat{r}$, where \hat{r} is a residual term to be determined later on.

A well suited optimization method for a least squares problem is the so-called Gauss-Newton optimization [4]. This scheme exploits the specific properties of this problem type by calculating a quadratic approximation to the Hessian $\nabla^2 J \approx d\hat{r}^\top d\hat{r} =: \hat{H}$ at each step, where $d\hat{r}$ is the Jacobian of the residual function \hat{r} . This formulation discards the computation of dense second order derivative parts of $\nabla^2 J$ and guarantees a symmetric positive semi-definite \hat{H} yielding a guaranteed descent direction u_k [4].

Commonly for solving (1), iterative solvers such as the conjugate gradient (CG) method are used [4]. As the size of $\hat{H} \in \mathbb{R}^{3MNO \times 3MNO}$ can be quite large, even at low resolutions sparse-matrix-based approaches may require extensive amounts of computation time and memory [18]. Fast computation of (1) using the CG method heavily depends on the computation of Hessian-vector multiplications $\hat{H}x = y$. Here, an efficient implementation can significantly increase algorithm performance. This requires an integrated approach, merging all the components described above to a single, parallelizable formula.

3.2. Problem specific calculation scheme

As has been described in Section 1, for D we use the NGF distance measure. To ensure smooth and plausible deforma-

tions, regularization is performed by using the curvature regularizer [19], which has been proven successful in a wide range of different applications [20, 21, 22]. An efficient gradient computation for the curvature regularizer has already been derived in [16]. As the curvature regularizer can be interpreted as a quadratic function, this efficient gradient computation also represents a matrix-vector multiplication with the Hessian and thus can directly be incorporated. The Hessian multiplication of the NGF distance measure, however, is far more complex and will be analyzed in detail in the following.

With $\|\cdot\|_\varepsilon = \sqrt{\langle \cdot, \cdot \rangle + \varepsilon^2}$, the NGF as defined in [17] can be written as

$$\begin{aligned} D_{\text{NGF}}(Y) &= \frac{\bar{h}}{2} \sum_{i=1}^{MNO} \left(1 - \left(\frac{\langle \nabla T_i(Y), \nabla R_i \rangle + \tau \varrho}{\|\nabla T_i(Y)\|_\tau \|\nabla R_i\|_\varrho} \right)^2 \right) \\ &=: \frac{\bar{h}}{2} \sum_{i=1}^{MNO} (1 - r_i^2) \end{aligned}$$

where τ, ϱ are parameters which allow for filtering of noise, $\bar{h} \in \mathbb{R}$ is the product of grid spacings in each direction and $\nabla T_i, \nabla R_i \in \mathbb{R}^3$ are approximations to the image gradients at index i using short finite differences as in [17].

Using these definitions and the reduction function $\psi : \mathbb{R}^{MNO} \rightarrow \mathbb{R}$, $(r_1, \dots, r_{MNO})^\top \mapsto \frac{\bar{h}}{2} \sum_{i=1}^{MNO} (1 - r_i^2)$, the NGF calculation can be formulated as a concatenation of functions

$$D_{\text{NGF}} : \mathbb{R}^{3MNO} \rightarrow \mathbb{R}, \quad Y \mapsto \psi(r(T(Y))).$$

This formulation enables the computation of the NGF gradient using the chain rule as $\nabla D_{\text{NGF}}(Y) = \frac{\partial \psi}{\partial r} \frac{\partial r}{\partial T} \frac{\partial T}{\partial Y}$, with $\frac{\partial r}{\partial T} \in \mathbb{R}^{MNO \times MNO}$ and $\frac{\partial T}{\partial Y} \in \mathbb{R}^{MNO \times 3MNO}$.

Abbreviating $\frac{\partial r}{\partial T} =: dr$, using the quadratic Gauss-Newton approximation yields

$$\nabla^2 D_{\text{NGF}}(Y) \approx H := \bar{h} \frac{\partial T}{\partial Y}^\top dr^\top dr \frac{\partial T}{\partial Y}.$$

As described in detail in [16], dr has a sparse band diagonal matrix structure with up to seven diagonals. In column dr_i there are seven nonzero elements with offset $m \in \mathcal{M}$ from the main diagonal element. Using $\mathcal{M} := (-MN, -M, -1, 0, +1, +M, +MN)$ these can be directly computed from the input data using the matrix-free computations in [16]. This can now be used to derive a matrix-free variant of the Hessian-vector multiplication. Let

$$\mathcal{N} := \left\{ m_1 + m_2 \mid m_1 \in \mathcal{M}, m_2 \in \mathcal{M} \right\}.$$

This set can be described as follows. Considering the i -th column of dr , abbreviating $dr_i =: v_i$, the matrix product $dr^\top dr$ consists of scalar products $\langle v_i, v_k \rangle$ of all combinations of $i, k = 1, \dots, MNO$. As can be seen in Figure 1 and is described in [16], v_i and v_k exhibit the same shifted

$j \in \mathcal{N}$	η_j	$\eta_j + j$
$-2MN$	(MN)	$(-MN)$
$-MN - M$	(M, MN)	$(-MN, -M)$
$-MN - 1$	$(1, MN)$	$(-1, -MN)$
$-MN$	$(0, MN)$	$(-MN, 0)$
$-MN + 1$	$(-1, MN)$	$(-MN, 1)$
$-MN + M$	$(-M, MN)$	$(-MN, M)$
$-2M$	(M)	$(-M)$
$-M - 1$	$(1, M)$	$(-M, -1)$
$-M$	$(0, M)$	$(-M, 0)$
$-M + 1$	$(-1, M)$	$(-M, 1)$
-2	(1)	(-1)
-1	$(0, 1)$	$(-1, 0)$
0	\mathcal{M}	\mathcal{M}

Table 1. First column: offsets j , where $\langle v_i, v_{i+j} \rangle \neq 0$; second column: offsets of nonzero elements of v_i contributing to $\langle v_i, v_{i+j} \rangle$; third column: offsets of nonzero elements of v_{i+j} contributing to $\langle v_i, v_{i+j} \rangle$. Because of symmetry, the table shows only the first 13 indices, remaining indices are equivalent with negative sign.

pattern of nonzero elements due to the band matrix structure of dr with seven nonzero diagonals. Calculating the scalar product $\langle v_i, v_{i+j} \rangle$, especially those offset indices j are of interest, where $\langle v_i, v_{i+j} \rangle \neq 0$. Choosing different indices j results in a shift of the nonzero pattern by j elements. While naive computation gives $7 \cdot 7 = 49$ possible indices j where $\langle v_i, v_{i+j} \rangle \neq 0$, closer examination considering \mathcal{N} yields only 25 elements, which are shown in the first column in Table 1 (because of symmetry only 13 rows are shown). Let η_j be the vector of offsets j of nonzero elements of v_i contributing to $\langle v_i, v_{i+j} \rangle$, shown in the second column in Table 1. Then, we define $v_i^{\eta_j} := v_i(i + \eta_j)$ as the vector of elements at indices $i + \eta_j$ (shown in Table 1, third column), with element wise addition. A single element of $dr^\top dr$ can now be written as

$$(dr^\top dr)_{i,i+j} = \begin{cases} \langle v_i^{\eta_j}, v_{i+j}^{\eta_j} \rangle & \text{if } j \in \mathcal{N}, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

This means that every nonzero element in (2) is composed of a scalar product of only two short vectors with one, two or (for $j = 0$) seven elements. This new formulation tremendously decreases memory consumption and computational costs. Additionally, (2) implies that a single row of $dr^\top dr$ contains at most 25 nonzero elements, and their column indices are exactly known. Instead of naively calculating MNO scalar products with MNO summands each for a single row of $dr^\top dr$, only 25 scalar products with at most seven summands each need to be considered.

Extending this formulation to include the left and right side multiplications of $\frac{\partial T}{\partial Y}$, we obtain a block matrix structure, where $dr^\top dr$ is weighted with combinations of all three directional derivatives in $\frac{\partial T}{\partial Y}$, see Figure 1.

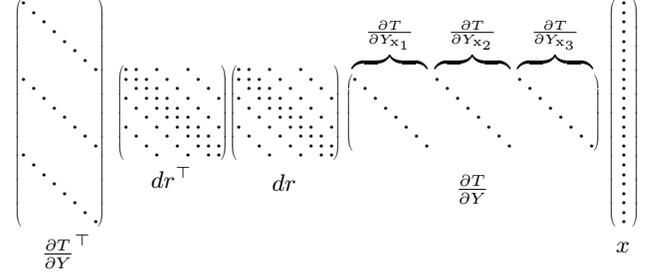


Fig. 1. Schematic view of Hessian-vector multiplication $Hx = \bar{h} \left(\frac{\partial T}{\partial Y}^\top dr^\top dr \frac{\partial T}{\partial Y} \right) x$.

To compute the final Hessian-vector multiplication $Hx = y$, using (2), in each step three elements of the result y can be fully computed from the input data. Defining

$$H_{l,k}(i, j) := \left(\frac{\partial T_i}{\partial Y_{x_l}} \right) (dr^\top dr)_{i,i+j} \left(\frac{\partial T_j}{\partial Y_{x_k}} \right),$$

with $l, k \in \{1, 2, 3\}$ and x_1, x_2, x_3 are the spacial directions, gives

$$y_{(l-1)MNO+i} = \bar{h} \sum_{j \in \mathcal{N}} \sum_{k=1,2,3} H_{l,k}(i, j) x_{(k-1)MNO+i+j}.$$

Note that for each i , the three values for $l = 1, 2, 3$ can be computed in one thread and MNO threads can do these result computations in parallel without conflicts.

4. EVALUATION

The proposed method was evaluated with two different scopes. First, the Hessian-vector multiplication itself was evaluated on random images of different size. Here, scalability was validated by performing the same calculations single-threaded as well as multi-threaded on a 12-core dual CPU Intel Xeon E5645 workstation. Additionally, time and memory consumption were compared against a standard sparse-matrix-based approach.

Second, the whole registration algorithm was considered. Here, a challenging registration of two pelvis datasets from radiotherapy was computed. An intra-session cone-beam CT was registered to a planning CT image, see Figure 2. As generally motion in pelvis data is strongly locally varying and many different structures have to be taken into account, high deformation resolution is required. To be able to compare the new scheme to the matrix-based approach, the images were down sampled to $128 \times 128 \times 81$ voxels image resolution and a multi-level scheme with $64 \times 64 \times 40$ and $32 \times 32 \times 20$ image resolution [18]. Deformation resolutions were discretized on a nodal grid, see [16] for details, with resolutions of $65 \times 65 \times 41$ and $33 \times 33 \times 21$.

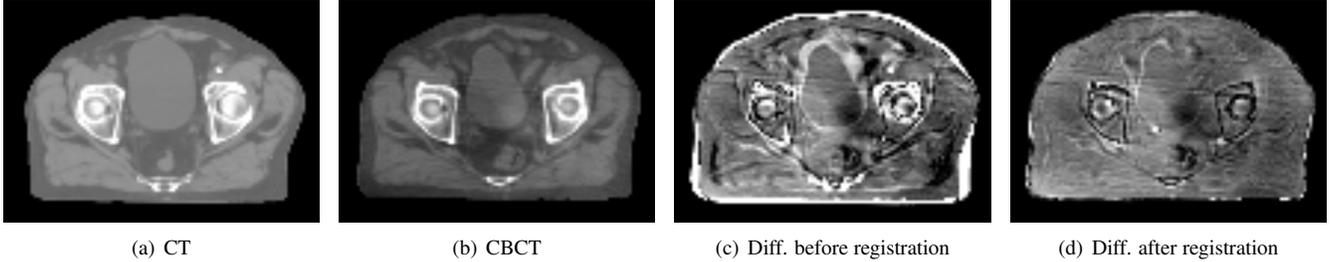


Fig. 2. Registration of three-dimensional CT and cone-beam CT images, a single axial slice is shown. Figures (a) and (b) show the original images; (c) and (d) show the absolute difference before and after registration, respectively.

resolution	t_{mb}	t_{mf}	m_{mb}	m_{mf}
9^3	0.05	0.03	2.77	0.02
17^3	0.32	0.02	23.36	0.10
25^3	1.13	0.03	80.09	0.32
33^3	2.75	0.05	191.34	0.73
41^3	5.00	0.08	375.43	1.40
49^3	8.92	0.09	650.72	2.39
57^3	14.09	0.12	1035.56	3.78
65^3	20.90	0.14	1548.27	5.62
73^3	28.66	0.21	2207.22	7.98
81^3	39.82	0.28	3030.74	10.92

Table 2. Performance measurements of Hessian-vector multiplication for different deformation resolutions. Times t are given in seconds for the matrix-free (mf) and matrix-based (mb) approach, while memory consumption m is given in megabytes. For the matrix-based approach these values include the calculation of the matrix entries as well as the multiplication itself. For the matrix-free scheme, as described, these steps are combined.

5. RESULTS

Evaluation of the new Hessian-vector multiplication gave an average runtime of 9.83s running single-threaded versus 0.81s running on 12 cores (averaged over 10 executions) with a deformation resolution of 129^3 . This implies a speedup factor of 12.14 and thus perfect linear scalability. The results of the evaluation of the Hessian-vector multiplication itself are shown in Figure 3 and Table 2. Only resolutions up to 81^3 could be compared, since the matrix-based approach exceeded 32GB of RAM available on the workstation. While the runtime of both algorithms still exhibits the same complexity class, the matrix-free approach has a vastly shorter runtime. Also, as can be seen in Table 2, the new approach exhibits a significantly lower memory consumption.

Results of the CT to cone-beam CT registration are shown in Figure 2. While the matrix-based approach took 2771.85s to complete, the matrix-free scheme finished in 136.37s, resulting in a speedup of the whole registration algorithm of 20.3.

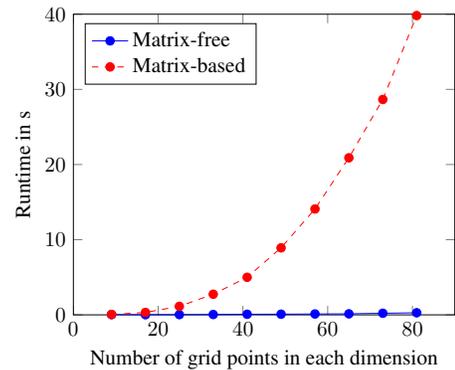


Fig. 3. Runtime of Hessian-vector multiplication depending on the resolution of the computed deformation

6. DISCUSSION

The results have shown that by simplification of calculations through algebraic elimination of intermediate steps using knowledge of the problem structure, computational costs dramatically decreased both in terms of runtime and memory consumption. The practice of on-the-fly calculations of needed intermediate values provides full parallelizability and results in full linear scalability for the presented Hessian-vector multiplication. While the matrix-based approach only worked for comparably small deformation resolutions before exceeding available memory, the matrix-free approach can handle resolutions well above, ultimately only limited by runtime, not memory consumption.

The presented approach extends existing methods to schemes with higher complexity and shows that even with a large number of intermediate results, when using methods that involve matrices of known structure, exploiting this structure yields great benefits. The new scheme enables the calculation of new deformation resolutions that were not possible before due to time and especially memory constraints. By this, the new method enables the computation of fast and accurate results in radiotherapy problems.

7. REFERENCES

- [1] Thörnqvist S., Petersen J.B.B., Høyer M., Bentzen L.N., and Muren L.P., “Propagation of target and organ at risk contours in radiotherapy of prostate cancer using deformable image registration,” *Acta Oncologica*, no. 49, pp. 1023–1032, 2010.
- [2] Zhong H., Weiss E., and Siebers J.V., “Assessment of dose reconstruction errors in image-guided radiation therapy,” *Phys. Med. Biol.*, vol. 719, no. 3, 2008.
- [3] E. Haber and J. Modersitzki, “Intensity gradient based registration and fusion of multi-modal images,” *Methods of Information in Medicine*, vol. 46, pp. 292–9, 2007.
- [4] J. Nocedal and S.J. Wright, *Numerical optimization*, Springer, 1999.
- [5] R. Shams et al., “A survey of medical image registration on multicore and the GPU,” *Signal Processing Magazine, IEEE*, vol. 27, no. 2, pp. 50–60, 2010.
- [6] R. Berg, L. König, J. Rühaak, R. Lausen, and B. Fischer, “Highly efficient image registration for embedded systems using a distributed multicore DSP architecture,” *Journal of Real-Time Image Processing*, 2014.
- [7] H. Wu and Y. Kim, “Fast wavelet-based multiresolution image registration on a multiprocessing digital signal processor,” *International Journal of Imaging Systems and Technology*, vol. 9, no. 1, pp. 29–37, 1998.
- [8] A. Köhn, J. Drexl, F. Ritter, M. König, and H.-O. Peitgen, “GPU accelerated image registration in two and three dimensions,” in *Bildverarbeitung für die Medizin*, pp. 261–265. Springer, 2006.
- [9] A. Khamene, R. Chisu, W. Wein, N. Navab, and F. Sauer, “A novel projection based approach for medical image registration,” in *Biomedical Image Registration. 2006*, vol. 4057 of *Lecture Notes in Computer Science*, pp. 247–256, Springer Berlin Heidelberg.
- [10] P. Muyan-Ozcelik, J.D. Owens, Junyi Xia, and S.S. Samant, “Fast deformable registration on the GPU: A CUDA implementation of demons,” in *Computational Sciences and Its Applications, 2008. ICCSA '08. International Conference on*, June 2008, pp. 223–233.
- [11] X. Gu, H. Pan, Y. Liang, R. Castillo, D. Yang, D. Choi, E. Castillo, A. Majumdar, T. Guerrero, and S.B. Jiang, “Implementation and evaluation of various demons deformable image registration algorithms on a GPU,” *Physics in Medicine and Biology*, vol. 55, no. 1, pp. 207, 2010.
- [12] R. Membarth, F. Hannig, J. Teich, M. Körner, and W. Eckert, “Frameworks for GPU accelerators: A comprehensive evaluation using 2D/3D image registration,” in *Application Specific Processors (SASP), 2011 IEEE 9th Symposium on*, June 2011, pp. 78–81.
- [13] T. Rohlfing and C.R. Maurer Jr., “Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, no. 1, pp. 16–25, March 2003.
- [14] R. Membarth, F. Hannig, J. Teich, M. Körner, and W. Eckert, “Frameworks for multi-core architectures: A comprehensive evaluation using 2D/3D image registration,” in *Architecture of Computing Systems - ARCS 2011*, vol. 6566 of *Lecture Notes in Computer Science*, pp. 62–73. Springer Berlin Heidelberg, 2011.
- [15] L. Hömke, “A multigrid method for anisotropic PDEs in elastic image registration,” *Numerical Linear Algebra with Applications*, vol. 13, no. 2-3, pp. 215–229, 2006.
- [16] L. König and J. Rühaak, “A fast and accurate parallel algorithm for non-linear image registration using normalized gradient fields,” in *Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on*, 2014, pp. 580–583.
- [17] J. Rühaak, L. König, M. Hallmann, N. Papenberg, S. Heldmann, H. Schumacher, and B. Fischer, “A fully parallel algorithm for multimodal image registration using normalized gradient fields,” in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, pp. 572–575.
- [18] J. Modersitzki, *FAIR: Flexible Algorithms for Image Registration*, vol. 6, Society for Industrial and Applied Mathematics (SIAM), 2009.
- [19] B. Fischer and J. Modersitzki, “Curvature based image registration,” *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 81–85, 2003.
- [20] T. Polzin, J. Rühaak, R. Werner, H. Handels, and J. Modersitzki, “Lung registration using automatically detected landmarks,” *Methods of Information in Medicine*, vol. 53, pp. 250–256, 2014.
- [21] J. Rühaak, S. Heldmann, T. Kipshagen, and B. Fischer, “Highly accurate fast lung CT registration,” in *SPIE Medical Imaging 2013: Image Processing*, Lake Buena Vista, USA, 2013.
- [22] L. König, T. Kipshagen, and J. Rühaak, “A non-linear image registration scheme for real-time liver ultrasound tracking using normalized gradient fields,” in *Proc. MICCAI CLUST14*, Boston, USA, 2014, pp. 29–36.