A FULLY PARALLEL ALGORITHM FOR MULTIMODAL IMAGE REGISTRATION USING NORMALIZED GRADIENT FIELDS

J. Rühaak[⋆]*, L. König*[⋆]*, M. Hallmann*[⋆]*, N. Papenberg*[⋆]*, S. Heldmann*[⋆]*, H. Schumacher*[†]*, and B. Fischer*[⋆]

[⋆] Fraunhofer MEVIS Project Group Image Registration, Lübeck, Germany
[†] MiE Medical Imaging Electronics GmbH, Seth, Germany

ABSTRACT

We present a super fast variational algorithm for the challenging problem of multimodal image registration. It is capable of registering full-body CT and PET images in about a second on a standard CPU with virtually no memory requirements.

The algorithm is founded on a Gauss-Newton optimization scheme with specifically tailored, mathematically optimized computations for objective function and derivatives. It is fully parallelized and perfectly scalable, thus directly suitable for usage in many-core environments.

The accuracy of our method was tested on 21 PET-CT scan pairs from clinical routine. The method was able to correct random distortions in the range from -10 cm to 10 cm translation and from -15° to 15° degree rotation to subvoxel accuracy. In addition, it exhibits excellent robustness to noise.

***Index Terms*—** Image registration, Computed tomography, Positron emission tomography, Computational efficiency, Parallel algorithms

1. INTRODUCTION

Multimodal image registration has a long history in science and engineering [1, 2]. The integration of complementary information, e.g. fusion of anatomical CT or MR scans with functional data from nuclear medicine such as PET and SPECT, can provide more detailed insight into biological processes and their exact location in the body. This integration is valuable in various clinical situations such as radiation therapy planning or tissue biopsy [3], and a lot of research has been performed on the topic [4]. Hybrid systems may circumvent the need for registration, but there are numerous clinical situations in which separate scans are beneficial, see e.g. the discussion in [3]. Hence, the registration of these modalities is of unchanged importance.

In many clinical workflows, image registration forms the basis for joint reading of complementary image modalities. A medical expert willing to focus on the patient images and their interpretation needs a registration that is robust, accurate, and fast. Unfortunately, many advanced registration methods miss the speed requirement by far and are thus currently no option on contemporary hardware in the clinic.

In this work, we present a multimodal image registration algorithm that meets the clinical requirements. It is designed for both rigid and affine-linear registration and combines accuracy, high speed and extremely low memory consumption to a very attractive package. The main principles are directly extensible to non-linear transformation models. From a computational point of view, mainly two aspects of our algorithm are of interest. First, the algorithm is fully parallelized with no relevant serial parts, resulting in excellent scalability with increasing number of computational cores. Secondly, it requires only a minimal amount of memory as no intermediate results such as image derivatives or grids are stored. Given the ever-increasing size of images, both properties are highly desirable now and in the future.

Our main contribution is a detailed analysis of the structure of the normalized gradient fields objective function and its derivatives. Our results allow for a reformulation of the objective function which drastically simplifies function value and especially derivative computations. The approach can be generalized to other distance measures and is directly suitable for efficient use in many-core systems.

2. RELATED WORK

The key ingredient to successful multimodal image registration is the distance measure. As images from different modalities offer complementary information, a straightforward comparison of intensities is bound to fail [5]. The arguably most prevalent distance measure is mutual information [6] which aims at maximizing the stochastical dependence between the intensities in the two images. The general formulation allows for usage in a variety of situations, yet the resulting objective function is highly nonconvex and hence difficult to optimize. Moreover, the estimation of the intensity distributions and their derivatives is nontrivial and computationally expensive [5]. A fast and robust alternative, the normalized gradient fields, was proposed in [7].

As high runtimes are common to almost all registration methods [8], the parallelization of image registration algorithms has received significant research attention [9]. Most

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

approaches, however, are restricted to a parallelization of selected components such as the interpolator or the distance measure computations, thereby limiting performance scalability with increasing number of computational cores.

The successful use of GPUs has been reported e.g. in [10]. As powerful GPUs with sufficiently large memory resources are rarely available in a clinical setting, the classical CPU is currently the most important target hardware for medical applications. Our method, however, is by design directly usable on massively parallel architectures such as GPUs.

3. METHOD

In order to provide an optimized algorithmic scheme, we first take a closer look at the general registration framework.

3.1. Registration framework

The goal of image registration is to find a reasonable transformation φ that maps a three dimensional template image T to a reference image R, so that both are similar in terms of a distance measure D.

The images are interpreted in a continuous model as functions $T : \mathbb{R}^3 \to \mathbb{R}$ and $R : \mathbb{R}^3 \to \mathbb{R}$ with compact support in domains $\Omega_R \subseteq \mathbb{R}^3$ and $\Omega_T \subseteq \mathbb{R}^3$, respectively. The transformation $\varphi : \Omega_R \to \mathbb{R}^3$ maps the reference image domain to the template domain and thus allows comparison of reference image R and deformed template $T(\varphi) := T \circ \varphi$. The distance measure $D(T(\varphi), R)$ depends on fixed R and the transformed template $T(\varphi)$, thus $D(T(\varphi), R) =: D(\varphi) \overset{\varphi}{\to} \min$ in order to find a plausible alignment of the images. Here, φ allows for affine-linear transformations that map a single point $\mathbf{x} = (x, y, z)^\top \in \mathbb{R}^3$ with $\varphi : \mathbf{x} \mapsto A\mathbf{x}+b$, $A \in \mathbb{R}^{3\times3}, b \in \mathbb{R}^3$. By using a specifically chosen transformation matrix A, this transformation model can also easily be restricted to e.g. rigid transformations, i.e. rotations and translations.

Minimization of the objective function is achieved using numerical optimization. The used Gauss-Newton approach requires the computation of the gradient ∇D and an approximation of the Hessian $\nabla^2 D$. It features fast convergence and can be implemented in a computationally efficient way [5].

3.2. Normalized gradient fields distance measure

In multimodal imaging, the most prominent features that persist over different modalities are image edges. As proposed in [7, 5] we use a distance measure based on normalized image gradient fields (NGF) which is both well suited for optimization and fast computation. The main idea is to measure the angle between two image gradients and to align these in either a parallel or antiparallel fashion. Our approach is embedded in a numerical optimization scheme which relies on minimizing a discretized version of the objective functional

$$D_{\mathrm{NGF}}(\varphi) = \frac{1}{2}\int_{\Omega_R} 1 - \left(\frac{\langle \nabla T(\varphi), \nabla R\rangle_{\varrho,\tau}}{\|\nabla T(\varphi)\|_\tau \|\nabla R\|_\varrho}\right)^2 \mathrm{d}x,$$

i.e. measuring the sine between the gradients of the transformed template image $T(\varphi)$ and the reference image R, where the inner product $\langle\cdot,\cdot\rangle_{\varrho,\tau} := \langle\cdot,\cdot\rangle + \varrho\tau$ and $\|\cdot\|_* = \langle\cdot,\cdot\rangle_{*,*}$. The values $\varrho, \tau > 0$ represent modality dependent parameters which allow filtering of noise.

Let $[0, M], [0, N], [0, P]$ denote the ranges of indices in x-, y- and z-direction for a discretization of Ω_R. We then define $T_i(\varphi)$ as the deformed template, interpolated on the reference image domain at point $i = x + My + MNz$ using trilinear interpolation. Applying the midpoint quadrature rule and finite differences we obtain a discretized distance measure. Defining $h_{\pm MN}, h_{\pm M}, h_{\pm 1}$ as the stepwidth in x-, y-, z-direction, respectively, as well as $\bar{h} := h_{\pm MN}h_{\pm M}h_{\pm 1}$, the discretization of the distance measure can be written as

$$\begin{aligned} D_{\mathrm{NGF}}(\varphi) &\approx \frac{\bar{h}}{2}\sum_{i=1}^{MNP}\left(1 - \left(\frac{\langle g(T_i(\varphi)), g(R_i)\rangle_{\varrho,\tau}}{\|g(T_i(\varphi))\|_\tau \|g(R_i)\|_\varrho}\right)^2\right) \\ &=: \frac{\bar{h}}{2}\sum_{i=1}^{MNP}\left(1 - r_i^2\right) =: \psi(r(T(\varphi))), \end{aligned} \tag{1}$$

where g is the approximation to the image gradient of two images I, J at point i. It is defined with forward/backward finite differences $g_+(I_i), g_-(I_i)$ respectively and $\langle g(I_i), g_i(J_i)\rangle := \frac{1}{2}\left(\langle g_+(I_i), g_+(J_i)\rangle + \langle g_-(I_i), g_-(J_i)\rangle\right)$, with the corresponding induced norm $\|g(I_i)\|$.

3.3. Problem specific derivative computation

The function value evaluation of D can be straightforwardly implemented by parallelizing over all voxels. For using the Gauss-Newton optimization scheme, the computation of the gradient and Hessian approximation is needed. To be able to fully parallelize those as well, we need to analyze their mathematical composition to derive a specialized version. The cascaded formulation in (1) can be differentiated by the product rule as $\mathrm{d}\psi = \frac{\partial\psi}{\partial r}\mathrm{d}r = \frac{\partial\psi}{\partial r}\frac{\partial r}{\partial T}\frac{\partial T}{\partial\varphi}$. With the set of indices $\mathcal{M} = \{-MN, -M, -1, 0, 1, M, MN\}$ we can first define

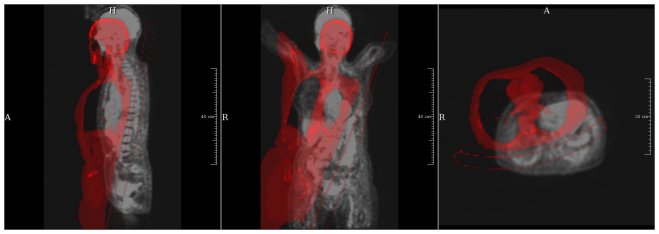
$$(\hat{r}_i)_l = \frac{1}{2h_l}\left(\frac{R_{i-l} - R_i}{\|g_i(R)\|_\varrho \|g_i(T(\varphi))\|_\tau} - \frac{\langle g_i(T), g_i(R)\rangle_{\varrho,\tau}\left(T_{i-l}(\varphi) - T_i(\varphi)\right)}{\|g_i(R)\|_\varrho \|g_i(T(\varphi))\|_\tau^3}\right)$$

and then the lth component of the derivative $\frac{\partial r}{\partial T}$ at point i as

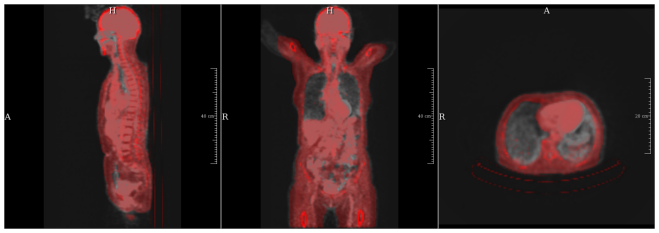
$$\left(\frac{\partial r_i}{\partial T}\right)_l = \begin{cases} (\hat{r}_i)_l, & \text{if } l \in \mathcal{M}\setminus 0 \\ -\sum_{l\in\mathcal{M}\setminus 0}\left(\frac{\partial r_i}{\partial T}\right)_l, & \text{if } l = 0 \\ 0, & \text{otherwise} \end{cases},$$

so that the $(k + 3j)$th component of $\mathrm{d}r$ can be written as

$$(\mathrm{d}r_i)_{k+3j} = \sum_{l\in\mathcal{M}}\left(\frac{\partial r_i}{\partial T}\right)_{i-l}\frac{\partial T_{i-l}(\varphi)}{\partial \mathbf{x}_{i-l+j\cdot MNP}}\mathbf{x}_{i-l+k\cdot MNP},$$
$$k \in \{0, \ldots, 3\},\ j \in \{0, \ldots, 2\},$$



(a)



(b)

Fig. 1: PET reference and CT template image before (a) and after (b) registration. The CT image is displayed in red.

with $\mathbf{x}_{i-l+3\cdot MNP} = 1$. Here the term $\frac{\partial T_{i-l}(\varphi)}{\partial \mathbf{x}_{i-l+j\cdot MNP}}$ denotes the derivative of the interpolant of the deformed template image at point $(i-l)$ with respect to coordinates $\mathbf{x}_{i-l+j\cdot MNP}$. Using the previous notation and $\frac{\partial \psi}{\partial r} = -\bar{h} r^\top$, the gradient of the distance measure then finally results in

$$\nabla D(\varphi) = \bar{h} \sum_{i=1}^{MNP} -r_i \mathrm{d}r_i. \tag{2}$$

To restrict φ to e.g. rigid transformations, another partial derivative of the transformation would enter the term $\mathrm{d}r$. This can easily be done and thus will be neglected here. As mentioned above, the Hessian can then be approximated as

$$\nabla^2 D(\varphi) = \bar{h} \sum_{i=1}^{MNP} \mathrm{d}r_i \cdot \mathrm{d}r_i^\top, \tag{3}$$

so that the derivatives (2) and (3) can be formulated in a matrix-free way in terms of their basic components T and R. This eliminates inter-point dependencies in intermediate results and enables full pointwise parallelization.

Our approach is embedded in a multi-level scheme where the objective function is first discretized and optimized on a smooth, coarse level and then successively on finer discretizations [5]. Note that computations are only performed up to a finest level of 128^3 voxels, because finer resolutions did not improve registration accuracy of our method in repeated experiments. To ensure a plausible alignment of sub-volumes (e.g. lung CT and full body PET), prior to registration we perform a principle component axis alignment of the images [11] and a z-direction grid-search to identify an initial alignment with the lowest value of D.

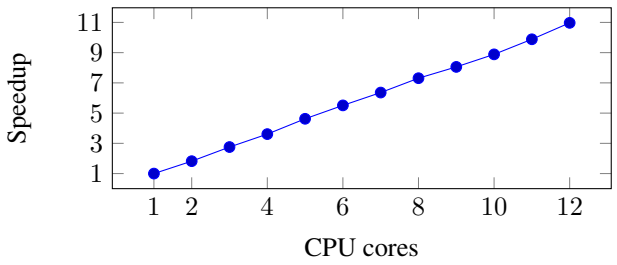


Fig. 2: Performance using an increasing number of CPU cores in multiples of one-core computation speed

3.4. Implementation Highlights

Our method is implemented in C++ and parallelized using standard OpenMP reduction techniques. The code was optimized using conventional profiling tools and serial optimization strategies. At the cost of seemingly higher computational work our formulation avoids slow memory write accesses and enables direct benefits from increased computational power, without memory bandwidth being a bottleneck. This ensures linear scalability with increased CPU core count and increases computational performance also on single-core systems compared to a sparse-matrix based computation that uses more memory accesses but less computations.

4. EVALUATION

The evaluation of registration methods is difficult. In general, the correspondence function φ is unknown, and it is hence a priori unclear how to compare two or more registration approaches. Common surrogates include the use of landmarks, segmentation masks, and deformation vector field analysis for non-linear algorithms [8]. All these criteria, however, can only express necessary conditions to the desired solution.

For this reason, we decided to perform an initial evaluation study on data with known ground truth deformation. We chose a set of 21 PET-CT scans from various hospitals and scanners stemming directly from clinical routine (image courtesy MiE GmbH, Seth), ranging from image sizes $128\times128\times144$ (PET) to $512\times512\times844$ (CT) voxels. As these data origins from hybrid PET-CT systems, we assume that the PET and CT scans are inherently registered. For all registrations, the PET is chosen as fixed and the CT as moving image.

Subsequently, we applied randomly selected artificial deformations in the range from -10 cm to 10 cm translation and from $-15°$ to $15°$ degree rotation to the CT scan, deformed it accordingly and selected the original PET scan together with the deformed CT as input to our registration algorithm. An example is shown in figure 1.

In a second study, we added Gaussian noise of different standard deviation to both scans and repeated our experiments. All computations were performed on a 3.4 GHz Intel i7-2600 quad-core on Ubuntu Linux.

Case	Avg. distance	Avg. runtime	Voxelsize (mm^3)
1	$1.53mm$	$1.55s$	$5.31\times5.31\times2.50$
2	$2.02mm$	$1.87s$	$5.31\times5.31\times2.30$
3	$2.12mm$	$1.59s$	$5.31\times5.31\times2.30$
4	$3.25mm$	$1.50s$	$4.00\times4.00\times4.00$
5	$2.14mm$	$1.43s$	$5.31\times5.31\times2.50$
6	$1.62mm$	$1.58s$	$5.31\times5.31\times2.30$
7	$2.58mm$	$1.54s$	$5.31\times5.31\times2.30$
8	$1.85mm$	$1.74s$	$5.31\times5.31\times2.30$
9	$1.56mm$	$1.64s$	$5.31\times5.31\times2.50$
10	$1.56mm$	$1.60s$	$5.31\times5.31\times2.30$
11	$2.13mm$	$1.80s$	$5.31\times5.31\times2.30$
12	$1.50mm$	$1.45s$	$5.31\times5.31\times2.50$
13	$1.86mm$	$1.38s$	$5.31\times5.31\times2.50$
14	$1.45mm$	$1.75s$	$5.31\times5.31\times2.50$
15	$1.62mm$	$1.80s$	$5.31\times5.31\times2.30$
16	$1.98mm$	$1.56s$	$5.31\times5.31\times2.30$
17	$1.52mm$	$1.62s$	$5.31\times5.31\times2.50$
18	$2.78mm$	$1.41s$	$5.31\times5.31\times2.50$
19	$1.89mm$	$1.75s$	$5.31\times5.31\times2.30$
20	$4.50mm$	$2.78s$	$5.31\times5.31\times7.00$
21	$3.18mm$	$2.31s$	$5.31\times5.31\times7.00$

Table 1: Evaluation results for ten random deformations each on 21 clinical PET-CT datasets. The voxelsize corresponds to the reference image (PET).

σ	0	200	400	600	800	1000
Error (mm)	2.12	2.70	2.89	3.08	3.20	3.25

Table 2: Average registration error for increasing noise levels.

In order to assess the computational efficiency of our method in more detail, we additionally performed registrations on a dual CPU twelve-core workstation. The scalability of our method was analyzed by varying the number of active cores from 1 to 12.

5. RESULTS

The results of the artificial deformations experiment are shown in table 1. The average registration error was reduced from $133.72mm$ to $2.12mm$ which is below the PET image resolution. An average runtime of $1.69s$ was measured.

The registration accuracy for various noise levels is shown in table 2. Adding Gaussian noise with $\sigma = 1000$ led to an error of $3.25mm$ instead of $2.12mm$. The results of the scalability measurements are shown in figure 2.

5.1. Discussion

We have presented a highly efficient multimodal image registration algorithm based on normalized gradient fields. The method is fully parallelized and exhibits a minimal memory footprint. It is founded on well-known mathematical and numerical principles.

Our method was evaluated with randomly chosen artificial deformations on 21 PET-CT datasets from clinical routine. It achieved very accurate results with computation times of just about 1.7 seconds on average on a standard PC. Moreover, it exhibits excellent robustness to noise – even at a very high noise level, the algorithm still achieved subvoxel accuracy. Hence, we believe that our method has the realistic potential to be routinely used in the clinic.

We will extend our scheme to variational approaches for nonlinear image registration in the future. Using the same reformulation techniques as presented here, comparably fast algorithms also for non-linear approaches come within reach.

6. REFERENCES

[1] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

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

[3] A. Alavi et al., "Is PET-CT the only option?," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 34, pp. 819–821, 2007.

[4] I.N. Bankman, *Handbook of Medical Image Processing and Analysis*, Academic Press, 2008.

[5] J. Modersitzki, *FAIR: Flexible Algorithms for Image Registration*, vol. 6, Society for Industrial and Applied Mathematics (SIAM), 2009.

[6] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual-information-based registration of medical images: a survey," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 8, pp. 986–1004, 2003.

[7] 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.

[8] K. Murphy et al., "Evaluation of registration methods on thoracic CT: The Empire10 challenge," *Medical Imaging, IEEE Transactions on*, vol. 30, pp. 1901–20, 2011.

[9] 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.

[10] A. Köhn et al., "GPU accelerated image registration in two and three dimensions," *Bildverarbeitung für die Medizin 2006*, pp. 261–265, 2006.

[11] J. Modersitzki, *Numerical Methods for Image Registration*, Oxford University Press, 2004.