Pulmonary Image Registration with elastix using a Standard Intensity-Based Algorithm

Marius Staring[1*], Stefan Klein[2], Johan H.C. Reiber[1], Wiro J. Niessen[2], and Berend C. Stoel[1]

[1] Division of Image Processing (LKEB), Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands

[2] Biomedical Imaging Group Rotterdam, Depts. of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands

Abstract. Accurate registration of thoracic CT is useful in clinical terms and also challenging due to the elastic nature of lung tissue deformations. The goal of the EMPIRE10 challenge (Evaluation of Methods for Pulmonary Image Registration 2010), a workshop of the MICCAI 2010 conference, is to provide a platform for in-depth evaluation and fair comparison of available registration algorithms for this application.

To this end we registered to the challenge with team RubberBand. The goal of our submission is to determine what a standard, but fully automatic, intensity-based image registration algorithm can achieve compared to the competition.

The algorithm, implemented in `elastix`, optimises the normalised correlation criterion, using a fast, parameter-free and robust stochastic optimisation procedure. A combination of an affine and two nonrigid B-spline transformations models the spatial relationship. The approach is embedded in a multi-resolution framework for both the image data and the transformation. No explicit regularisation is used.

Of the 34 submitted algorithms, our contribution achieved the 7-th place with an average rank of 13.13 (best 8.03, worst 31.46). The incorporation of a regularisation term may improve the ranking of the algorithm, since our final score was most negatively influenced by the score for folding.

Key words: pulmonary image registration, evaluation, `elastix`

1 Introduction

The registration of pulmonary CT data has drawn considerable interest from many research groups. In addition, many registration algorithms already exist, of which it is currently unclear which perform best on pulmonary data, or which parts of algorithms are beneficial for robustness, precision and accuracy.

In order to compare the performance of the several algorithms, the EMPIRE10 team organised a challenge. The EMPIRE10 website (`http://empire10.isi.uu.nl`) states: There are a number of benefits to comparing algorithms in this way:

* Corresponding author: `m.staring@lumc.nl`

– All algorithms will be applied to exactly the same set of data.
– Any algorithm parameters or settings will be chosen by those familiar with the algorithm and expert in its configuration.
– The resulting registrations will be independently evaluated using the same criteria for all participants.

The authors registered to the challenge under the team name "RubberBand", referring to the registration package `elastix`, developed previously by the authors [1]. Executables and source code of `elastix` are publicly available from the website `http://elastix.isi.uu.nl`, under the BSD license, which allows free academic and commercial use and permits modification of the source code. A manual for `elastix` and an example of usage can also be downloaded. In addition, we created a "parameter file database", which is a collection of parameter files that proved successful, together with a short description of the clinical application for which they were used. The parameter file database can be found through the website[3] and `elastix`-users are encouraged to upload their own settings. A default parameter file can also be found here.

The goal of our contribution is to determine what a standard and generic, but fully automatic, intensity-based image registration algorithm can achieve compared to the competition. How will this relatively simple, general purpose algorithm compare to more advanced registration algorithms that are tailored to the specific application?

2 Methods

In this paper we adopt the formulation of image registration as an optimisation problem:

$$\widehat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \mathcal{C}(\boldsymbol{T}_{\boldsymbol{\mu}}; I_F, I_M), \tag{1}$$

where I_F and I_M denote the fixed and moving image, respectively, $\boldsymbol{T}_{\boldsymbol{\mu}}$ is the spatial transformation relating the two and parameterised by the vector of parameters $\boldsymbol{\mu}$, and $\mathcal{C}$ is the cost function or similarity measure that defines the quality of alignment.

Image registration is performed in three stages:

1. Affine registration using the original data, without the use of lung masks. This is done to get a coarse global alignment of the entire anatomy. Lung masks are not used, to exploit all anatomy.
2. Nonrigid registration using the processed data (see "Masking" below for an explanation), without the use of lung masks. Our experiments revealed that the use of lung masks at an early stage had a negative impact on lung boundary alignment in case of large offsets (i.e. large differences in inspiration level).

[3] `http://elastix.isi.uu.nl/wiki.php`

3. Nonrigid registration using the processed data, and with the use of the lung mask of the fixed image. From our experiments we learned that the match of smaller structures within the lung is improved by using a lung mask.

Several choices for the different registration components are made:

Cost function $\mathcal{C}$: Normalised Correlation Coefficient (NCC), which is suitable for mono-modal image registration, but can compensate for global intensity differences due to differences in inspiration level. It is defined as:

$$\mathrm{NCC}(\boldsymbol{\mu}; I_F, I_M) = \frac{\sum\limits_{\boldsymbol{x}_i} \left(I_F(\boldsymbol{x}_i) - \overline{I_F}\right)\left(I_M(\boldsymbol{T_\mu}(\boldsymbol{x}_i)) - \overline{I_M}\right)}{\sqrt{\sum\limits_{\boldsymbol{x}_i} \left(I_F(\boldsymbol{x}_i) - \overline{I_F}\right)^2 \sum\limits_{\boldsymbol{x}_i} \left(I_M(\boldsymbol{T_\mu}(\boldsymbol{x}_i)) - \overline{I_M}\right)^2}}, \qquad (2)$$

where $\boldsymbol{x}_i$ are samples drawn from the fixed image, with $\overline{I_F} = \frac{1}{|\Omega_F|}\sum_{\boldsymbol{x}_i} I_F(\boldsymbol{x}_i)$ the average grey-value, and similarly for $\overline{I_M}$.
It is possible to include regularisation in the cost function: $\mathcal{C} = \mathrm{NCC} + \alpha\mathcal{R}$, with a suitable choice for $\mathcal{R}$, for example the bending energy, penalising the second order derivatives [2]. For the specific application at hand, regularisation would have been beneficial for some data sets (appearance of smearing effects). However, it requires manual setting of an additional data-dependent parameter α, which is not a trivial choice, and additionally it increases the computation time. Therefore, for the sake of simplicity regularisation was omitted, at the cost of a deduction of points in the evaluation (singularities in the deformation field).

Transformation: An affine registration is performed prior to nonrigid registration to accommodate for global offset and differences in inspiration level. Subsequent nonrigid transformations are modelled by B-splines [2], embedded in a multi-grid setting. The latter require the setting of the grid spacing and the definition of a multi-grid strategy. For stage 2 the grid spacing was set to 80, 80, 40, 20, and 10 mm in each direction for each resolution, respectively. For stage 3 we used 80, 40, 20, 10, and 5 mm. See Table 1 for an overview.

Optimisation: To optimise (1) we opt for an iterative procedure, called adaptive stochastic gradient descent (ASGD) [3]:

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - \gamma(t_k)\widetilde{\boldsymbol{g}}_k, \qquad (3)$$

Table 1. Parameter settings for stages 1-3, resolution levels R1-R5.

Stage	Iterations		R1	R2	R3	R4	R5
1. Affine	1000						
2. Nonrigid without mask	1000	Grid spacing (mm)	80	80	40	20	10
		Downsample factor	16	8	4	2	1
3. Nonrigid with mask	2000	Grid spacing (mm)	80	40	20	10	5
		Downsample factor	4	3	2	1	1

with k the iteration number, and $\widetilde{\boldsymbol{g}}_k$ an approximation of the cost function derivative $\partial\mathcal{C}/\partial\boldsymbol{\mu}_k$. The derivative is approximated by random sampling of the fixed image with a relatively small number of samples [4]. The scalar $\gamma(t_k)$ determines the step size, where $\gamma(\cdot)$ is a decaying function and t_k is defined by $t_{k+1} = \left[t_k + f(-\widetilde{\boldsymbol{g}}_k^T \widetilde{\boldsymbol{g}}_{k-1})\right]^+$, with $[\cdot]^+ = \max(\cdot, 0)$, and $f(\cdot)$ a sigmoid function. The use of the inner product $\widetilde{\boldsymbol{g}}_k^T \widetilde{\boldsymbol{g}}_{k-1}$ for determining the step size makes the optimisation procedure adaptive (dependent on the estimated progress), and the registration more robust. Details can be found in [3]. The stop condition is a user-defined maximum number of iterations K. We used $K = 1000, 1000, 2000$ iterations for each of the three stages, respectively (see Table 1). Other settings of ASGD were left to their defaults.

Sampling strategy: A relatively small number of samples (2000) are drawn randomly each iteration from the fixed image (off the voxel grid), to compute $\widetilde{\boldsymbol{g}}_k$.

Interpolation: During registration a linear interpolator is used to compute the spatial derivative of the moving image $\partial I_M/\partial\boldsymbol{x}$, required for computing $\widetilde{\boldsymbol{g}}_k$.

Hierarchical strategy: For the image data Gaussian pyramids are used with sub-sampling, to increase robustness. For the B-spline transform a multi-grid approach is used, starting with a coarse control point grid in the first resolution, only capable of modelling coarse deformations. In subsequent resolutions the B-spline grid is gradually refined, thereby introducing the capability to match smaller structures. For all stages 5 resolutions were used, with isotropic down-sampling of the data, since the data was mostly isotropic. For stage 1 and 2 the images were down-sampled with a factor of 16, 8, 4, 2, and 1, for each resolution, respectively. For stage 3 the factors were 4, 3, 2, 1, and 1. See Table 1 for an overview.

Masking: In the last step of the registration procedure we have used lung masks, which were created automatically by the EMPIRE10 organisers. The fixed and moving image data were processed in the following way: all voxels that are more than 2 voxels outside the lung segmentation (provided by EMPIRE10) are given the intensity value 0. This was done to mitigate the effect the ribs have on the transformation within the lungs, observed in earlier experiments. The ribs move discontinuously from the lung field during inspiration and can therefore not be used as guidance. The image gradient $\partial I_M/\partial\boldsymbol{x}$ is relatively high at the ribs, influencing the deformation field, and propagating its effect into the lungs causing misalignment of fissures and vessels. This preprocessing step eliminates that problem.

The above image registration algorithm is fully automatic. Exact `elastix` settings that were used in the experiments have been made available via the parameter file database, see `http://elastix.isi.uu.nl/wiki.php`, see `par0011`. The registration settings for each experiment can be inspected in detail, and the parameter files can be downloaded for reproducing our results or for use in other applications.

3 Experiments and Results

Data sets and the scoring methodology are described in [5].

3.1 Runtime

All registration were performed on an Intel Xeon W3520 @ 2.66 GHz, 4GB RAM, Windows 7 64 bit. The mean run time of `elastix` for each stage is given in Table 2. On average the registration took 16 minutes, 20 seconds, of which 5 minutes, 18 seconds were spent to automatically compute the optimisation parameters by the ASGD optimiser. The computation time in stage 3 is longer than that of stage 2 due to a doubling in the number of iterations and a more involved computation of $\boldsymbol{T}_{\boldsymbol{\mu}}(\boldsymbol{x})$, since at stage 3 $\boldsymbol{T}_{\boldsymbol{\mu}}(\boldsymbol{x})$ is a composition of three transforms.

3.2 Results

Visual inspection of the affine registration showed that all scans were successfully matched globally. Automatic scoring was performed on the final result after stage 3, by the EMPIRE10 organisers. The results are given in Table 3. They are divided in four categories: lung boundary match, fissure match, landmark precision, and the presence of singularities in the deformation field. A comparison to other participants can be found at `http://empire10.isi.uu.nl/mainResults.php`. Overall, of the 34 submitted algorithms, our contribution achieved the 7-th place with an average rank of 13.13 (best 8.03, worst 31.46).

The results for the lung boundary match are very good. Never an overall error of more than 0.006% was obtained, with an average of 0.001%. This resulted in a final rank of 11.15/34, at the same level of competing algorithms, except for the two with final rank 1 and 2. Two poor scores for lung boundary overlap are suspected to be due to breathing artifacts in the CT scan (scans 13, 16).

With respect to the fissures our errors are slightly higher than those of other submissions with final rank 1 - 10. Scans 07 (noise), 14 (diffuse areas), 18 and 20 were especially difficult. An average error of 0.50% was obtained, with corresponding rank 13.62 (min 9.52, max 16.52 in 1-10).

Table 2. Average runtime in seconds for each stage of the registration.

	registration			ASGD		
stage	mean	min	max	mean	min	max
1. Affine	56	30	70	5	3	8
2. Nonrigid without mask	216	150	286	84	45	138
3. Nonrigid with mask	708	511	1279	229	70	720
total	980			318		

	Lung Boundaries		Fissures		Landmarks		Singularities	
Scan Pair	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
01	0.00	9.00	0.00	3.00	1.72	7.00	0.00	11.50
02	0.00	11.00	0.00	15.00	0.33	1.00	0.01	29.00
03	0.00	13.00	0.00	12.50	0.34	7.00	0.00	12.00
04	0.00	11.00	0.00	16.50	0.79	2.00	0.00	14.00
05	0.00	13.00	0.00	16.00	0.02	15.00	0.00	29.00
06	0.00	16.00	0.00	7.00	0.32	9.00	0.00	14.00
07	0.00	3.00	1.01	15.00	3.15	17.00	0.02	27.00
08	0.00	4.00	0.36	21.00	0.59	4.00	0.00	12.50
09	0.00	11.00	0.00	21.00	0.54	6.00	0.00	27.00
10	0.00	15.00	0.00	15.00	0.96	4.00	0.00	27.00
11	0.00	3.00	0.14	18.00	0.60	1.00	0.01	27.00
12	0.00	23.00	0.00	13.50	0.00	1.50	0.00	14.50
13	0.00	23.00	0.07	10.00	1.03	17.00	0.21	32.00
14	0.00	5.00	3.60	17.00	3.32	16.00	0.11	28.00
15	0.00	16.00	0.00	7.00	0.63	8.00	0.00	12.50
16	0.00	21.00	0.00	6.50	0.76	4.00	0.02	28.00
17	0.00	6.50	0.05	19.50	0.71	8.00	0.01	29.00
18	0.00	2.00	2.74	16.00	2.28	11.00	0.00	10.50
19	0.00	14.00	0.00	12.00	0.40	1.00	0.00	14.50
20	0.00	3.50	2.13	11.00	1.36	6.00	0.00	10.50
Avg	0.00	11.15	0.50	13.62	0.99	7.27	0.02	20.47
Average Ranking Overall								13.13
Final Placement								7

Table 3. Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms.

Regarding the landmark distance error our submission performs very well: there are 6 algorithms that perform better in terms of average score, and only 3 in terms of average rank. This resulted in a final rank of 7.27.

The weakest point of our submission is the presence of singularities in the deformation field as measured by a negative determinant of the Jacobian $\partial \boldsymbol{T}/\partial \boldsymbol{x}$. Our submission reaches an average rank of 20.47 out of 34, which is about 7 worse than submissions with a final rank around ours (1 - 10). A more detailed analysis revealed that 10/20 scans score poor, and the remaining 10 obtained a perfect score of 0% folding. Of the 10 poor results 1/10 was due to a streaking artifact in the CT scan (scan 13), 5/10 were due to higher noise levels and especially diffuse scan areas (scans 07, 10, 11, 14, 16), and the remaining 4/10 (scans 02, 05, 09, 17) showed no visual appearance of a smearing effect.

4 Discussion and Conclusion

The goal of our contribution was to determine what a standard and generic, but fully automatic, intensity-based image registration algorithm can achieve compared to the competition. The results are good in terms of lung boundary alignment, landmark alignment, and fissure alignment. The worst scores were obtained in the category 'Singularities'. This is due to the omission of a regularising term in the cost function. We expect that the use of a regularisation term will bring the folding score up to par with competing algorithms, which would improve the final placement from 7 to around 5. This will however increase the complexity of the algorithm somewhat, and introduce an extra parameter. Another possibility would be to include hard constraints on the determinant of the Jacobian of the transformation, as in [6].

There seems little a-priori performance bias to certain scan protocols. We observed not much difference between the registration quality of low dose (LD) scan pairs, or LD-ULD scan pairs. Additionally, the two sheep datasets and the artificially warped data were registered with similar quality as the other scans. The algorithm however performed less well on data containing diffuse areas, or areas with little structure, which is not a surprising observation.

In conclusion, a standard, fully automatic, intensity-based image registration algorithm achieved a ranking of 7 out of 34, with room for improvement in the category 'Singularities'. The implementation is publicly available from `http://elastix.isi.uu.nl`.

References

1. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1) (January 2010) 196 – 205
2. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging **18**(8) (1999) 712 – 721
3. Klein, S., Pluim, J.P.W., Staring, M., Viergever, M.A.: Adaptive stochastic gradient descent optimisation for image registration. International Journal of Computer Vision **81**(3) (2009) 227–239
4. Klein, S., Staring, M., Pluim, J.P.W.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. IEEE Transactions on Image Processing **16**(12) (December 2007) 2879 – 2890
5. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Pluim, J.: Evaluation of methods for pulmonary image registration: The EMPIRE10 study. (2010)
6. Sdika, M.: A fast nonrigid image registration with constraints on the jacobian using large scale constrained optimization. IEEE Transactions on Medical Imaging **27**(2) (2008) 271 – 281