DIRBOOST: AN ALGORITHM FOR BOOSTING DEFORMABLE IMAGE REGISTRATION

Sascha E.A. Muenzing[⋆]*, Bram van Ginneken*[†]*, Josien P.W. Pluim*[⋆]

[⋆] Image Sciences Institute, University Medical Center Utrecht, The Netherlands
[†] Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Center, The Netherlands

ABSTRACT

We introduce a novel boosting algorithm to boost – i.e. improve on – existing methods for deformable image registration. The proposed DIRBoost algorithm is inspired by the theory on hypothesis boosting, well-known in the field of machine learning. DIRBoost involves a classifier for landmark-based Registration Error Detection (RED). Based on these RED predictions a Voronoi tessellation is generated to obtain a dense estimate of local image registration quality. All areas predicted as erroneous registration are subjected to boosting, i.e. undergo iterative registrations by employing boosting masks on both the fixed and moving image. We evaluated the DIRBoost algorithm on five CT pulmonary breathhold inspiration and expiration scan pairs, employing the NiftyReg registration algorithm. DIRBoost could boost about 50% of the wrongly registered areas which in turn also improved the average landmark registration error by 24%.

Index Terms— Deformable image registration, Boosting, Ensemble Learning, Pattern recognition

1. INTRODUCTION

Accurate registration of medical images is key to medical image analysis. Recently an evaluation study on pulmonary intra-patient CT registration has evaluated and ranked state-of-the-art registration algorithms on a common data set [1]. Although many of those registration algorithms achieve excellent results on average, usually there are few scan pairs or distinct areas within one image where the registration fails. In this paper we propose a boosting algorithm for deformable image registration (DIRBoost) that allows to detect locally wrong registrations and subsequently improves on those areas in an iterative manner, similar to the well-established boosting of classifiers.

Deformable Image Registration. Registration of a moving image $I_M(\mathbf{x}) : \Omega_M \subset \mathbb{R}^D \mapsto \mathbb{R}$ to a fixed image $I_F(\mathbf{x}) : \Omega_F \subset \mathbb{R}^D \mapsto \mathbb{R}$, both of dimension D, is the problem of finding a displacement $\mathbf{u}(\mathbf{x})$ that makes $I_M(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ spatially aligned to $I_F(\mathbf{x})$. We define the obtained transformation field $\mathbf{T}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$. The quality of alignment is defined by a distance or similarity measure, such as the normalized mutual information (NMI). If the underlying transformation model allows local deformations, i.e. nonlinear fields $\mathbf{T}(\mathbf{x})$, then we call it deformable image registration (DIR).

Boosting. The underlying idea of boosting is to combine simple "rules" to form an ensemble such that the performance of the single ensemble member is improved, i.e. "boosted". Let $h_1, h_2, \ldots, h_N$ be a set of hypotheses, and consider the composite ensemble hypothesis $f(\mathbf{x}) = \sum_{n=1}^{N} \alpha_n h_n(\mathbf{x})$. Here α_n denotes the coefficient with which the ensemble member h_n is combined; α_n and the hypothesis h_n are to be learned within the boosting procedure [2]. While boosting is not algorithmically constrained, most boosting algorithms consist of an iterative procedure where a hypothesis is learned with respect to (adaptively [7]) changed data distributions.

We hypothesize that the boosting theory can be transferred into the domain of deformable image registration. In this paper we present an iterative and adaptive algorithm for boosting deformable image registration (DIRBoost).

2. METHODS

The DIRBoost algorithm is described in Figure 1. DIRBoost is an iterative and adaptive algorithm. In order to boost a DIR algorithm, we require an estimate of local registration accuracy. For this assessment we utilize a landmark-based classifier (RED). Figure 2 illustrates the application of RED within DIRBoost. In the following we describe the employed RED system and explain further details of the DIRBoost algorithm.

We employ an extended version of the method for automatic detection of registration errors described in [4]. The method is based on supervised learning of local alignment patterns, which are captured by statistical image features at distinctive landmark points lm_{ref}. A classifier cascade, using an optimal multi-feature model, classifies local alignments into three quality categories: correct (CA), poor (PA), and wrong alignment (WA). The quality categories are based on the landmark registration error (LRE), i.e. the amount of misalignment between corresponding landmark positions in the fixed and registered moving image; and defined as follows:

$$\begin{aligned} \text{CA} &\doteq \{\text{LRE} \leq 2 \text{ mm}\} \\ \text{PA} &\doteq \{2 < \text{LRE} < 5 \text{ mm}\} \\ \text{WA} &\doteq \{\text{LRE} \geq 5 \text{ mm}\} \end{aligned}$$

Algorithm DIRBoost
Input: I_F, I_M, and initial deformation field $\mathbf{u}_1$
binary masks M_{FROI}, M_{MROI} defining the region of interest
registration error detection method **RED**
deformable image registration algorithm **DIR**
integer N specifying number of iterations

Initialize the boosting masks:

$$M_{FBoost} = M_{FROI}$$

$$M_{MBoost} = M_{MROI}$$

Do for $n = 1, 2, ..., N$

1. Set weight field

$$\mathbf{a}_n(\mathbf{x}) = \begin{cases} \mathbf{1} : M_{FBoost}(\mathbf{x}) == 1 \\ \mathbf{0} : \text{otherwise.} \end{cases}$$

2. Call **DIR** algorithm

$$\text{DIR}(\mathbf{u}_n, I_F \cdot M_{FBoost}, I_M \cdot M_{MBoost});$$

get back hypothesis $h_n : \mathbf{u}_{n+1} \mapsto \Omega(I_M)$.

3. Update hypothesis

$$\mathbf{u}_{n+1} \leftarrow \mathbf{u}_{n+1} \cdot \mathbf{a}_n \wedge \mathbf{u}_n \cdot \mathbf{a}_n^c,$$

with $\mathbf{u}_{n+1}$ and $\mathbf{u}_n$ being folding-free, the folding preventing join operation($\wedge$) yields a topology preserving field update.

4. Call **RED** to estimate the error

$$\text{REDpred} = \text{RED}(I_F, I_M, \mathbf{u}_{n+1}, M_{ROI}),$$

where REDpred is a registration error assessment at every landmark lm_{boost} estimated by the quality classes: CA, PA, WA (Correct, Poor and Wrong Alignment).

5. Compute Voronoi tessellation
REDvoron = 3D Voronoi tessellation based on REDpred.

6. Generate boosting masks
a) REDvoron(WA) = WA areas of REDvoron.
b) M_{FBoost} = expansion of REDvoron(WA) to the nearest neighbor non-WA landmarks.
c) $M_{MBoost} = M_{FBoost} \circ \mathbf{T}_{n+1}^{-1}$

Output hypothesis:
Select the best hypothesis h_f according to REDpred(WA), and in case of a tie take REDpred(CA) into account

$$h_f = \mathbf{u}_{f+1},$$

where h_f evolves from the hypotheses sequence:

$$H_{n=1}^{f} := \mathbf{u}_{n+1} \cdot \mathbf{a}_n \wedge \mathbf{u}_n \cdot \mathbf{a}_n^c,$$

Here, $\cdot$ denotes the Hadamard product with entry-wise multiplications of field vectors $\mathbf{u}$ and $\mathbf{a}$, and with $\mathbf{a}^c$ being the binary complement of $\mathbf{a}$.

Fig. 1. The deformable image registration boosting algorithm.

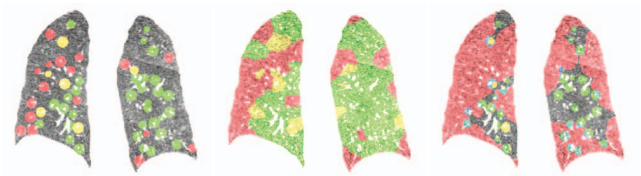


Fig. 2. Visualization (from left to right) of the registration error detection, Voronoi tessellation and mask expansion to nearest non-WA landmarks.

To automatically classify a previously unseen registration, a set of lm_{boost} landmarks is automatically detected on I_F, and the trained classifier is employed to predict for each landmark sample one of the above defined quality categories. That way, an assessment of local image registration quality is obtained. In the following we refer to the entity of this supervised Registration Error Detection method by RED. We denote the registration quality predictions made on lm_{boost} by REDpred. That is, REDpred(WA) = number of lm_{boost} landmarks predicted with an estimated misalignment of more than 5mm.

In classifier boosting each ensemble member is weighted by a coefficient α_n that is usually related to the accuracy of the single hypothesis h_n (cp. Section 1). In adaptive boosting this means the weight of previously correctly classified instances decreases and the weight of misclassified instances increases. The aim of this adaptive weighting is to focus the training of subsequent classifiers on these difficult, yet incorrectly predicted instances. We propose a binary weighting scheme to achieve a similar effect for boosting of deformable image registration, and we define a weight field $\mathbf{a}_n(\mathbf{x})$ such that each voxel $(\mathbf{x})$ in the region of interest (ROI) is assigned the binary weight one or zero, i.e. was previously misaligned or correctly registered. In practice that yields simple binary masks which can be applied to the input images I_F and I_M in order to focus the registration on previously incorrectly mapped voxels. Each generated deformation field $\mathbf{u}_{n+1}$ is based on the previous deformation field $\mathbf{u}_n$. Due to the binary weighting scheme the update is simply an addition of $\mathbf{u}_{n+1}$ to $\mathbf{u}_n$ weighted by $\mathbf{a}_n$ and the complement of $\mathbf{a}_n$ respectively. However, irregularities in the underlying topology of the updated field should be avoided. Given that the DIR algorithm generates folding-free deformation fields, only the join borders are prone to irregularities. If applicable post registration (cp. [6]), the same regularization scheme employed within a DIR algorithm can be imposed (locally) on the updated field. Further, we propose a weighting function that can simply be stated by: "expand to the nearest non-WA landmark". This means that each non-WA Voronoi cell neighboring a WA voxel is merged halfway to the existing boosting mask. See Figure 2 for a visualization of the boosting mask expansion.

3. EXPERIMENTS

This section describes the application of the proposed DIRBoost algorithm on a set of five pulmonary CT breath-hold inspiration-expiration scan pairs. The RED classifier was trained on this data (leave-one-scan pair-out), and the base registrations were computed using the NiftyReg registration package. DIRBoost was then applied on the computed base registrations, employing the trained RED classifier and the NiftyReg DIR algorithm.

3.1. Image Data

Five patients (male, ages 51-75yrs) were chosen randomly from a lung cancer screening database, each with a breathhold inspiration and a breathhold expiration CT scan, made in the same session. The inspiration scan was created using a low-dose protocol (30mAs) whereas the expiration scan was ultra-low-dose (20mAs), both with a slice spacing of 0.70 mm and pixel spacing between 0.63mm and 0.77mm.

3.2. Landmark Sets

We use two different landmark sets: a reference landmark set with landmarks denoted lm_{ref}, and a boosting landmark set with landmark lm_{boost}.

Reference landmark set A set of 100 distinctive landmark points lm_{ref} was automatically defined in the fixed image (expiration scan) of each scan pair. Each landmark in the fixed image was then matched with the corresponding point in the moving image. All landmark correspondences were manually established by the first author using the publicly available matching tool described in [3]. In this work the reference landmark set is used for the evaluation of DIRBoost, and the training and cross-validation of the RED classifier.

Boosting landmark set The lm_{boost} landmarks are acquired by the same method [3] as lm_{ref} landmarks but parameter settings are chosen such that theoretically 1,000 landmarks can be determined. In practice only about 500 landmarks could be determined, which can be explained by the relatively strong image noise, and the occurrence of severe emphysema appearing as diffuse lung tissue or air-filled bullae. Because of the adapted parameter settings, lm_{ref} is practically disjunct to lm_{boost}. Over all scan pairs, the averaged mean distance of lm_{ref} landmarks to nearest lm_{boost} landmarks is 6.8mm, with only in scan pair 4 and 5 one identical landmark position, and an averaged maximum distance of 12.9mm.

3.3. DIR algorithm

We opted for the DIR algorithm of the NiftyReg package [6] (version 1.3) for our boosting experiments.

NiftyReg contains a global and local registration algorithm. The global registration is based on a block-matching technique and the local registration is based on a B-Spline deformation model. Control point positions are optimized using a conjugate gradient ascent optimizer. The objective function is composed of the normalized mutual information as a metric and optionally, the bending energy (BE) and the squared Jacobian determinant as a penalty terms. The optimizer can be run until convergence of the registration parameters or until a maximum number of iterations is reached. NiftyReg includes a folding correction scheme that can be conducted concurrently with the registration process or separately as a post-registration process.

Configuration and application of NiftyReg. In DIRBoost we distinguish between the base registration (iteration $n = 0$), which possibly involves an affine registration, and the actual boosting registration starting at $n = 1$.

Computation of base registration: We use the registration set-up and parametrization proposed in [5] but perform all registrations until convergence.

DIR set-up for boosting: We use the same parametrization as employed in the base registration but a reduced set-up consisting of two registration steps and a folding correction step. In addition we require auxiliary external computations: (i) Transformation of $\mathbf{u}$ into a coarse B-spline grid $\mathbf{b}$. (ii) Gaussian smoothing at join border (i.e. between $\mathbf{u}_{n+1}$ and $\mathbf{u}_n$) of size of B-spline grid-spacing (6 voxel) with $\sigma = 3$voxels prior to transformation into B-Spline representation to ease folding correction.

We list the registration steps using the following abbreviations:
-maxit: Maximal number of iterations per level.
-ln: Number of levels to perform.
-lp: Only perform the first levels [ln].
-be: Weight of the bending energy penalty term.
-jl: Weight of log of the Jacobian determinant penalty term.

1. Coarse alignment step: input as listed in Fig. 1 with $\mathbf{b}$ from auxiliary method (i), and parameters -ln 4 -lp 3 -be 0.0001 -jl 0.001 -maxit 500.

2. Fine alignment step: input as listed in Fig. 1 using in step 1 generated output $\mathbf{b}$ as input and parameters -ln 3 -lp 2 -be 0.001 -jl 0.001 -maxit 500.

3. Folding correction step: here we first perform the auxiliary computation (ii), then apply the folding correction scheme on the final resolution -ln 1.

4. EVALUATION & RESULTS

The results of our experiments are depicted in Table 1 and Figure 3. Whereas Figure 3 is based on lm_{boost} landmarks, the results in Table 1 are obtained on the separate reference landmark set lm_{ref}. In addition to the LRE measure, the amount of folding and boundary overlap is calculated, both displayed in number of voxel occurrences. For the boundary evaluation we computed $|M_{FROI} - (M_{FROI} \circ \mathbf{T}_f)|$, and allowed two voxel tolerance to both inside and outside the ROI to account for possible inaccuracies of the lung segmentations. The boosting results in Table 1(b) are based on the hypothesis selection rule described in Figure 1. Note: Classification bias in RED is avoided by the use of a rotation training set (leave-one-scan pair-out). DIR bias towards lm_{ref} landmarks is avoided by using separate landmark sets for boosting (lm_{boost}), moreover the DIR algorithm does not involve any landmark matching, optimization is purely intensity-based.

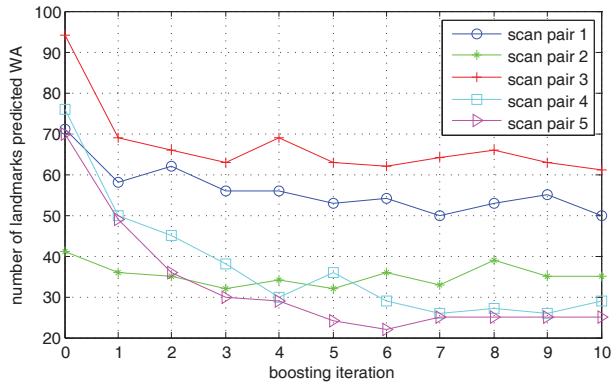


Fig. 3. Number of landmarks with predicted misalignment larger than 5mm (REDpred(WA)) plotted over progress of boosting. Iteration 0 denotes the base registration.

The proposed DIRBoost algorithm could establish for most scan pairs within about six iterations a final hypothesis. On average this yielded a corrected registration of about 50% of the initially wrongly registered areas. While boundary alignment and topology preservation were maintained, DIRBoost could also improve on the mean landmark registration error, on average by 24% (cp Table 1 and Figure 1).

5. DISCUSSION & CONCLUSION

In Table 1 one particular case stands out: in scan pair 2, the maximum LRE value of the boosted registration is larger than originally in the base registration. We looked up the location of this particular reference landmark, and found that RED predicted this region as PA (poor alignment). We assume that the RED classification accuracy can be improved by an extended reference landmark set, such that per scan pair more landmarks at noisy and less distinctive areas are included in the training set. However, we believe that the current RED classification accuracy was not significantly limiting the boosted registration results.

The presented DIRBoost algorithm can boost any DIR algorithm within the given properties of the employed DIR algorithm such as similarity measure and transformation model. Those possibly limiting properties might be also addressed within DIRBoost, similar to modified versions of AdaBoost.

In conclusion, an algorithm for boosting of deformable image registration was presented. In our experiments on intra-patient breathhold inspiration-expiration CT scans, DIRBoost boosts about 50% of wrong alignments and improves on the average LRE of the base registration by 24%. This improvement should be appreciated in light of the already outstanding performance of the applied NiftyReg base registration.

(a) Base Registration

scan pair	LRE min	max	mean	std	>5	>10	>20	folding	overlap
1	0.00	15.58	1.61	1.81	4	1	0	0	0
2	0.00	21.27	1.97	3.09	10	3	1	0	0
3	0.00	15.17	1.53	2.20	7	2	0	0	0
4	0.00	16.20	2.60	2.89	10	4	0	0	0
5	0.00	34.19	3.93	6.91	17	11	6	0	0
avg	0.00	20.48	2.33	3.38	48	21	7	0	0

(b) Boosted Registration

scan pair	LRE min	max	mean	std	>5	>10	>20	folding	overlap
1	0.00	5.13	1.46	1.06	1	0	0	0	0
2	0.00	28.37	1.88	3.37	6	3	1	0	0
3	0.00	12.21	1.31	1.67	2	1	0	0	0
4	0.00	15.85	2.36	2.41	8	2	0	0	0
5	0.00	20.41	1.83	2.71	6	3	1	0	0
avg	0.00	16.39	1.77	2.24	23	9	2	0	0

Table 1. Evaluation of base and boosted registration. Assessment of landmark registration error (LRE) is based on the reference landmark set. Decimal values are in [mm]. >x columns count landmarks with misalignment larger than x mm. Folding and boundary overlap values depict voxel occurrences.

6. REFERENCES

[1] K. Murphy et al, "Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge," *IEEE TMI*, 2011.

[2] R. Meir and G. Raetsch, "An introduction to boosting and leveraging," in *LNCS*, 2003, pp. 118–183.

[3] K. Murphy, B. van Ginneken, S. Klein, M. Staring, B.J. de Hoop, M.A. Viergever, and J.P.W. Pluim, "Semi-automatic construction of reference standards for evaluation of image registration," *MIA*, 2011.

[4] S.E.A. Muenzing, K. Murphy, B. van Ginneken, and J.P.W. Pluim, "Automatic detection of registration errors for quality assessment in medical image registration," in *SPIE*, 2009, vol. 7259, pp. 72590K–72590K–9.

[5] M. Modat, J. McClelland, and S. Ourselin, "Lung registration using the NiftyReg package," in *Med. Img. Anal. for the Clinic*, 2010, pp. 33–42.

[6] M. Modat, G.R. Ridgway, Z.A. Taylor, M. Lehmann, J.Barnes, D.J. Hawkes, N.C. Fox, and S. Ourselin, "Fast free-form deformation using graphics processing units," *CMPM*, vol. 98, no. 3, pp. 278 – 284, 2010.

[7] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *JCSS*, vol. 55, pp. 119–139, August 1997.