# Deformable Registration of Chest CT Scans with Adaptive Local Mutual Information

Dante De Nigris[1], D. Louis Collins[2], and Tal Arbel[1]

[1] McGill University, Centre for Intelligent Machines,
{dante,tal}@cim.mcgill.ca,
[2] McGill University, Dept. of Biomedical Engineering,
louis@bic.mni.mcgill.ca

**Abstract.** We propose a novel method for non-rigid chest CT registration based on a new, adaptive local measure which evaluates gradient orientation similarity, and an adaptive pixel sampling scheme. A computationally efficient estimation of gradient orientations is also proposed based on voxel neighborhood rigidity. We have applied our method to the Evaluation of Methods for Pulmonary Image Registration 2010 (EMPIRE10) Challenge[3] datset, where we have shown good image correspondence results in terms of lung boundaries, fissures and expert selected landmarks.

## 1 Introduction

In this work, we propose a registration framework which embeds a new localized similarity metric expressed as an orientation similarity measure based on a local approximation to Mutual Information (MI). Other local MI metrics have been devised [1, 2], however, our metric is adaptive, in that it can loosen or tighten its constraints depending on the modalities being registered and the scale at which the images are being registered. We also propose an adaptive multi-scale pixel sampling scheme and a computationally efficient method for estimating gradient orientations.

Even though our method was originally developed with the context of brain MRI/US deformable registration in mind[3], we consider it important to study the advantages and disadvantages of our approach in different registration contexts. In this work, we evaluate the performance of our method in the context of chest Computer Tomography (CT) deformable registration. The challenge of such context is considerably different and lies in obtaining a transformation that accurately matches lung boundaries, fissures and bronchioles in a physically plausible fashion. Even though the registration is performed on the same modality, the scans are from different time points and the quality and dynamic intensity range of the scans can be considerably different. Hence, it remains of critical importance to implement a registration method robust to different intensity responses and significant levels of noise.

[3] http://empire10.isi.uu.nl/

## 2 Method

### 2.1 Local Mutual Information - Previous Work

In [1], the authors developed a local similarity metric based on the analytical limit of MI as the window of observation approaches the voxel size. The expression for local MI (LMI) is obtained by modeling both images with a first-order Taylor expansion and is a monotonically decreasing function of the angle, $\theta$, between the gradient orientations,

$$LMI(\theta) = C_d + \log_2 |\sin(\theta)| \tag{1}$$

where $C_d$ is a constant that depends on the dimension of the image.

It is not possible to build an energy function by simply summing Eqn. 1 over a set of points in the image, since an extremum would appear whenever any of the points has a minimal inner angle. To circumvent this issue, the authors simplify the expression to,

$$LMI_2(\theta) = \frac{1}{2}\cos^2(\theta) \tag{2}$$

which has a smoother shape and whose energy function exhibits an extremum only when there is a collective coherence in terms of orientation similarity.

In related work [2], the authors preserve the dynamics of the original expression and eliminate the singularity by including an $\epsilon$ factor,

$$LMI_3(\theta) = \log_2 (\epsilon + |\sin(\theta)|) \tag{3}$$

The two localized metrics have significantly different *coherence-selectivity* trade-offs. The metric expressed by Eqn. 2 represents a coherent metric in the sense that it is maximal when a majority of points show some degree of image correspondence. In contrast, a selective metric, such as Eqn. 3, has stricter correspondence constraints and therefore exhibits an extremum as soon as a few points comply with such constraints. A measure that favors coherence will tend to be smoother but less accurate because it will effectively average over all sampled pixels.

### 2.2 Adaptive Local Mutual Information

Our work builds on previous work [1, 2] in which MI is locally approximated by a metric appropriately derived from the energy function. We propose an adaptive local orientation-based similar metric (ALMI) that addresses the coherence-selectivity trade-off directly as follows:

$$ALMI(\theta; K, \theta_c) = 2 - \frac{1}{1 + e^{-K(\theta - \theta_c)}} - \frac{1}{1 + e^{-K(\pi - \theta - \theta_c)}} \tag{4}$$

The selectivity of this smooth sigmoid-based function is adaptable by varying the curvature, $K$, and cutoff angle, $\theta_c$. The cutoff angle identifies the angle at

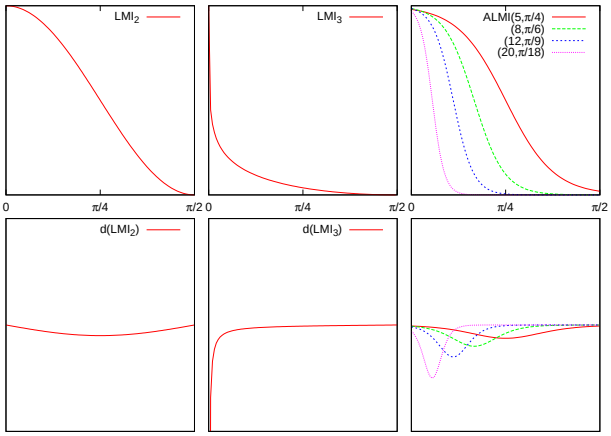



Fig. 1: The top row shows the two previous proposed localized measures and our proposed measure as a function of $\theta$. The bottom row shows their respective derivatives.

which the localized metric decays to half its maximum value. The curvature, $K$, defines the gradient of the curve. For example, a high curvature value will characterise a steep slope with a short span.

Figure 1 illustrates how the two previously proposed measures compare to various configurations of our proposed expression. An important advantage of ALMI is that it saturates smoothly to a maximum as $\theta \to \{0, \pi\}$. This property allows the energy function to reach a stable state as the optimizer approaches an extremum.

Since all the experiments in this work are performed with CT images of relatively high quality which preserve the structures of interest, we set the parameters of ALMI to $(K = 15, \theta_c = \frac{\pi}{9})$, which is a relatively selective profile.

### 2.3 Gradient Orientation Approximation

A metric based on gradient orientations incurs a computational cost related to evaluating the gradient at each point of interest. We propose a simplifying implementation in which the gradients are computed only once for both the fixed and the moving image. Subsequent gradient orientations of the transformed moving image are estimated from the initial orientations.

In order to achieve non-rigid registrations, we make the simplifying assumption that the transformation in the voxel neighborhood can be well approximated by a locally rigid transformation. Hence, we can estimate the gradient orientation of a point by multiplying the gradient orientation of the point prior to transformation with an estimated rotation matrix as follows,

$$\begin{aligned}\theta(\mathbf{x}) &= \angle(\nabla I_f(\mathbf{x}), \nabla I_m(T(\mathbf{x}))) \\ &\approx \angle(\nabla I_f(\mathbf{x}), R \cdot \nabla I_m(\mathbf{x}' = T(\mathbf{x})))\end{aligned}$$

Such a scheme eliminates the effect of intensity-interpolation artifacts and minimizes the expense of using high quality gradient operators.

### 2.4 Pre-Processing

Our method makes use of pre-segmented lung masks. Pixels within the lung boundaries have a value of 1, while pixels outside the lung boundaries have a value of 0. Since our local similarity measure tends to match borders and blobs, it is necessary to dilate the initial lung masks by a couple of pixels to provide sufficient image support for the lung boundaries.

Dilation is performed by blurring the initial masks with a Gaussian kernel with a standard deviation equal to the size of the voxels and then selecting the voxels with intensity above a threshold value of 0.01. The dilated lung masks are then used as initial sampling masks for the fixed and moving image.

Additionally, the initial volumes are down-sampled by a factor of two to reduce the computational load. Down-sampling is performed by convolving the original image with a Gaussian kernel with a standard deviation equal to twice the pixel spacing and then resampling the image with a corresponding pixel grid of increased pixel spacing. The down-sampled volumes are then registered in a multi-scale framework, see Section 2.5. The result of this registration is then used as an initial point for a second registration which makes use of the full resolution volumes and is only performed at the original image level (i.e. no image pyramid). The first registration stage represents most of the processing time and is expected to account for most of the final deformation. The second registration stage has a reduced number of iterations and is mainly expected to refine the final deformation.

### 2.5 Adaptive Sampling

In addition to the metric proposed above, we propose an adaptive multiscale pixel selection scheme for the first registration stage. Rather than adopting a standard low-pass image pyramid [4] in which the images are both blurred and down-sampled at each level, we use a smoothing image pyramid in which only the scale of Gaussian blurring is increased at each level. Figure 2 illustrates how the gradient orientations change with different blurring scales. As the scale of diffusion is increased, the spatial support of a structure (i.e. the extent of image area where the gradient orientations accurately indicate the orientation of a structure of reference) tends to increase proportionately to the ratio of the current scale to the previous scale.

Performance gains can be attained by focusing on a cleverly selected subset of voxels in the images. Our conjecture is that the gradient magnitude is a valid indicator of the *reliability* of an estimated gradient orientation, and therefore high gradient voxels are selected. Furthermore, pixels with low gradients typically appear in homogeneous regions and thus do not have structure to drive the local non-linear registration. Hence, we adapt the sampling mask, $M(\mathbf{x})$, in relation

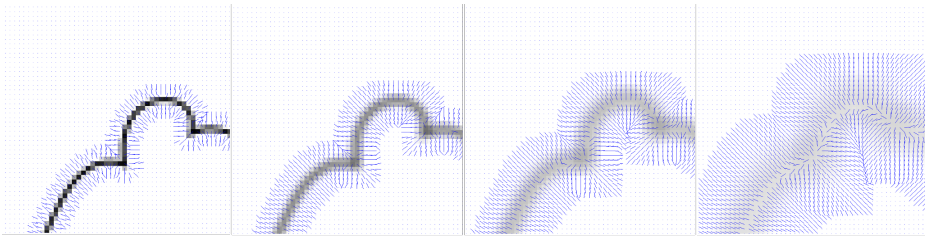


Fig. 2: Gradient Orientations (short blue lines) at different blurring scales, $\sigma \in 0, 1, 2, 4$

to the current Gaussian scale, as follows,

$$M(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad |\nabla I_f(\mathbf{x})| > g(p_\sigma) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

where $g(p)$ is the threshold value that captures the top $p$ percentile of the fixed image gradient magnitude and $p_\sigma$ satisfies the inequality $p_{\sigma_n+1} > p_{\sigma_n}$.

In this work, we chose the percentile for each scale based on a qualitative assessment of how well the relevant features are captured. In other words, at the finest scale we search for the percentile that best captures the lung boundaries, major fissures and bronchioles. For coarser scales the percentile is increased to account for the spatial support related to such features. Table 1 lists some of the relevant parameters of the registration method.

### 2.6 Optimization

Our proposed method was implemented by extending the Elastix Toolbox [5]. We chose a BSpline transformation to characterize the deformation with a uniform grid of knots separated by a spacing of 8mm x 8mm x 8mm. The grid of knots is also adapted in relation to the image pyramid level. In other words, there are fewer knots at coarser image levels. Such a scheme reduces the computational load since the number of transformation parameters is greatly reduced. The schedule for the grid of knots is specified in Table 1. It is of interest to note that no regularization was embedded in our method. In other words, the energy function does not incorporate a penalty for non-regular deformations.

The optimization is performed by a gradient decent optimizer with an adaptive gain[4] which operates until reaching a specified number of iterations. Furthermore, instead of using all the pixels found in the pixel selection mask, a subset of the pixels is randomly chosen[5] at every iteration of the optimizer, thereby further reducing the computational load while preserving robustness to local minima.

[4] Identified in Elastix under the name of *AdaptiveStochasticGradientDescent*
[5] Identified in Elastix under the name of *RandomSparseMask*

| Stage | Level | Gaussian Blurring Std Deviation (in mm) | Sampling Percentile $p_\sigma$ | BSpline Grid Downsampling Factor |
|---|---|---|---|---|
| 1 | 6 | 32 x 32 x 32 | 100 % | 32 |
| | 5 | 16 x 16 x 16 | 100 % | 8 |
| | 4 | 8 x 8 x 8 | 100 % | 4 |
| | 3 | 4 x 4 x 4 | 80 % | 2 |
| | 2 | 2 x 2 x 2 | 60 % | 1 |
| | 1 | 1 x 1 x 1 | 40 % | 1 |
| 2 | 0 | 1 x 1 x1 | 40 % | 1 |

Table 1: Relevant parameters of multi-scale registration. The image pyramid is processed in decreasing order of Gaussian blurring. Hence, level 6 is processed first, and level 0 is processed last. The pixel selection percentile indicates the top percent of high gradient magnitude pixels selected from the fixed image. The Bspline grid downsampling factor effectively defines the grid of knots at each level. For example, a downsampling factor of 2 and a final grid spacing of 8 x 8 x 8, results in a grid spacing of 16 x 16 x 16.

## 3 Results and Discussion

We tested our method with the 20 scan pairs provided by the EMPIRE10 Challenge, a transparent approach for comparing multiple methods on a common dataset. Table 2 shows the detailed results of our method for each scan pair and each scoring criteria. For details on the scoring criteria, the reader is invited to refer to [6]. For a comprehensive list of rankings and scores, the reader is invited to visit the EMPIRE10 website[6].

In terms of scores, our best results were for the landmark criteria (like bronchioles) where our method shows an average localization error of 1.10 mm and an average ranking of 10.37 (of 34). Our top 10 best registration results in this category show an average error of less than 0.72 mm, which corresponds to one pixel or less. In our case, we believe that these results are due to both the nature of the similarity metric employed in our framework, and the adaptive sampling scheme proposed. The sampling scheme allows the algorithm to focus on the features of interest, whereas the similarity metric provides robustness to intensity response variations across images by matching gradient orientations, as opposed to maximizing some notion of intensity correspondence.

The registration error for landmarks is evaluated as the distance to whichever observer's point is closest. Furthermore, the landmark points chosen by the observers must be within 3 mm to be considered valid. Hence, the scoring critera evaluates how well the landmark is registered to *any* of the observers' points. The justification is that all observers are correct as long as they exist within 3 mm of each other and therefore represent various "true" solutions. Hence, the problem is assumed to be ill-posed with multiple solutions, potentially more than

[6] http://empire10.isi.uu.nl/mainResults.php

| | Lung Boundaries | | Fissures | | Landmarks | | Singularities | |
|---|---|---|---|---|---|---|---|---|
| **Scan Pair** | **Score** | **Rank** | **Score** | **Rank** | **Score** | **Rank** | **Score** | **Rank** |
| **01** | 0.10 | 22.00 | 0.37 | 15.00 | 1.92 | 8.00 | 0.09 | 29.00 |
| **02** | 0.00 | 11.00 | 0.00 | 15.00 | 0.33 | 2.00 | 0.00 | 12.50 |
| **03** | 0.00 | 25.00 | 0.00 | 12.50 | 0.32 | 3.00 | 0.00 | 12.00 |
| **04** | 0.00 | 22.00 | 0.00 | 16.50 | 1.39 | 17.00 | 0.00 | 14.00 |
| **05** | 0.00 | 28.00 | 0.00 | 16.00 | 0.00 | 5.50 | 0.00 | 31.00 |
| **06** | 0.00 | 16.00 | 0.00 | 25.00 | 0.34 | 12.00 | 0.00 | 28.00 |
| **07** | 0.28 | 24.00 | 2.51 | 22.00 | 1.98 | 10.00 | 0.18 | 29.00 |
| **08** | 0.02 | 22.00 | 0.39 | 22.00 | 0.71 | 6.00 | 0.07 | 30.00 |
| **09** | 0.00 | 14.00 | 0.05 | 28.00 | 0.49 | 1.00 | 0.00 | 28.00 |
| **10** | 0.01 | 21.00 | 0.00 | 15.00 | 1.53 | 11.00 | 0.00 | 13.50 |
| **11** | 0.21 | 23.00 | 0.00 | 5.00 | 0.63 | 2.00 | 0.09 | 28.00 |
| **12** | 0.00 | 10.00 | 0.00 | 13.50 | 0.01 | 8.00 | 0.00 | 14.50 |
| **13** | 0.10 | 33.00 | 0.27 | 30.00 | 1.38 | 31.00 | 0.16 | 31.00 |
| **14** | 0.40 | 25.00 | 3.23 | 15.00 | 3.15 | 15.00 | 0.04 | 27.00 |
| **15** | 0.00 | 29.00 | 0.00 | 26.00 | 0.61 | 4.00 | 0.00 | 27.00 |
| **16** | 0.33 | 34.00 | 0.92 | 28.00 | 2.28 | 30.00 | 0.55 | 32.00 |
| **17** | 0.08 | 34.00 | 0.03 | 6.00 | 1.13 | 21.00 | 0.28 | 33.00 |
| **18** | 0.16 | 23.00 | 0.54 | 7.00 | 1.89 | 8.00 | 0.16 | 28.00 |
| **19** | 0.00 | 14.00 | 0.00 | 26.00 | 0.46 | 6.00 | 0.00 | 14.50 |
| **20** | 0.07 | 22.00 | 2.50 | 13.00 | 1.40 | 7.00 | 0.22 | 29.00 |
| | | | | | | | | |
| **Avg** | 0.09 | 22.60 | 0.54 | 17.82 | 1.10 | 10.37 | 0.09 | 24.55 |
| | | | | | | | | |
| **Average Ranking Overall** | | | | | | | | 18.83 |
| **Final Placement** | | | | | | | | 22 |

Table 2: Results for each scan pair, per category and overall. Rankings and final placement are from a total of 34 competing algorithms.

three. We argue that, instead of evaluating the distance to the closest candidate point, it would be more reasonable to evaluate the landmark registration error based on some notion of agreement between observers (e.g. a weighted distance from the landmark identified by each observer or explicitly cross-validating landmarks with multiple observers) or to take the inherent ill-posedness into account. Figure 3 illustrates a potential worst-case scencario where two considerably different registration results (shown in red) lead to the same localization error even though one shows a more satisfactory agreement with all the observers.

As in any other registration framework, it is necessary for coarse scale structures like lung boundaries to be accurately resolved so that fine scale structures like bronchioles can be pulled to their corresponding landmarks. Our proposed method shows good results in terms of lung boundary matching, with an average percentage error of 0.09, meaning 0.09% of tested pixels near the lung bound-

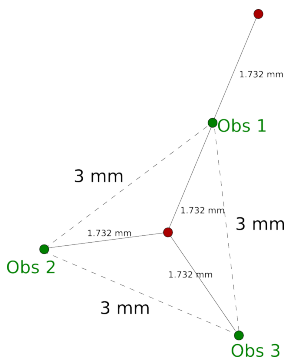



Fig. 3: Landmark Localization Error. Three observers identify different landmark localizations (green circles) with a maximum allowed distance between them of 3 mm. One method registers the landmark to the center of the triangle (red circle), while the other registers the landmark considerably farther from all but one observer. However, both methods share the same localization error, since the distance to the closest candidate point is the same (1.732 mm).

aries were penalized (unregistered). Furthermore, the top four scan pairs show a near-perfect registration with a percentage error less than 1e-5. In other words, less than 0.00001% of the lung boundary pixels suffered a penalization. It is interesting to note that most of the participating methods had very competitive lung boundary registration results. For example, even though in Case 16 our method has a percentage error less than 1e-5, it is actually ranked $16^{\text{th}}$.

Nonetheless, for some of the scan pairs we had lower scores for the lung boundary tests, which can be explained by mainly two factors. First, we observed that lung boundary mismatches tend to occur in the bottom section of the lung. In particular, the method seems to have difficulty in accurately resolving "corner" regions of high curvature such as in the case shown in Figure 4. In addition, we found that such behaviour typically occurs only when the corner is found in the fixed image and the corresponding region in the moving image has to compress into the corner, as can be seen in Figure 4.

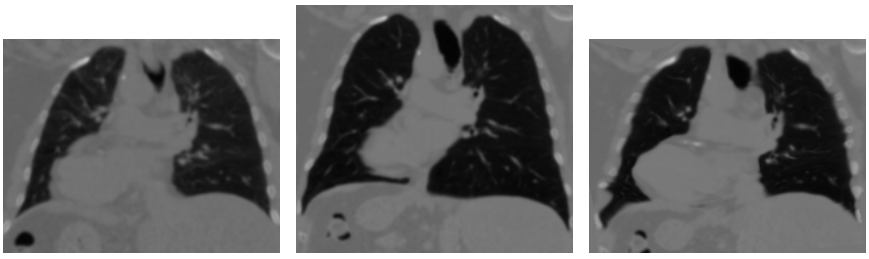
The inverse case, where a tight corner in the moving image has to expand towards a lower curvature region in the fixed image, tends to be well resolved, as shown in Figure 5. We also found that the cases where the lung boundaries have considerable mismatch also had a relatively higher average landmark error. Therefore, improving the registration of such lung boundary regions will likely improve the landmark registration results. We believe a considerable extent of those challenges can be addressed by improving the symmetry of the pixel sampling method, that is, by selecting pixels corresponding to features of the fixed image as well as pixels corresponding to moving image features.

We noticed that cases 16 and 17 showed particularly poor results. Both cases are characterized by considerably coarse pixel spacings: Case 16 has a pixel grid

of 0.97656 mm x 0.97656 mm x 2.5 mm and Case 17 has a grid of 0.97656 mm x 0.97656 mm x 2.0 mm. Larger, non-isotropic voxel pixels sizes are clearly detrimental to our method. In particular, since most of the processing is performed on the down-sampled volumes, the initial pixel grids for cases 16 and 17 are extremely coarse, leading to a degradation of the gradient orientation approximation. To avoid such losses in performance, it would be of interest to either avoid the down-sampling stage entirely and perform the multi-scale registration starting with the original volumes, or to use a fixed common voxel size for the down-sampling stage of all scan pairs.

Finally, in our implementation, we chose to avoid imposing additional regularization constraints in our energy function. This led to penalties in terms of deformation singularities, with an average ranking of 24.52 and an average score of 0.09

Future work will explore using the penalty term proposed in [7] which provides an excellent option for this context since it addresses the scoring criteria directly and also penalizes deformations which are locally non-rigid, thereby supporting the assumptions made for our approximation of gradient orientations.

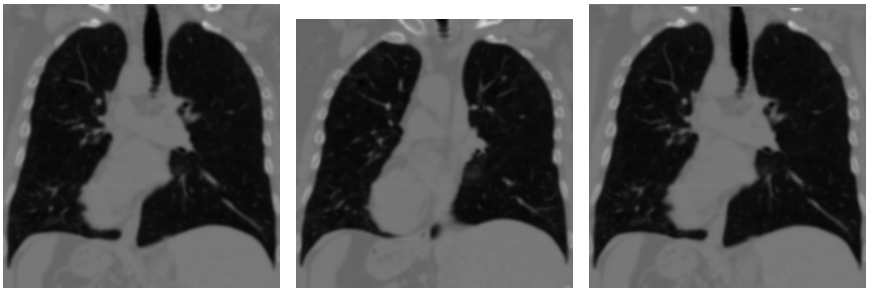


(a) Fixed (b) Moving (c) Deformed Moving

Fig. 4: Example of a Mismatch in Bottom Lung Boundaries. Notice how the bottom-left corner of the lung boundary does not show good correspondence in images (a) and (c)

## 4 Conclusions

We have presented a new method for non-rigid registration of CT chest scans using a similarity metric based on gradient orientations and an adaptive pixel sampling scheme. We have shown good results in terms of lung boundary, landmark (i.e. bronchioles) and fissure matching. In particular, we have shown high landmark accuracy, comparable or better than the pixel size. We believe that our method can be further improved by adopting a symmetric pixel sampling method. Finally, incorporating a regularization term can serve to both enforce



(a) Fixed (b) Moving (c) Deformed Moving

Fig. 5: Example of a Good Match in Lung Boundaries. The moving image, (b), exposes a high curvature corner that is deformed to match the respective fixed image region with lower curvature.

the assumptions behind our approach to gradient orientation approximation and to favor locally rigid deformations.

## References


1. Karacali, B.: Information theoretic deformable registration using local image information. IJCV **72** (2007) 219–237
2. Biesdorf, A., Wörz, S., Kaiser, H.J., Rohr, K.: Hybrid spline-based multimodal registration using local measures for joint entropy and mutual information. In: MICCAI. (2009) 607–615
3. De Nigris, D., Mercier, L., Del Maestro, R., Collins, D.L., Arbel, T.: Hierarchical multimodal image registration based on adaptive local mutual information. In: MICCAI. (2010)
4. Lindeberg, T.: Scale-space theory in computer vision. Springer (December 1993)
5. 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) (2010) 196–205
6. 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. In: Grand Challenges in Medical Image Analysis. (2010)
7. Staring, M., Klein, S., Pluim, J.P.W.: A rigidity penalty term for nonrigid registration. Medical Physics **34** (2007) 4098–+