Detection and Correction of Inconsistency-based Errors in Non-Rigid Registration

Tobias Gass, Gabor Szekely, and Orcun Goksel

Computer Vision Lab, ETH Zurich, Switzerland.
{gasst, szekely, ogoksel}@vision.ee.ethz.ch

Abstract

In this paper we present a novel post-processing technique to detect and correct inconsistency-based errors in non-rigid registration. While deformable registration is ubiquitous in medical image computing, assessing its quality has yet been an open problem. We propose a method that predicts local registration errors of existing pairwise registrations between a set of images, while simultaneously estimating corrected registrations. In the solution the error is constrained to be small in areas of high post-registration image similarity, while local registrations are constrained to be consistent between direct and indirect registration paths. The latter is a critical property of an ideal registration process, and has been frequently used to asses the performance of *registration algorithms*. In our work, the consistency is used as a *target criterion*, for which we efficiently find a solution using a linear least-squares model on a coarse grid of registration control points. We show experimentally that the local errors estimated by our algorithm correlate strongly with true registration errors in experiments with known, dense ground-truth deformations. Additionally, the estimated corrected registrations consistently improve over the initial registrations in terms of average deformation error or TRE for different registration algorithms on both simulated and clinical data, independent of modality (MRI/CT), dimensionality (2D/3D) and employed primary registration method (demons/Markov-random-field).

1 Introduction

Image registration is a key technology in medical image analysis. It enables applications such as atlas-based segmentation [1], statistical model building, and automatic landmark detection. In many of these tasks, high registration accuracy is critical, but difficult to achieve since the registration problem is ill-posed [2]. Common approaches use multi-resolution grid hierarchies [3] or transitive/symmetric energy terms [4, 5] to enable robust pairwise or triplet registration. Group-wise registration methods [6, 7] are also often employed for robust registration . Still, there is no guarantee that even a robust registration method will yield accurate results in individual cases. It would therefore be immensely valuable if the accuracy of non-rigid registrations in an existing work-flow could reliably be estimated.

In this context, it was shown that voxel similarity metrics are indeed not good indicators of registration quality due to homogeneous tissue regions, partial volume effects, and anatomies of similar appearance [8, 9]. While independent labelling of landmarks or anatomic regions can be used to measure registration fidelity, this requires invaluable time and effort of trained medical personnel, which an *automatic* registration was supposed to avoid in the first place. To estimate the uncertainty in single pair-wise registrations, different methods such as the Cramér-Rao bound [10] or bootstrap re-sampling [11] have been proposed, increasing the complexity of a registration algorithm by orders of magnitude. Other studies estimated registration accuracy based on intensity

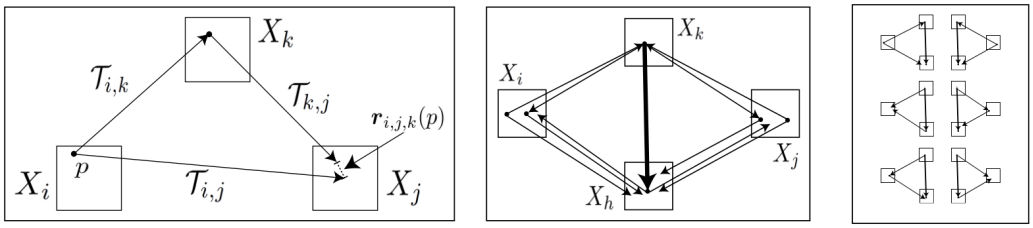


Figure 1: Illustration of an inconsistent registration (left), the residual $\vec{r}_{i,j,k}(p)$ indicates the amount of *inconsistency*. Multiple such loops (center) create *redundancy* that is exploited in CLERC. For instance, the registration in bold is part of 6 such individual loops (right) in this simple example.

measures, e.g. using the multiple Gaussian state-space [12] or voxel-statistics based on active appearance models from registered images [13].

A different line of research investigates the *consistency* of multiple registrations as a measure of registration fidelity [14, 15, 16, 9, 17]. The most common application of this approach has been the evaluation and comparison of registration *algorithms* by computing the residual norm of inconsistencies in pair-wise registrations. Recently, a method was proposed to exploit redundancy in multiple registration circles for estimating the spatial location and magnitude of errors in pair-wise registrations [17]. This method uses a simplified assumption of registration error accumulation, based on a point-wise addition/multiplication of error magnitudes, in order to estimate a dimensionless measure for local registration error, which was validated in a basic synthetic setup. However, it is apparent that such error accumulation assumptions cannot model actual errors of real-world scenarios, as also demonstrated later in our results and comparisons.

In this paper, we introduce a novel method of detecting local error magnitudes and directions, while efficiently correcting such local errors in non-rigid registration. We dub this CLERC for *correcting local errors using registration consistency*. Other work in the literature such as [17] suggested approximating spatial errors assuming their linear accumulation through consecutive registrations. This merely yields (direction-less) values that *may correlate* with the error magnitudes. In contrast, our method estimates error *vectors* with their expected magnitude. We formulate this as a linear least-square problem of minimizing registration inconsistency, which we define based on registration transitivity. Our method requires neither a reference image nor invertible deformations and instead it works as a post-registration filter that can be applied to the collective set of all pairwise registrations obtained from *any* deformable registration method. CLERC then computes estimations of registration errors and corrections efficiently on a coarse grid of control points.

2 Method

We define a registration $\mathcal{R}$ as the set of all pairwise registrations between images of a set $\mathcal{X} = \{X_1, X_2, \cdots, X_N\}$. An image X is a function that maps points in the D-dimensional spatial domain Ω to a space $\mathcal{F}$ of image intensities, e.g. CT Hounsfield units. A non-rigid deformation $\mathcal{T}$ is a mapping from Ω to Ω which is based on a displacement field $\mathcal{D}$ such that $\mathcal{T}(p) = p + \mathcal{D}(p)$ for all points p in Ω. Note that both X and $\mathcal{D}$ are commonly defined on a discrete (Cartesian) regular grid, where non-grid values can be obtained by interpolation. In this notation, deforming an image is a function composition $X \circ \mathcal{T} = X(\mathcal{T})$. Composing two deformations is also a function composition: $\mathcal{T}_2 \circ \mathcal{T}_1(p) = p + \mathcal{D}_2(\mathcal{T}_1(p)) + \mathcal{D}_1(p)$. As introduced in [14], a perfect registration leads to consistent correspondences over the entire set. Multiple definitions for consistency are used in the literature, most notably inverse consistency (or symmetry), transitivity [18] and circle consistency [17]. In our method we use transitivity as a measure of consistency, since it yields information cues from all associated triplets as illus-

trated in Fig. 1(right). In this paper, we use the transitivity as a goal function in order to estimate the local errors in a set of given non-rigid registrations, computed using an arbitrary pair-wise registration method. Assume that each observed $\mathcal{T}_{i,j} \in \mathcal{R}$, which registers image X_i to X_j, is composed of an unknown true deformation $\widehat{\mathcal{T}}_{i,j}$ and an unknown error displacement field $\widehat{\Delta}_{i,j}$ such that $\widehat{\mathcal{T}}_{i,j} + \widehat{\Delta}_{i,j} = \mathcal{T}_{i,j}$. CLERC finds an estimate for all $\widehat{\mathcal{T}}$ and $\widehat{\Delta}$ by minimizing the inconsistency of deformation triplets. The transitivity-based consistency criterion can be defined as follows:

$$\mathcal{C}(\{\widehat{\mathcal{T}}\}) = 1/Z_{\mathcal{C}} \sum_i \sum_{j \neq i} \sum_{k \neq \{i,j\}} \sum_p ||\widehat{\mathcal{T}}_{k,j} \circ \widehat{\mathcal{T}}_{i,k}(p) - \widehat{\mathcal{T}}_{i,j}(p)||^2, \tag{1}$$

where $Z_{\mathcal{C}}$ is a normalization constant (the number of all summands). To incorporate image-based information, we employ the assumption that errors $\widehat{\Delta}$ are relatively small in areas of higher local post-registration similarity. Intuitively, minimizing the criterion

$$\mathcal{E}(\{\widehat{\Delta}\}) = 1/Z_{\mathcal{E}} \sum_i \sum_{j \neq i} \sum_p ||\Psi_{ij}(p) \widehat{\Delta}_{i,j}(p)||^2 \tag{2}$$

penalizes large errors in areas of high post-registration similarity $\Psi_{ij}(p)$, which is the local similarity between the transformed image $X_i(\mathcal{T}_{i,j})$ and X_j. We use local normalized cross correlation (LNCC) [19] as such similarity metric. We set $\Psi_{i,j} = (0.5 + 0.5 \cdot \text{LNCC}\left(X_i\left(\mathcal{T}_{i,j}\right), X_j, \sigma\right))^{\gamma}$, where σ is the window size and γ controls the distribution of this weight.

The CLERC optimization problem is then formulated as follows:

$$\min_{\hat{\mathcal{T}}, \hat{\Delta}} \quad \mathcal{C}(\{\widehat{\mathcal{T}}\}) + \lambda_{\text{err}} \mathcal{E}(\{\widehat{\Delta}\}) + \lambda_{\text{def}} \mathcal{F}(\{\widehat{\mathcal{T}}\}, \{\widehat{\Delta}\}) \tag{3}$$

where λ_{err} weights the influence of the local error term, and λ_{def} acts as a Lagrange multiplier enforcing the the definition of the observed deformation being the true deformation plus the error:

$$\mathcal{F}(\{\widehat{\mathcal{T}}\}, \{\widehat{\Delta}\}) = 1/Z_{\mathcal{F}} \sum_i \sum_{j \neq i} \sum_p ||\widehat{\mathcal{T}}_{i,j}(p) + \widehat{\Delta}_{i,j}(p) - \mathcal{T}_{i,j}(p)||^2 \tag{4}$$

For an efficient solution, we approximate $\mathcal{C}$ by substituting the unknown true deformation $\widehat{\mathcal{T}}_{i,k}$ with the observed deformation $\mathcal{T}_{i,k}$ in the displacement composition as:

$$\begin{aligned} \widehat{\mathcal{T}}_{k,j} \circ \widehat{\mathcal{T}}_{i,k} - \widehat{\mathcal{T}}_{i,j} &= \widehat{\mathcal{D}}_{k,j}(\widehat{\mathcal{T}}_{i,k}) + \widehat{\mathcal{D}}_{i,k} - \widehat{\mathcal{D}}_{i,j} \\ &\approx \widehat{\mathcal{D}}_{k,j}(\mathcal{T}_{i,k}) + \widehat{\mathcal{D}}_{i,k} - \widehat{\mathcal{D}}_{i,j}. \end{aligned} \tag{5}$$

With a Lagrange multiplier λ_{sim} for the constraint in (3), this optimization problem is a linear least-squares problem of the form $\min |\vec{A}\vec{x} - b|^2$, where x is a vector containing all displacement and error variables, i.e. $\hat{\mathcal{T}}$ and $\hat{\mathcal{D}}$. A is a structured sparse matrix in which each row corresponds to one inner sum from either $\mathcal{C}, \mathcal{E}$ or $\mathcal{F}$. A then has the following non-zero entries:

- $1/Z_{\mathcal{C}}$ at positions corresponding to the constraints in (1);
- $\lambda_{\text{err}}/Z_{\mathcal{E}} \Psi_{i,j}(p)$ at positions to the corresponding the constraints in (2);
- $\lambda_{\text{def}}/Z_{\mathcal{F}}$ at positions corresponding to the constraints in the last term (4);

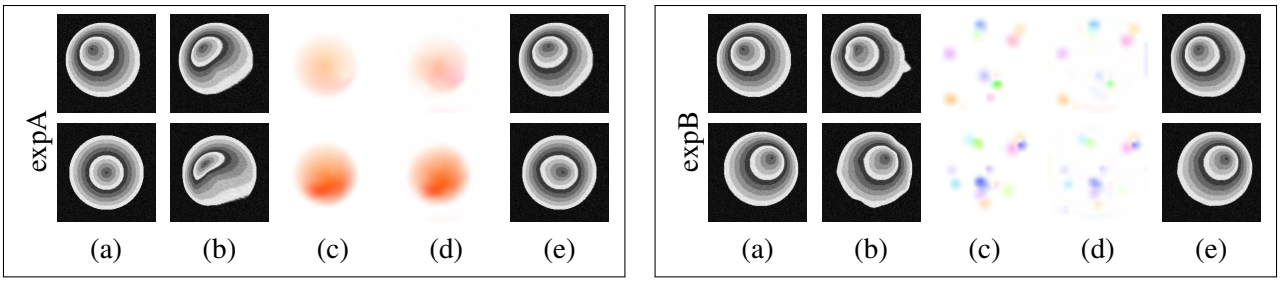


Figure 2: Two samples (each row) synthetic data with (left) large and (right) small errors. (a) target image, (b) erroneously deformed source image, (c) known ground-truth deformation error, (d) CLERC estimate of deformation error and (e) source image deformed with estimated true deformation. Hue corresponds to the orientation and saturation to the magnitude of errors.

b consists of zeros for all terms corresponding to (1) and (2), and of the observed displacements for the rows corresponding to the rows in (4). In this formulation, the D-dimensional components of the displacement and error vectors $\widehat{\Delta}(p)$ and $\widehat{\mathcal{D}}(p)$ are independent. We therefore find a least-squares solution for each of the D systems of equations in (3) using the reflective Newton trust-region method [20], which was chosen due to its computational and memory efficiency. As commonly utilized in medical image registration, we solve (3) on a coarse grid for efficiency and then interpolate the results. We use linear interpolation of the variables when $\widehat{\mathcal{D}}_{k,j}(\mathcal{T}_{i,k})$ is not on a grid point. In all our experiments, we observed $\vec{A}$ to be of full rank.

3 Experimental Results

We evaluate our method in three experiments, starting from a fully controlled environment using synthetic data, where we can have known ground-truth and demonstrate the utility of our method. We then present results on clinical datasets.

3.1 Simulated Deformations and Error

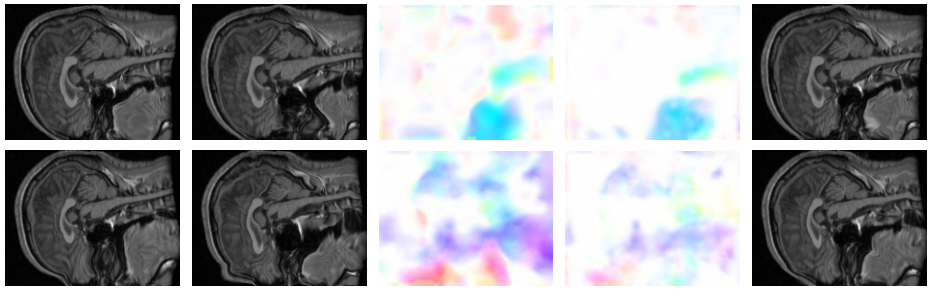
Nine 199×199 synthetic images with known pairwise registrations were first created by deforming a reference image. Simulated, random errors were then added to all registrations. We used two different parameterizations of the error-generation process, namely *few large errors* (expA) and *many small errors* (expB). The mean magnitude of displacement errors is 15 px for expA and 5 px for expB. The number of errors in each registration also follows a normal distribution, with the mean error frequency being 1 for expA and 20 for expB. Examples of typical error distributions can be seen in Fig. 2. We then ran CLERC for each setup independently, on a coarse grid of 25×25 control points. We present the average deformation error (ADE) defined as the mean of the magnitudes of local differences between estimated and known ground-truth deformations. We also report the inconsistency $\mathcal{C}$. Tab. 1 shows the improvements in ADE for both experiments, as well as a significant reduction in registration inconsistency in both cases.

3.2 Simulated Data, Real Registrations

For the second experiment, we simulated medical images with known correspondences. We used 19 mid-saggital slices of brain MRI with 481×374 px^2 resolution and 0.3 mm^2 spacing. We first registered one randomly chosen image to the remaining 18 images using a Markov-random field (MRF) based registration technique [21]. The computed registrations were then used to deform the source image. Subsequently, the 18 deformed images plus

Table 1: Average deformation error (ADE), target registration error (TRE) and inconsistency $\mathcal{C}_1$ in pixels (synthetic data) and mm (clinical data) for the experiments using simulated deformations before and after CLERC.

	synthetic data				2D MRI				3D CT			
	expA		expB		demons		MRF		demons		MRF	
	ADE	$\mathcal{C}$	ADE	$\mathcal{C}$	ADE	$\mathcal{C}$	ADE	$\mathcal{C}_1$	TRE	$\mathcal{C}$	TRE	$\mathcal{C}$
Post-registration error	3.33	5.34	0.67	1.47	3.50	2.53	1.92	1.85	9.62	8.14	5.25	5.67
Improved by CLERC	0.99	1.13	0.39	0.42	2.66	0.86	1.38	0.65	7.78	1.60	4.96	1.34



(a) (b) (c) (d) (e)

Figure 3: Registration errors of MRF (top) and demons (bottom) on the semi-simulated data. (a) target image, (b) deformed source, (c) true registration error, (d) estimated registration error, (e) source image deformed by the estimated true registration. Hue corresponds to orientation, and saturation to the magnitude of errors.

the source image were used as a new dataset. This dataset thus not only exhibits true anatomical variability, but also the ground-truth registrations are then known from the deformation process or can be analytically computed. We used diffeomorphic demons [22] and MRF-based registration [21] to mutually re-register all 19 images, resulting in 342 registrations each. Subsequently, we solved CLERC for each of these sets using a coarse grid with 8mm spacing. Quantitative results are given in Tab. 1. It is seen that ADE improves by over 25%, while the inconsistency is reduced considerably. We show one example case for each registration method in Fig. 3. We also compare our estimates with our own implementation of the AQUIRC method presented in [17]. Their method yields a unit- and direction-less measure of local error magnitude that can be correlated with the true local error magnitude. For the demons registration, the error magnitude predicted by ACUIRC has a correlation coefficient of 0.47, while ours yields 0.74. For the MRF registration, the correlation coefficients are 0.64 (AQUIRC) and 0.85 (CLERC).

3.3 Clinical Data

The last experiment was performed using a dataset of 15 clinical 3D CT scans of the head of different individuals, with $160\times160\times129\,\mathrm{px}^3$ resolution and unit voxel spacing. All images were mutually registered using demons and MRF-based registrations, and post-registration TRE was computed using 12 manually placed landmarks on the jawbone and the skull. In this inter-patient experiments, dense correspondences and thus consistency was not guaranteed. In fact, the presence and the number of teeth vary substantially among the elder population in this dataset. Similarly as above, we solved CLERC using a coarse grid with 8 mm spacing. The results are given in Tab. 1, in which a significant improvement in post-registration TRE using CLERC is observed.

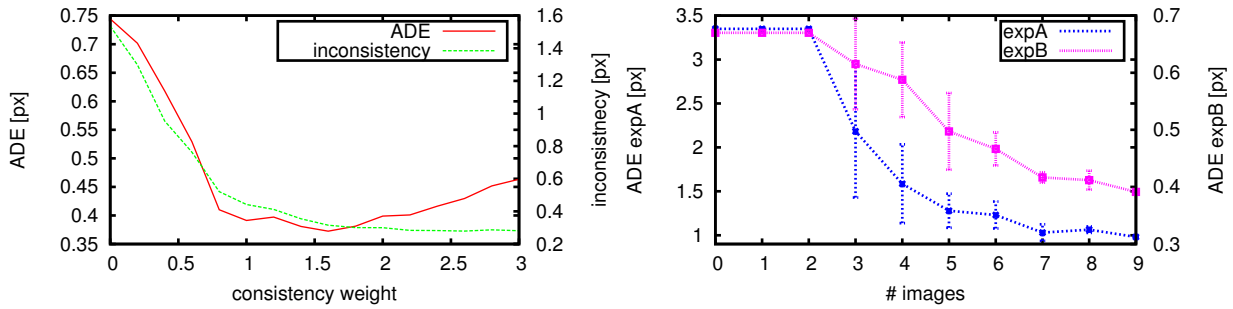


Figure 4: Influence of consistency weight on ADE and inconsistency achievable in CLERC in expB (left). The achievable improvement by using CLERC is seen to converge at increased number of images used for solving CLERC (right). Each data-point is from an average of 10 repeated experiments with random images with the variability shown by error bars.

3.4 Discussion

We have demonstrated experimentally that CLERC can reliably detect local errors in a fashion that also allows for correcting such errors. This discriminates it from methods such as AQUIRC [17], which use a simplified error generation model and are thus only able to solve for a dimension-less measure of error magnitude. In contrast, our method yields an estimate of the true magnitude and orientation of the error. CLERC is also robust with respect to the parameterization, which is illustrated by the fact that all results presented in this work were achieved using the same parameters $\lambda_{\text{err}} = \lambda_{\text{def}} = 20$. We only varied the window size for LNCC slightly to take into account the imaged anatomy/structure. Accordingly, σ was 1 px for the synthetic examples, and 1 mm for the clinical images, while γ was 8 for the 2D MR images and 6 otherwise. Note that further experiment-specific improvement is possible, in particular λ_{err} can be used to incorporate prior knowledge on the expected error. Overall, CLERC was shown to reduce the average absolute registration error by up to 72% for simple simulated errors and by up to 19% for the clinical data we presented.

Computation time and memory requirement of CLERC are determined by the linear system in (3). Despite being sparse, this linear system may still become quite large for fine grid resolutions or with increased number of images. Using the proposed solver, the linear system for each component of the 3D experiment can be solved relatively fast in approximately 5 minutes. Nonetheless, the memory consumption is the bottleneck of our algorithm. For the presented experiments, we limited the maximum required memory at 32 GB in order to avoid swapping to hard-drive, which otherwise slows down computation significantly. For a limited memory footprint, one can resort to coarser grid resolutions. Fig. 4(right) indicates a pattern of convergence in the achievable improvement using CLERC with increasing number of images. This is a promising observation that a small, affordable subset of images might indeed already yield accurate error estimations.

Note that the accuracy of the approximation in (5) deteriorates with increasing post-registration error as can be seen in the experiments with clinical data. A possible solution is to solve CLERC repeatedly, thereby improving the approximation as long as the overall error becomes smaller. Encouraged by our preliminary results following this approach, further investigations will be carried out in this direction in the future.

4 Conclusions

In this paper, we have presented a novel method CLERC to predict spatial location, magnitude and orientation of inconsistency-based errors in *any* non-rigid registration. This is achieved by using an error term that jointly estimates consistent registrations and local registration errors over a set of images registered pairwise. To the

best of our knowledge, CLERC is the first method that detects inconsistency-based errors as *vectors* such that these can also be corrected subsequently.

References

[1] Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, C.: Quo vadis, atlas-based segmentation? Handbook of Biomedical Image Analysis (2005) 435–486

[2] Fischer, B., Modersitzki, J.: Ill-posed medicine - an introduction to image registration. Inverse Problems **24**(3) (2008) 1–19

[3] Rueckert, D., Aljabar, P., Heckemann, R.a., Hajnal, J.V., Hammers, A.: Diffeomorphic registration using B-splines. In: MICCAI. (2006) 702–709

[4] Geng, X.: Transitive inverse-consistent image registration and evaluation. PhD thesis, University of Iowa (2007)

[5] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic Registration: a demons-based approach. In: MICCAI. Volume 11. (2008) 754–61

[6] Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage **23 Suppl 1** (2004) S151–60

[7] Skrinjar, O., Bistoquet, A., Tagare, H.: Symmetric and transitive registration of image sequences. Int J Biomed Imaging (2008)

[8] Crum, W., Griffin, L., Hill, D., Hawkes, D.: Zen and the art of medical image registration: correspondence, homology, and quality. NeuroImage **20**(3) (2003) 1425–37

[9] Rohlfing, T.: Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. IEEE Trans Med Imag **31**(2) (2012) 153–63

[10] Robinson, D., Milanfar, P.: Fundamental performance limits in image registration. IEEE Trans Imag Proc **13**(9) (2004) 1185–1199

[11] Kybic, J.: Bootstrap resampling for image registration uncertainty estimation without ground truth. IEEE Trans Imag Proc **19**(1) (2010) 64–73

[12] Crum, W., Griffin, L., Hawkes, D.: Automatic estimation of error in voxel-based registration. In: MICCAI. (2004) 821–28

[13] Schestowitz, R.S., Crum, W.R., Petrovic, V.S., Twining, C.J., Cootes, T.F., Taylor, C.J.: Non-rigid registration assessment without ground truth. In: ISBI. (2006) 836–839

[14] Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.C.: Automated image registration: I. General methods and intrasubject, intramodality validation. Journal of Computer Assisted Tomography **22**(1) (1998) 139–52

[15] Holden, M., Hill, D.L., Denton, E.R., Jarosz, J.M., Cox, T.C., Rohlfing, T., Goodey, J., Hawkes, D.J.: Voxel similarity measures for 3-D serial MR brain image registration. IEEE Trans Med Imag **19**(2) (2000) 94–102

[16] Christensen, G.E., Johnson, H.J.: Consistent image registration. IEEE Trans Med Imag **20**(7) (2001) 568–582

[17] Datteri, R., Dawant, B.: Automatic detection of the magnitude and spatial location of error in non-rigid registration. In: Biomedical Image Registration. (2012) 21–30

[18] Christensen, G.E.: Invertibility and transitivity analysis for nonrigid image registration. Journal of Electronic Imaging **12**(1) (2003) 106

[19] Cachier, P., Bardinet, E., Dormont, D., Pennec, X., Ayache, N.: Iconic feature based nonrigid registration: the PASHA algorithm. Comput Vis Image Und **89**(2-3) (2003) 272–298

[20] Coleman, T., Li, Y.: A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. SIAM Journal on Optimization **6**(4) (1996) 1040–58

[21] Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med Image Anal **12**(6) (2008) 731–741

[22] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage **45**(1 Suppl) (2009) 61–72