Registration of Developmental Image Sequences with Missing Data

Istvan Csapo[1] Yundi Shi[1] Mar Sanchez[2] Martin Styner[1] Marc Niethammer[1]
[1]UNC Chapel Hill [2]Emory University

Abstract

Longitudinal image registration is commonly used to establish spatial correspondence between images when investigating temporal changes in brain morphology. Most image registration methods have been developed to align images that are similar in appearance or structure. If such similarity is not given (e.g., in the case of neurodevelopmental studies, which is the target application of this paper), (i) local similarity measures, (ii) metamorphosis approaches, or (iii) methods modeling longitudinal intensity change can be used. Methods modeling longitudinal intensity change have the advantage of not treating images as independent static samples. However, missing or incomplete data can lead to poor model estimation and, in turn, poor registration. Therefore, incomplete longitudinal data sets are often excluded from analysis. Here, we propose a method to build a longitudinal atlas of intensity change and incorporate it as a prior into an existing model-based registration method. We show that using the prior can guide the deformable registration of longitudinal images of brain development with missing data and produce comparable registration results to complete data sets.

1. Introduction

Studying brain development is of great interest to investigate the normal growth process or the influence of various diseases on neurodevelopment. During the early stages of development the brain undergoes great changes in size and structure. Overall brain size nearly doubles during the first year of neonatal development and reaches 80-90% of its adult volume by age 2 [6]. This rapid increase in volume does not affect all brain structures equally. Gray matter, the cerebellum, and the lateral ventricles contribute the majority of the brain volume increase during the first year of development, while white matter volume increases only modestly [6]. In addition to morphological changes, the myelination of white matter proceeds rapidly during this same period leading to non-uniform changes in the magnetic resonance (MR) appearance of white matter throughout the brain [8].

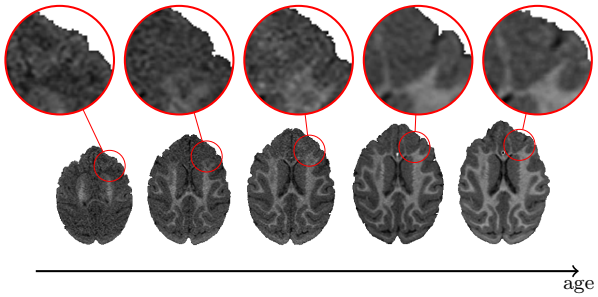

Figure 1. Axial slices of MR brain images of monkey at ages 2 weeks, 3, 6, 12, and 18 months. White matter appearance changes non-uniformly as axons are myelinated during brain development.

The combined morphological and appearance changes (see Fig. 1) make longitudinal analysis of brain development a challenging problem. Longitudinal analysis has to be flexible enough to allow for the large non-uniform morphological and appearance changes while at the same time provide temporally smooth solutions.

As the brain changes shape during its development, spatial correspondences need to be estimated over time. This is typically done by image registration methods. To capture localized deformations (as occur in neurodevelopment) necessitates deformable image registration methods. However, deformable image registration is a highly ambiguous problem as many transformations may result in visually good image alignment. Hence, regularity conditions are imposed to favor spatially smooth transformations. For longitudinal analysis methods, temporal regularization can be used to single out plausible solutions [3, 5, 13], but various brain structures develop at different rates and temporal regularization can introduce bias by restricting the temporal evolution of the solution.

However, in conjunction with a suitable image similarity measure successful deformable image registration can be achieved in many instances. As image appearance changes (in structural MR images; see Fig. 2) during neurodevelopment standard similarity measures (such as mutual information or cross correlation) are no longer appropriate and may lead to incorrect estimates of deformation that regularization alone cannot fully prevent. A way to improve alignment is to introduce meaningful constraints into the registration. In [2] white matter appearance over time is esti-

mated locally and then used to create a predicted target image for registration at a given time-point. By restricting intensities to conform to a parametric model an improved registration performance can be obtained. While an improvement over the standard registration model, the model in [2] is fully data-driven and parameter-estimation may therefore become unreliable if the number of images in the longitudinal data set is low. To improve performance we therefore propose to learn an atlas of expected models of intensity change and to use it as a prior to better guide the deformable registration for longitudinal data sets with missing time-points. In addition, the prior can also aid the model-based registration of complete data sets when the model cannot be reliably estimated due to corrupted data or poor initial alignment.

Furthermore, atlases are frequently used to aid various image processing methods, such as registration [9] and segmentation [4]. Therefore, the maturation atlas in itself could be useful for other methods.

Contributions.

- A *method to integrate the longitudinal intensity prior into a deformable registration approach* to aid the registration for data sets with missing time-points.
- An *atlas-building method and the resulting atlas* describing white matter maturation and its expected variation based on longitudinal data sets.
- A *synthetic test case* showing improved alignment for the proposed registration approach over a deformable registration method that includes a longitudinal intensity model without any priors [2].
- An *evaluation on a real data set* showing that the proposed method can achieve comparable registration results to registration with the full longitudinal image sequence even with missing time-points. The results are validated with manually selected landmarks.

Our method aims to recover biologically plausible changes in brain maturation. Sec. 2 integrates the atlas as a prior into a registration method. Sec. 3 describes the approach to build the atlas of brain maturation. Experimental results are shown in Sec. 4. The paper concludes with a discussion and outlook in Sec. 5.

2. Using the Maturation Information

Given a longitudinal set of images that should be registered we have (following [2]) the equivalent of the sum of squared difference intensity difference with respect to a fit logistic model over all images

$$SSR(\{I_i\}) = \sum_{i=0}^{n-1} \int_\Omega (I(\Phi_i(x), t_i) - \hat{I}(x, t_i))^2 \, dx, \quad (1)$$

where Ω is the image domain of the fixed image, $\hat{I}$ is the intensity-adjusted spatially fixed target image, and Φ_i is the map from the moving image I to the target image $\hat{I}$. As we want to make use of the prior information we augment this similarity measure to

$$SSRP(\{I_i\}) = \sum_{i=0}^{n-1} \int_\Omega (I(\Phi_i(x), t_i) - \hat{I}(x, t_i))^2 \, dx + \gamma \int_\Omega \int_t \frac{1}{\sigma^2(x,t)} (I_p(x,t) - \hat{I}(x,t))^2 \, dt \, dx, \quad (2)$$

where I_p is the intensity predicted by the prior model, σ^2 is the variance of the image intensity values with respect to the model intensity (see Sec. 3 for more details on the model estimation), and $\gamma > 0$ a chosen balancing constant. Hence, given a current estimate of the deformation we can compute an estimate for the model $\hat{I}$ locally by minimizing Eq. 2 with respect to the parameters $\hat{\theta}$ of $\hat{I}$. The remainder of the procedure stays unchanged.

3. Estimating the Prior from a Population

To build the atlas we make the following choices and assumptions:

1. **Locality of the model.** We define the model over a white matter mask only as the expected change in gray matter intensity throughout neurodevelopment is negligible compared to the white matter.

2. **Atlas space.** As white/gray-matter contrast is best for a fully matured brain, the atlas is defined on an average atlas image for the latest time-point in our data set – where an automatic white matter segmentation can be reliably obtained. The atlas model parameters are resampled in the anatomical space of each subject through registering the atlas to the latest time-point.

3. **Temporal model type.** Ideally, we would learn the model fully from the data. However, given the limited number of subjects typical for most neurodevelopment studies we use a logistic model (as in [2]) with fixed asymptotes (for minimal and maximal white matter intensity). This way the temporal model is parameterized by only two parameters capturing onset and rate of maturation.

4. **Spatial regularization.** In order to provide robust estimates of the model parameters with respect to image noise and small registration errors we enforced spatial regularity on the estimated group-wise model parameters via a $3 \times 3 \times 3$ median filter.

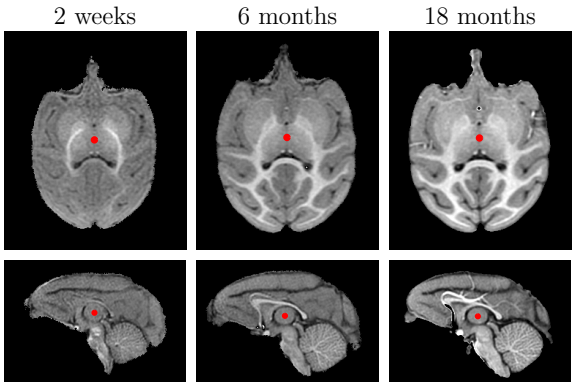

Figure 2. Example of an expert-defined anatomical landmark within the same subject (*top:* axial slices; *bottom:* sagittal slices).

Data set. We make use of synthetic and real data sets for evaluation and atlas-building. Our real data set consists of longitudinal MR images with five time-points (scanned at 2 weeks, 3, 6, 12, and 18 months) for 10 normal rhesus macaques acquired at the Yerkes Imaging Center (Emory University, GA). The subjects were scanned on a 3T Siemens Trio scanner at the Yerkes Imaging Center, Emory University with a high-resolution T1-weighted 3D magnetization prepared rapid gradient echo (MPRAGE) sequence (TR = 3,000ms, TE = 3.33ms, flip angle = 8 , matrix = 192 $\times$ 192, voxel size = 0.6mm^3, some images were acquired with TE = 3.51ms, voxel size = 0.5mm^3).

The images were inhomogeneity corrected [12] and registered into a standardized atlas space using an affine transformation with a local cross correlation metric (ANTS [1]). Skull stripping and tissue segmentation were obtained with an atlas-based segmentation method [10]. Furthermore, 35 anatomical landmarks were defined by an expert for all 5 time-points (see Fig. 2).

To build our maturation atlas we first perform intra-subject elastic deformable registration to the latest time-point using the previously developed longitudinal SSR registration [2]. Inter-subject registration is then established by registering the oldest time-point of each subject to a common atlas space. This establishes correspondences between all subjects for all time-points. We thin the white matter mask to avoid model estimation errors for the logistic intensity model near tissue boundaries due to partial voluming, segmentation errors, or registration errors. For each point in the thinned white matter mask the logistic models are then estimated as follows.

Model estimation. We point-wise estimate logistic curves. We considered (i) a direct fit of the logistic model to all data points (disregarding their longitudinal nature), (ii) a full longitudinal estimation procedure, and (iii) fitting individual logistic curves for each subject followed by the computation of the median of these logistic curves. Since different subjects of the same age may be at slightly different stages of neurodevelopment the first two approaches would require a simultaneous estimation of this time-shift to avoid underestimating the maturation rate. We therefore chose the latter approach as it is robust to such time-shifts, is simple, and provides a parametric model. We locally compute the median curve over the population using the approach proposed for calculating the functional boxplot [11], where we replace all the measured values by the values of their individual logistic model fits. As the median curve will be a curve from the data set we thus obtain a median logistic curve. (This is fundamentally different from computing a point-wise median.) The individual logistic models are fit using ordinary least squares[1] [$\hat{\theta_j} = \underset{\theta}{\operatorname{argmin}} \sum_i (I_j(t_i) - logistic(t_i, \theta))^2$], where $I_j(t_i)$ is the intensity at time-point t_i for subject j and θ denotes the coefficients for the logistic model. The logistic model is defined as

$$logistic(\mathbf{x}, t; \alpha, \beta(\mathbf{x}), k(\mathbf{x})) = \frac{\alpha}{1 + e^{-k(\mathbf{x})(t-\beta(\mathbf{x}))}} \ , \quad (3)$$

where α is a global parameter (intensity change between unmyelinated and myelinated white matter), β (onset: time of maximum rate of intensity change) and k (maximum rate of intensity change) are spatially varying model parameters (see Fig. 3). We further compute the variance of the actual measured values with respect to the median logistic model ($\bar{\theta}$) at our ages of interest [$\hat{\sigma}^2(t_i) = \frac{1}{N(t_i)} \sum_j (I_j(t_i) - logistic(t_i, \bar{\theta}))^2$], where $N(t_i)$ denotes the number of measurements at time-point t_i. This allows us to introduce local weights into the registration method (see Sec. 2).

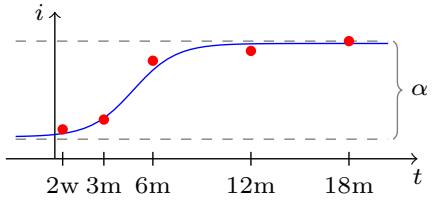

Figure 3. Logistic intensity model. The model parameters are estimated for five time-points.

As maturation is expected to be locally smooth we smoothly extend the estimated parameters of the logistic model to the full white matter mask. Fig. 4 shows sample slices with the color-coded coefficients for onset and maturation rate. The maturation atlas captures the expected maturation patterns [8] with areas deeper in the brain either reaching their matured intensity early or already being fully matured at two weeks of age (first time-point).

[1] This could be easily replaced by a weighted least squares model approximating for example a Rician noise model [7].

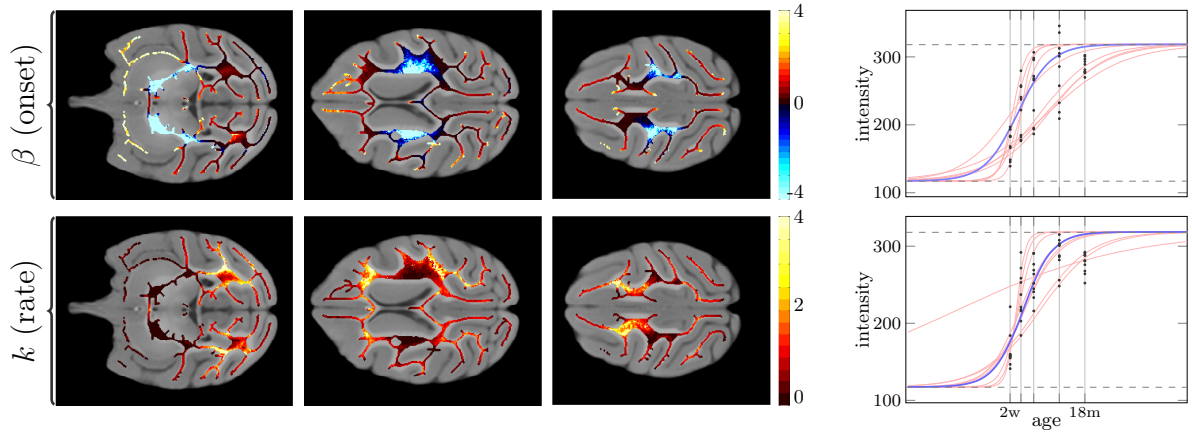

Figure 4. Estimated onset (in months) and rate of maturation overlaid on three axial slices of the structural atlas (*columns 1-3*: inferior to superior). Note that this is the initial estimate in the thinned white matter mask. The parameters follow an expected pattern of maturation. The plots show the individual (*red*) and median (*blue*) logistic curves and the intensity values (*black dots*) for the ten subjects for two different voxels.

Shortcomings of the model estimation. Time-shift uncertainties are only implicitly accounted for in our model as the computed variances will be a combination between image noise, registration errors, and time-shifts of the maturation trajectories. An improved estimation procedure could also estimate the time-shifts, however at the cost of greater model complexity. Our observations indicate that these shifts (if present) are relatively minor and therefore chose not to explicitly model them. However, the effect of time-shifts should be investigated as part of future work.

4. Experimental Results

To assess the utility of incorporating the white matter intensity prior into the longitudinal similarity measure SSR we i) validated the proposed method on a synthetic 2D data set, ii) for the real 3D monkey data, showed that affine registration alone is not sufficient to align developmental image sequences, and iii) tested the utility of the prior for incomplete real data sets.

4.1. Registering synthetic 2D data set

To test the influence of prior brain maturation information when used within the registration we used a synthetic test case shown in Fig. 5 and compared the model-based similarity measure SSR without prior to the new measure SSRP. The five synthetic images $I_1, \ldots, I_5$ represent simplified MR images of the developing brain at times $t_i = 0, 0.5, 0.6, 0.7, 1.0$ with logistic white matter intensity change over time (the onset and rate of intensity change spatially varies within the white matter) and a Rician noise model. A small longitudinal deformation was added to the moving images $I_1, \ldots, I_4$ to simulate a non-linear growth process (described in Sec. 4.2.1). The red outline shows the tissue gray/white matter boundary of the target image. The small misalignment of the tissue boundaries is representative of the misalignment seen in the real images after an initial affine registration. The intensity model estimation often fails to recover a reasonable model in these regions due to the mixed gray and white matter voxels in the same spatial location at the various time-points. The goal of the experiment is to recover the true deformation Φ_i^{-1} for each time-point $t_1, \ldots, t_4$.

Table 1. Registration error for synthetic data experiment with varying prior weights and number of time-points used for model estimation. The error is the average pixel-wise difference between the known true deformation and the recovered deformation within the mask of the target image (in pixels). The minimum for each row is highlighted in bold.

	prior weight (γ)				
time-points	0	0.01	0.1	0.5	1
1 5	1.78	0.96	0.61	0.24	**0.09**
1 2 5	0.92	0.69	0.39	0.27	**0.22**
1 2 3 5	0.49	0.47	0.38	0.31	**0.28**
1 2 3 4 5	0.25	0.25	0.25	0.24	**0.24**

In this experiment the true intensity model is known for the synthetic images and it is used as the prior. The registration experiments were run with varying the prior weight parameter $\gamma = 0, 0.01, 0.1, 0.5, 1$ (with $\gamma = 0$ SSRP reduces to the standard SSR registration; the results converge at $\gamma = 1$) and the number of time-points available for the model estimation from 2 up to all 5 (total of 20 experiments). Fig. 6 shows the registration results for two prior weights ($\gamma = 0, 1$) and two different number of time-points for the model estimation. Using only two time-points for the model estimation step without the prior results in a poor

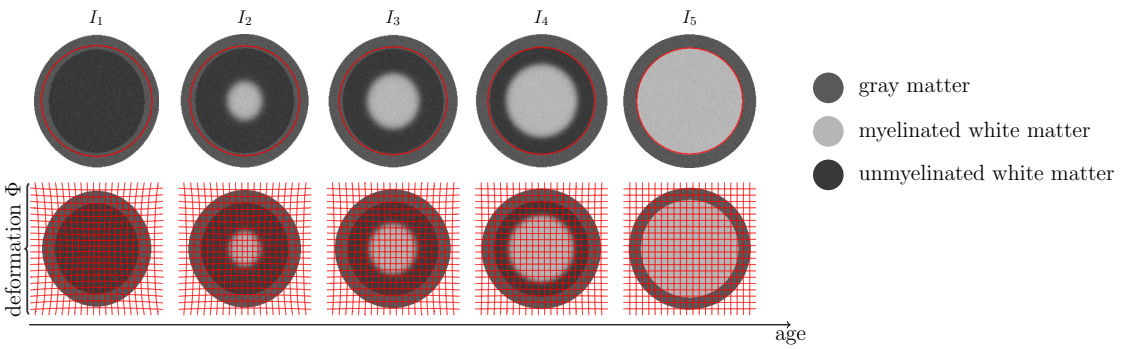

Figure 5. Synthetic data set with logistic white matter intensity change and longitudinal deformation over time. $I_1, \ldots, I_4$ are source images, I_5 is the target image. The red outline shows the white matter gray matter boundary of the target image (*top row*).

model and subsequently in poor registration. Adding the prior with a lot of weight results in a greatly improved model and registration. Note that when all the time-points are available for the model estimation the prior has less of an influence since the model can be estimated well from the data alone. Table 1 shows the registration errors for all the 20 experiments. The registration errors are computed as the average pixel-wise difference between the known true deformation and the recovered deformation within the mask of the target image. The prior considerably improves the registration results when the model is estimated from only a few time-points, but has less of an effect when all time-points are available. However, using the prior improved registration even with the complete data set. This is due to the improved model estimation near tissue boundaries where initial misregistration can result in mixing gray and white matter voxels and therefore lead to poor model estimation. Note that since the true prior is known for this experiment, the best results is obtained with largest overall weight on the prior (with two time-points the data is weighted less than with all the time-points).

4.2. Registering real 3D data set

The real monkey data set is described in Sec. 3 and shown in Fig. 1 for a single subject. We assessed the quality of the registration results using manually placed landmarks.

4.2.1 Accuracy of affine registration

While the non-uniform growth of various brain regions [6] indicates that affine registration should not be sufficient to align the longitudinal images and therefore a deformable registration method is necessary, our experience suggests that affine registration alone can produce a good alignment even between I_{2wk} and I_{18mo}. To test the within-subject appropriateness of a simple affine image-alignment, for each subject we first affinely registered the longitudinal images ($I_{2wk}, \ldots, I_{12mo}$ to I_{18mo}) using a normalized cross correlation similarity measure which has previously been found effective [2] even in the presence of image intensity changes as expected during brain maturation. Then we performed an additional deformable registration of the affinely aligned images using SSR [2].

The registration error was calculated as the distance between the transformed and the target landmarks. Significance was calculated with a paired-sample t-test at a significance level of $\alpha = 0.05$. The boxplots in Fig. 9 show the landmark alignment errors per time-point for both affine (a) and deformable (d) registration methods using all landmarks (left plot; 350 landmark pairs per time-point) and using only a subset of six landmarks (right plot; 60 landmark pairs per time-point).

What is clear from both plots is that the registration error for both affine and deformable registration is inversely proportional to age, but there is no clear benefit of using deformable registration when aggregating the results over all the landmarks (the mean registration error for $I_{2wk}, \ldots, I_{12mo}$ are 1.95, 1.57, 1.04 and 0.65 mm for affine and 1.82, 1.44, 0.98, 0.74 mm for deformable registration, respectively; the means of the two methods are significantly different for each age). This would suggest that the poor affine alignment of I_{2wk} is not the result of an insufficient transformation model, but possibly increasingly inaccurate landmark placement due to poor tissue contrast at earlier time-points (see Fig. 2).

However, taking a closer look at individual landmarks revealed that only a small subset was not well aligned after affine registration. Fig. 7 shows four landmarks after affine and deformable registration. Landmarks 28 and 4 were well aligned after affine registration and thus did not change after deformable registration; however, landmarks 7 and 32 moved closer to the target 18 month locations. Out of the 35 landmarks we chose a subset of 6 that have significantly changed after the deformable registration (improvement was not a criteria when choosing the subset). The high variability across subjects of certain landmarks may be attributed to the inaccuracy of the landmarks at the earlier time-points due to poor tissue contrast or anatomical variations in those brain regions.

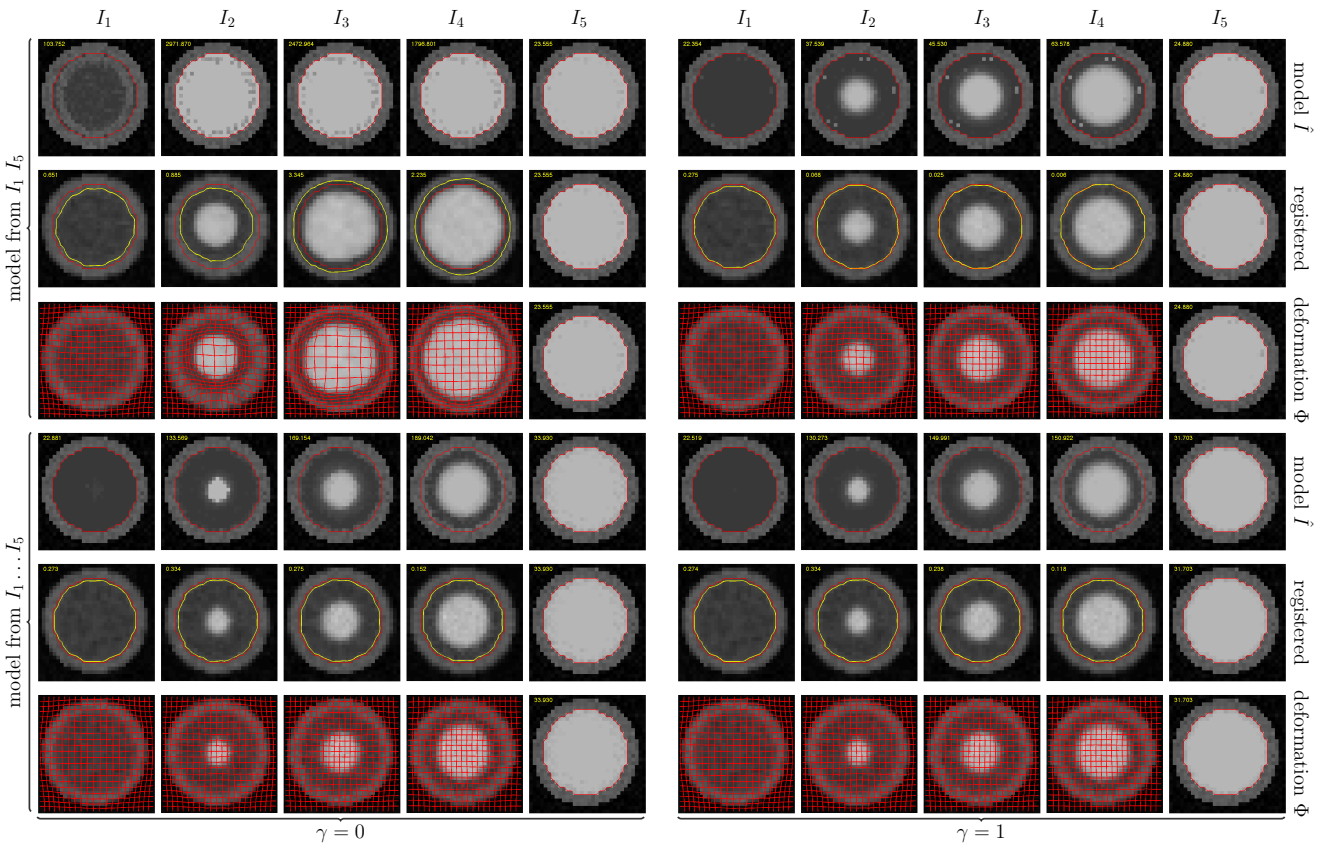

Figure 6. Registration results with no prior (left set of images) and use of the prior with weight 1 (right set of images) and the number of time-points used for the model estimation. For the two different number of time-points: *top row:* estimated model; *middle row:* the target tissue boundary (*red*) and the registered boundary (*yellow*) are overlaid on the registered images and should be close for good registration; *bottom row:* the deformation field field is overlaid on the registered images.

Once we restricted our analysis to a subset of the 6 landmarks previously described the benefit of the deformable registration became apparent (see right plot in Fig. 9; the mean registration error for $I_{2wk}, \ldots, I_{12mo}$ are 2.44, 2.03, 1.24, 0.59 mm for affine and 1.97, 1.24, 0.86, 0.72 mm for deformable registration, respectively; the means of the two methods are significantly different for each age). In fact, these results are consistent with the longitudinal deformations seen in the images. Fig. 8 shows the typical deformation of the brain seen in the data sets: an upward bending of the brain about the brain stem along the anterior-posterior line. This finding is consistent with the known developmental growth trajectories of the centrally located lateral ventricles and the cerebellum that contribute the majority of the subcortical brain volume increase during early postnatal development [6]. The landmarks that have changed significantly are located in the regions that the bending deformation affected the most (superior-anterior and inferior-posterior regions).

4.2.2 Influence of prior with missing data

In this experiment we tested the utility of the proposed method that incorporates a prior into the model-based deformable registration method.

We registered a subset of the four early time-points $I_{2wk}, \ldots, I_{12mo}$ to the target image I_{18mo} with the proposed model-based method for each of the 10 subjects. The subsets either contained $\{I_{2wk}\}$, $\{I_{2wk}, I_{3mo}\}$, $\{I_{2wk}, I_{3mo}, I_{6mo}\}$, or all four, $\{I_{2wk}, \ldots, I_{12mo}\}$, moving images (the last subset corresponding to no missing data). The missing time-points were not used for the model estimation. For each of the four subsets, the registration experiments were repeated with five different prior weights $\gamma = 0, 0.01, 0.1, 0.5, 1$ (with $\gamma = 0$ SSRP reduces to the standard SSR registration) totaling 20 experiments (as in Sec. 4.1).

As in Sec. 4.2.1, the registration error was computed as the distance between registered and the target landmarks. We used the most challenging registration between I_{2wk} and the I_{18mo} to assess the quality of the registrations (note that some of the immediate time-points were used for the

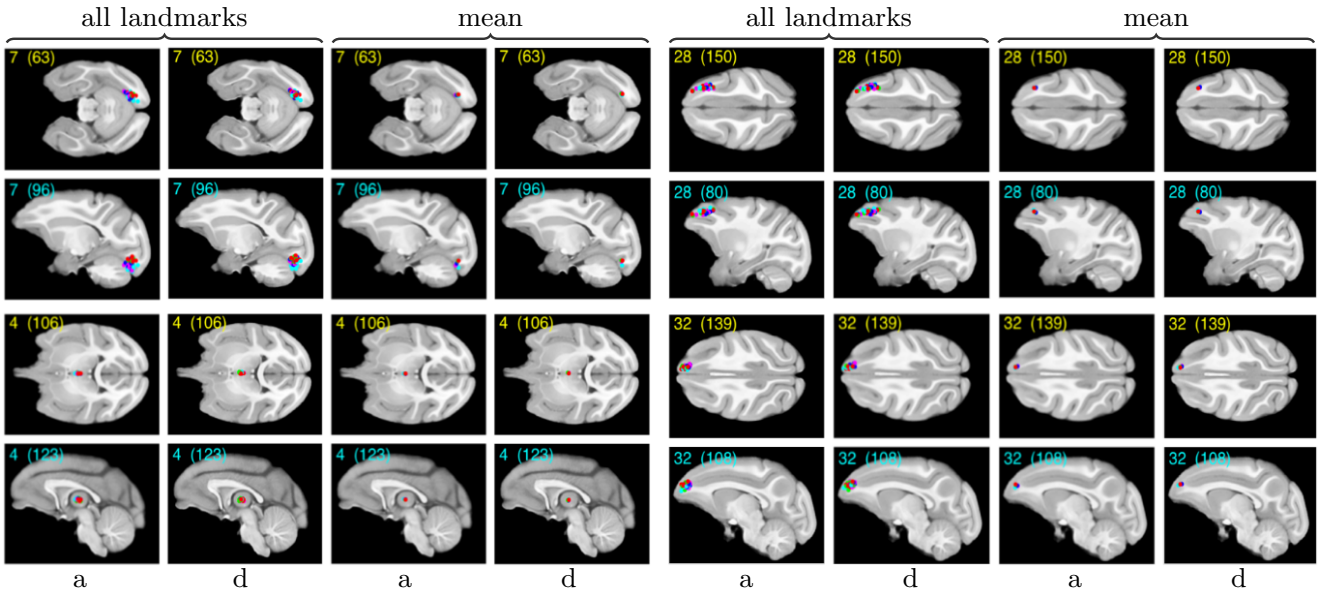

Figure 7. A subset of landmarks (7, 28, 4, 32) are shown on the structural atlas. For each landmark, we show the location on the axial (*top row*) and the sagittal (*bottom row*) plane. The landmark locations are color coded by age (2wk: cyan; 3mo: magenta; 6mo: green; 12mo: blue; 18mo: red). For each landmark, the first two columns show all 50 locations (10 subjects $\times$ 5 time-points) for affine (a) and deformable (d) registrations. Columns 3 and 4 show the mean location for each age for affine (a) and deformable (d) registrations. Note that the I_{18mo} (target) images were affinely registered to the atlas before the within subject affine and deformable registrations, but variations in anatomy and landmark placement across subjects can be seen as the spread in landmark locations.

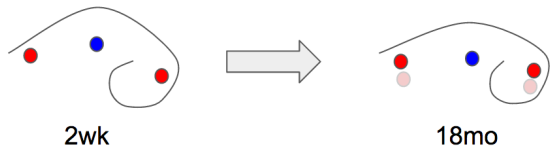

Figure 8. Cartoon depiction of the typical anterior-posterior deformation seen between the 2 week and the 18 month images (in the sagittal plane). The deformation is a result of the non-uniform growth trajectories of various brain regions. Landmarks located near the anterior and posterior regions (*red*) are affected more by the deformation than centrally located landmarks (*blue*).

model estimation for three of the subsets). Similarly to the previous experiment, we chose a new subset of 6 landmarks (different from the one used in Sec. 4.2.1) in the central region of the brain where using a prior had noticeable impact. The following results were computed using the subset of landmarks.

Table 2 shows the registration errors for all 20 experiments. For each moving image subset with a non-zero prior weight (columns 2-5) the registration errors were compared to the zero prior weight errors (column 1). The prior significantly improved the registration results by 0.21 mm for the data sets with three missing time-points (top row) at prior weight 1 (highlighted in bold). Additionally, there were significant improvements with the prior for two and no missing data sets. However, all improvements were at a sub pixel level.

Table 2. Registration error (in mm) for monkey data experiment with varying prior weights and number of time-points used for model estimation. The registration errors are computed for the I_{2wk} to I_{18mo} registrations for all 4 subsets of moving images and 5 prior weights. For each row, statistically significant improvements with respect to the first column ($\gamma = 0$) are highlighted in bold.

	prior weight (γ)				
time-points	0	0.01	0.1	0.5	1
1 5	2.08	2.11	2.07	2.02	**1.87**
1 2 5	1.96	1.91	**1.91**	**1.84**	**1.86**
1 2 3 5	1.88	1.89	1.91	1.86	1.86
1 2 3 4 5	1.90	1.93	**1.84**	1.88	1.89

The synthetic experiments show a clear benefit when using the prior with missing data, but the improvement for the real data is difficult to show globally. While the prior has a clear impact on the estimated model images ($\hat{I}$), the benefit of using a prior only becomes evident when restricting the landmarks to the central region of the brain where the model estimation is most influenced by the prior. Note that while these improvements are only measured at sparsely placed landmarks, when trying to measure subtle volumetric changes during development the small improvements are integrated over large structures.

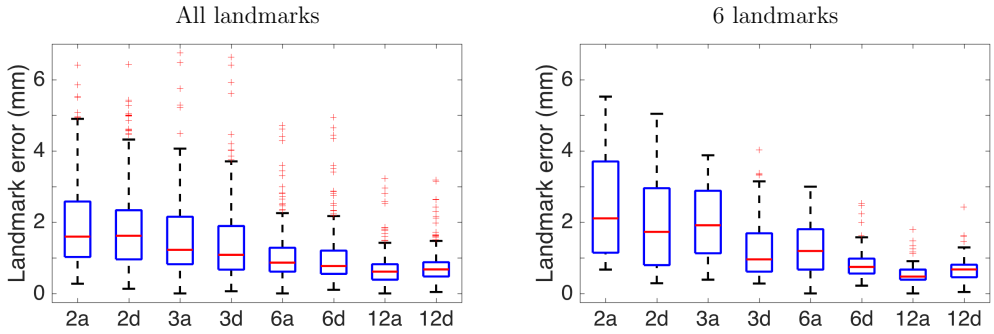

Figure 9. Landmark errors per time-point for all 10 subjects after registering the earlier time-points $I_{2wk}, \ldots, I_{12mo}$ to the oldest I_{18mo} image within subjects using affine (2a,...,12a) and deformable (2d,...,12d) registration. While the deformable registration slightly improves alignment for $I_{2wk}, \ldots, I_{6mo}$ (and worsens it for I_{12mo}), the changes are negligible when aggregating over all landmarks (*left plot*), but are more prominent when only considering a subset of the landmarks (*right plot*).

A more in-depth analysis of the subtle improvements should be part of future work, but we list a few possible explanations of why the benefit of using a prior is not as evident for the real data as was for the synthetic data set: (i) the landmarks are placed in easily identifiable regions with prominent structures, therefore these regions are easily registered even without a prior; (ii) the landmarks are inaccurate in the earlier time-points due to the poor tissue contrast and noisy images (see Fig. 2); (iii) the morphological changes, other than affine, between the 2 week and the 18 month images are too subtle to be measured by sparse landmarks and only affected a small subset of the 35 landmarks.

5. Conclusions

We proposed to incorporate a longitudinal intensity atlas into a model-based longitudinal similarity measure. The atlas is used as a prior and can guide the registration when the model cannot be reliably estimated due to missing data. While the similarity measure uses a logistic model, the formulation is general and other models can be used. We showed that using a prior can improve registration with missing data. In the future, we will investigate the benefit of using a prior with data sets containing larger non-linear morphological changes.

References

[1] B. B. Avants, N. J. Tustison, G. Song, and J. C. Gee. Ants: Open-source tools for normalization and neuroanatomy. 3

[2] I. Csapo, B. Davis, Y. Shi, M. Sanchez, M. Styner, and M. Niethammer. Longitudinal image registration with temporally-dependent image similarity measure. *IEEE Trans. Med. Imaging*, 32(10):1939–1951, 2013. 1, 2, 3, 5

[3] S. Durrleman, X. Pennec, A. Trouv, G. Gerig, and N. Ayache. Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. *Med Image Comput Comput Assist Interv*, 12(Pt 1):297–304, 2009. 1

[4] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, and et al. Whole brain segmentation. *Neuron*, 33(3):341–355, 2002. 2

[5] J. Fishbaugh, S. Durrleman, and G. Gerig. Estimation of smooth growth trajectories with controlled acceleration from time series shape data. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pages 401–408, 2011. 1

[6] R. C. Knickmeyer, S. Gouttard, C. Kang, D. Evans, K. Wilber, J. K. Smith, R. M. Hamer, W. Lin, G. Gerig, and J. H. Gilmore. A structural mri study of human brain development from birth to 2 years. *J Neurosci*, 28(47):12176–12182, 2008. 1, 5, 6

[7] R. Salvador, A. Pena, D. K. Menon, T. A. Carpenter, J. D. Pickard, and E. T. Bullmore. Formal characterization and extension of the linearized diffusion tensor model. *Human brain mapping*, 24(2):144–155, 2004. 3

[8] R. C. Sampaio and C. L. Truwit. *Myelination in the developing brain. In: Handbook of developmental cognitive neuroscience*, pages 35–44. MIT Press, 2001. 1, 3

[9] A. Serag, P. Aljabar, S. Counsell, J. Boardman, J. V. Hajnal, and D. Rueckert. Lisa: Longitudinal image registration via spatio-temporal atlases. *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012. 2

[10] M. Styner, R. Knickmeyer, C. Coe, S. J. Short, and J. Gilmore. Automatic regional analysis of DTI properties in the developmental macaque brain. *Medical Imaging 2008: Image Processing*, 2008. 3

[11] Y. Sun and M. G. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, 2011. 3

[12] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging*, 29(6):1310–1320, 2010. 3

[13] G. Wu, Q. Wang, and D. Shen. Registration of longitudinal brain image sequences with implicit template and spatial-temporal heuristics. *NeuroImage*, 59(1):404–421, 2012. 1