

Temporally-Dependent Image Similarity Measure for Longitudinal Analysis

Istvan Csapo¹, Brad Davis², Yundi Shi¹, Mar Sanchez³, Martin Styner¹, and Marc Niethammer¹

¹ University of North Carolina at Chapel Hill, NC

² Kitware, Inc., Carrboro, NC

³ Emory University, Atlanta, GA
icsapo@cs.unc.edu

Abstract. Current longitudinal image registration methods rely on the assumption that image appearance between time-points remains constant or changes uniformly within intensity classes. This assumption, however, is not valid for magnetic resonance imaging of brain development. Registration methods developed to align images with non-uniform appearance change either (i) locally minimize some global similarity measure, or (ii) iteratively estimate an intensity transformation that makes the images similar. However, these methods treat the individual images as independent static samples and are inadequate for the strong non-uniform appearance changes seen in neurodevelopmental data. Here, we propose a *model-based similarity measure* intended for aligning longitudinal images that locally estimates a temporal model of intensity change. Unlike previous approaches, the model-based formulation is able to capture complex appearance changes between time-points and we demonstrate that it is critical when using a deformable transformation model.

1 Introduction

The analysis of longitudinal images is important in the study of neurodevelopment and its disorders. If global measures are insufficient for analysis, change can be localized by establishing image correspondence via registration. The aim of the registration method is then to find a reasonable geometric transformation between the images according to some similarity measure and a model of spatial transformation. Although longitudinal registration has received some attention in recent years, most of the effort has been focused on the spatial extent of change (various formulations of large-deformation-diffeomorphic-mapping (LDDMM) registration [2,6]) while relying on conservative assumptions about the temporal changes in image appearance.

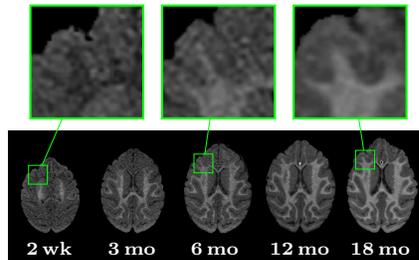


Fig. 1. MR images of the developing monkey brain (2 weeks through 18 months). Unmyelinated white matter in the early stages of development appears darker than the myelinated white matter in later stages.

Commonly used global similarity measures (sum of squared differences (SSD), normalized cross correlation (NCC), mutual information (MI) [10]), for instance, expect a one-to-one relationship between the spatially corresponding intensities (or intensity ranges in the case of mutual information due to histogram binning) of the different time-points. This is not a valid assumption for certain registration problems. In longitudinal magnetic resonance (MR) imaging studies of brain development, for example, the biological process of myelination causes a substantial shift in the MR appearance of white matter tissue that is both spatially and temporally non-uniform [1,8] (Fig. 1 shows MR images of the developing brain). As a result, deformable registration methods that use global similarity measures often fail to recover the correct alignment in this setting since inconsistencies in appearance can be resolved by introducing erroneous local deformations that are not supported by the underlying structural information [5].

The various approaches that have been proposed for aligning images with non-uniform appearance change (often for less severe intensity variation arising from magnetic field inhomogeneities) either (i) locally minimize some global similarity measure in overlapping subregions that are small enough to have near constant intensities within tissue types [9,5], or (ii) jointly with registration, estimate an intensity transformation that makes the images similar [3,7]. While local methods are appropriate for aligning images with spatially smooth and slowly varying intensity changes within tissue classes, the trade off between registration accuracy and subregion size means that they are inadequate for the strong intensity gradients seen in myelinating white matter tissue. Intensity transform methods either have similar spatial limitations due to the slowly varying basis functions used to approximate the intensity transform [3], or discard spatial information and therefore cannot capture complex intensity transformations [7].

Here, we formulate a model-based similarity measure (mSM) that estimates local appearance change over time. Once the temporal model is estimated, existing deformable registration methods can also be used with the model to recover the correct alignment by changing the appearance of one image to match the other. After formulating our method in the following section, we first demonstrate in Sect. 3 that MI and our approach both perform well with an affine transformation model in the presence of non-uniform appearance change, but then show that using the model-based approach is critical for deformable registration. This method can either (i) estimate the temporal intensity change model without any prior, or (ii) use a known model for the initial alignment.

2 Modeling Appearance Change

We introduce a spatio-temporal model of appearance change into a general registration framework via a model-based similarity measure. To motivate the model-based similarity measure consider the standard sum of squared differences (SSD) similarity term. If a transformation model is involved the SSD can be written as

$$SSD(I_0, I_1) = \int_{\Omega} (I_0(x) \circ \Phi(x) - I_1(x))^2 dx, \quad (1)$$

where $\Phi(x)$ is the transformation that maps the coordinate system of image I_0 to that of image I_1 . If we consider the SSD measure a simple model-based registration method, where the image model is simply the given image $I = I_0$, then SSD aims to minimize the squared residuals to this model subject to the sought for transformation. We will therefore consider SSD a special case of a sum of squared residual (SSR) model. With a time-dependent image model and a generalization to multiple images, we can write the corresponding SSR as

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

where $I_i(\cdot, t_i)$ denotes the measured image at time-point t_i and $\hat{I}(\cdot, t_i)$ the model (estimate) at the same time-point. Note that for two time-points and $\hat{I} = I_1$ the model simplifies to the standard SSD. For simplicity, consider a quadratic (in-time) appearance model

$$\hat{I}(x, t) = \alpha(x)t^2 + \beta(x)t + \gamma(x), \quad (2)$$

where α , β , and γ are spatially varying model coefficients.

2.1 Transformation Model

Assume that we aim to estimate the affine transform of the form $Ax + b$ for each image back to the coordinate system of the model \hat{I} and denote the set of these transformations as $\{A_i, b_i\}$. Then the registration model becomes

$$SSR(\{I_i\}, \{A_i, b_i\}, \alpha, \beta, \gamma) = \sum_{i=0}^{n-1} \int_{\Omega} (I_i(A_i x + b_i, t_i) - \hat{I}(x, t_i))^2 dx, \quad (3)$$

resulting in the point-wise linear system

$$\sum_{i=0}^{n-1} \begin{pmatrix} t_i^4 & t_i^3 & t_i^2 \\ t_i^3 & t_i^2 & t_i \\ t_i^2 & t_i & 1 \end{pmatrix} \begin{pmatrix} \alpha(x) \\ \beta(x) \\ \gamma(x) \end{pmatrix} = \sum_{i=0}^{n-1} I_i(A_i x + b_i, t_i) \begin{pmatrix} t_i^2 \\ t_i \\ 1 \end{pmatrix},$$

which amounts to a local fitting of the quadratic model (2) given the current estimate of the affine transformation parameters. Any other model could be substituted here. Note that the image-comparison terms in (3) are strictly with respect to the model (2) which is estimated jointly. Estimation of the model and the affine parameters can then be accomplished by alternating model fitting and transformation parameter estimation steps. In the extreme case one (i) estimates the affine transforms given the current model then (ii) re-estimates the model and then repeats these two steps to convergence. Note also that there is a rotational ambiguity, so one of the coordinate systems should be fixed, e.g., $A_0 = x, b_0 = 0$.

Here, to introduce the decoupled registration and model estimation steps, we started with an affine transformation model. However, the same principle can be applied to more flexible registration methods, such as the deformable elastic registration used for the experiments in Sect. 3.

2.2 Spatial Regularization

Instead of estimating the appearance model parameters independently for each voxel, we can get a more robust estimate by estimating the parameters over subregions of the image, where the subregions are defined on a template (atlas) image. This, however, still leaves the problem of choosing the image subregions. If tissue segmentation is available, one reasonable choice would be to estimate the parameters for each tissue class. Using this approach, for each assumed to be uniform template region R_l the parameter fitting equations for the quadratic case become

$$|R_l| \sum_{i=0}^{n-1} \begin{pmatrix} t_i^4 & t_i^3 & t_i^2 \\ t_i^3 & t_i^2 & t_i \\ t_i^2 & t_i & 1 \end{pmatrix} \begin{pmatrix} \alpha_l \\ \beta_l \\ \gamma_l \end{pmatrix} = \sum_{i=0}^{n-1} \int_{R_l} I_i(A_i x + b_i, t_i) dx \begin{pmatrix} t_i^2 \\ t_i \\ 1 \end{pmatrix},$$

everything else stays the same. Here, $|R_l|$ denotes the cardinality of the set R_l and α_l , β_l , and γ_l denote the model parameters valid (constant) on R_l . This approach, however, assumes that the intensity change within a tissue class is spatially uniform.

2.3 Model Parameter Estimation

For the current implementation of the method spatial regularization of the model parameters was achieved by a subregion based approach shown in Fig. 2. Since the intensity change in the white matter occurs dominantly in the posterior-anterior (PA) direction (see Fig. 3) we chose subregions, R_l , that span across the white matter perpendicular to the posterior-anterior direction. So far, this is the same as the subregion method described above. However, instead of estimating the model parameters from R_l only, we use a neighborhood N_l of width w centered on R_l and use robust statistics to estimate the model parameters for R_l . The neighborhoods for adjacent R_l are overlapping and therefore encourage spatial regularization in the posterior-anterior direction.

Estimation of the model and affine parameters can be accomplished by alternating the model fitting and transformation parameter estimation steps. In the extreme case we can first estimate the transformation given the model as a separate registration step and then re-estimate the model and repeat these two steps until convergence. In fact, taking this idea even further, one can change the

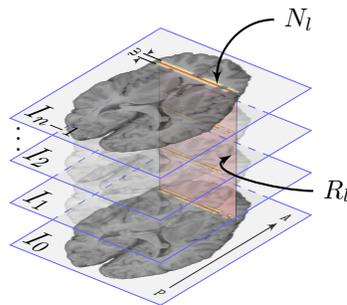


Fig. 2. Subregion based model estimation. Each subregion R_l (red) is defined on the available white matter segmentation (usually at the last-time point). The subregions are perpendicular to the PA direction. The model parameters for R_l are estimated from a neighborhood N_l of width w (yellow).

appearance of the estimated image according to the model and use any registration method to estimate the transformation parameters. Here, we use the latter approach which allows the testing of existing registration algorithms with mSM. The algorithm for iterative registration and estimation is set up as follows:

- 0) Initialize model (\hat{I}) parameters to $\alpha = \alpha_0, \beta = \beta_0, \gamma = \gamma_0$.
- 1) Affinely pre-register images $\{I_i\}$ to \hat{I} .
- 2) Estimate the appearance of \hat{I} at times $\{t_i\}$, giving $\{\hat{I}(t_i)\}$.
- 3) Estimate displacement fields $\{u_i\}$ by registering images $\{I_i\}$ to $\{\hat{I}(t_i)\}$.
- 4) Estimate model parameters α, β, γ from the registered images $\{I_i \circ u_i\}$
- 5) Repeat from step 2 until convergence.

Convergence was achieved when the change in the registration energy function between iterations was below tolerance (typically less than 5 iterations).

3 Experimental Results

The similarity measures were tested by registering pairs of 2D synthetic images with a known ground truth transformation between them. The registration accuracy was determined by computing the distance between the ground truth and the recovered transformation. The root mean squared (RMS) error of the voxel-wise distance within the mask of the target image then yielded the registration error. All experiments are in 2D, but the method generalizes to 3D.

3.1 White Matter Intensity Distributions from Real Data

An important part of the registration experiments is testing the similarity measures on realistic appearance change while knowing the ground truth deformations. To this end, we calculated the spatial and temporal intensity changes from the MR images of 9 rhesus monkeys during the first 12 months of life. The white matter intensity trajectories acquired from the real monkey data were then used to generate the simulated brain images for Experiment 2 (Sect. 3.3).

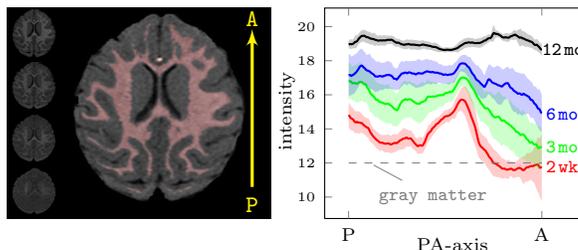


Fig. 3. Spatio-temporal distribution of white matter intensities in 9 monkeys. A single slice from each time-point is shown in order in the left column (2 week at the bottom), and the white matter segmentation (red) at 12 months is shown in the middle. Plotted, for each time-point, the mean (line) ± 1 standard deviation (shaded region) of the spatial distribution of the white matter intensities averaged over the whole brain of each monkey in the PA direction. The images were affinely registered and their gray matter intensity distributions matched.

The spatial white matter distributions were calculated for each time-point (2 week, 3, 6, 12 month) of the 9 monkeys. The early time-points have low gray-white matter contrast, therefore the white matter segmentation of the 12 month image was transferred to the earlier time-points (this is often the case for longitudinal studies where good tissue segmentation might only be available at the latest time-point). Due to the few images available at this stage of the study, we averaged the white matter intensity change of the whole brain in a single dimension along the posterior-anterior direction (most of the intensity change is along this direction [4]). Figure 3 shows the mean and variation of the white matter intensity profiles from all four time-points. Myelination starts in the posterior and central regions of the white matter and continues towards the periphery and, dominantly, towards the anterior and posterior regions. These findings agree with existing studies on myelination [4]. Of note is the strong white matter intensity gradient in the early time-points due to the varying onset and speed of the myelination process.

3.2 Experiment 1: Synthetic Data

In this experiment, we created sets of 64×64 2D synthetic images. Each set consisted of 11 time-points (I_i , $i = 0, \dots, 10$). I_0 was designated as the target image and all subsequent time-points as the source images. The gray matter intensities of all 11 images were fixed ($I_i^{\text{gm}} = 80$). For the source images, I_1, \dots, I_{10} , we introduced two types of white matter appearance change:

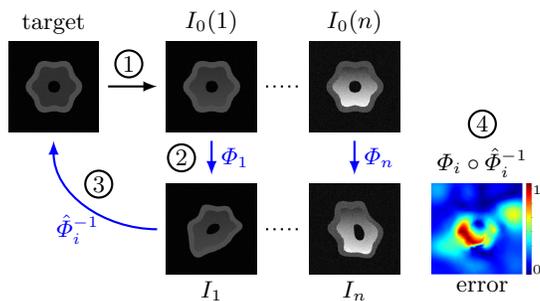


Fig. 4. Experimental setup: 1) Increasing white matter intensity gradient is added to the target, I_0 . 2) Adding known random deformations yields the source images, 3) which are registered back to the target. 4) Registration error is calculated from the known (Φ_i) and recovered ($\hat{\Phi}_i^{-1}$) transformations.

- i) Uniform white matter appearance change over time, starting as dark (unmyelinated) white matter ($I_1^{\text{wm}} = 20$) and gradually brightening (myelinated) white matter ($I_{10}^{\text{wm}} = 180$) resulting in contrast inversion between gray and white matter. The target white matter intensity was set to 100.
- ii) White matter intensity gradient along the posterior-anterior direction with increasing gradient magnitude over time. The target image had uniform white matter ($I_0^{\text{wm}} = 50$). For the source images the intensity gradient magnitude increased from 1 to 7 intensity units per pixel (giving $I_1^{\text{wm}} = \{50, \dots, 70\}$ up to $I_{10}^{\text{wm}} = \{50, \dots, 200\}$). These gradients are of similar magnitude as observed in the real monkey data.

We tested the similarity measures for two types of transformation models: affine; and deformable with elastic regularization. Figure 4 shows the experimental

setup with deformable transformation model (for the affine registration experiments Φ_i was an affine transform; for the deformable registration experiment Φ_i was a spline deformation with 20 control points). The aim of the experiment was to recover the ground truth inverse deformation, Φ_i^{-1} , by registering the 10 source images to I_0 (giving $\hat{\Phi}_i^{-1}$) with each of the four similarity measures (SSD, NCC, MI, mSM). We repeated each experiment 100 times for each transformation model with different random deformations giving a total of 16000 registrations (2 white matter change \times 2 transformation model \times 10 source image \times 4 measure \times 100 experiment). Significance was calculated with Welch’s t -test (assuming normal distributions, but unequal variances) at a significance level of $p < 0.01$.

Note that the synthetic images have longitudinal intensity changes over time, but the random spatial deformations do not have a temporal model. This is intended to be a challenging scenario for the parameter estimation, as true longitudinal data is much less spatially variable, and avoid bias towards any particular longitudinal growth model. Next, we describe the results of the experiments for each transformation model.

Affine transformation model. Affine registration is often appropriate for images from the same adult subject. In our case, it is only a preprocessing step to roughly align the images before a more flexible, deformable registration. Nevertheless, the initial alignment can greatly affect the initial model estimation and the subsequent deformable solution. Therefore we first investigate the sensitivity of affine registration to white matter appearance changes separately from deformable registration.

Figure 5 shows the results for registering I_1 through I_{10} to the target image I_0 from multiple sets ($n = 100$, giving 1000 pair-wise registrations for each similarity measure) of longitudinal images with both uniform and gradient spatial white matter intensity profiles. A registration error of less than 1 voxel can be considered good alignment.

With uniform white matter, all four measures performed well when the contrast of the source image was close to the contrast of the target image (near 0 white matter intensity difference in the first plot of the median root means squared registration error). The results for the gradient white matter profiles show that the performance of both SSD and NCC declined as the gradient magnitude increased, while MI and mSM aligned the images well even with the strongest gradient. Overall, mSM significantly outperformed SSD and NCC but not MI, however, for individual time-points mSM did significantly better for 3, . . . , 10.

The experiments suggest that affine registration can be reliably achieved by MI or mSM, but for simplicity MI should be used if affine alignment is the only objective.

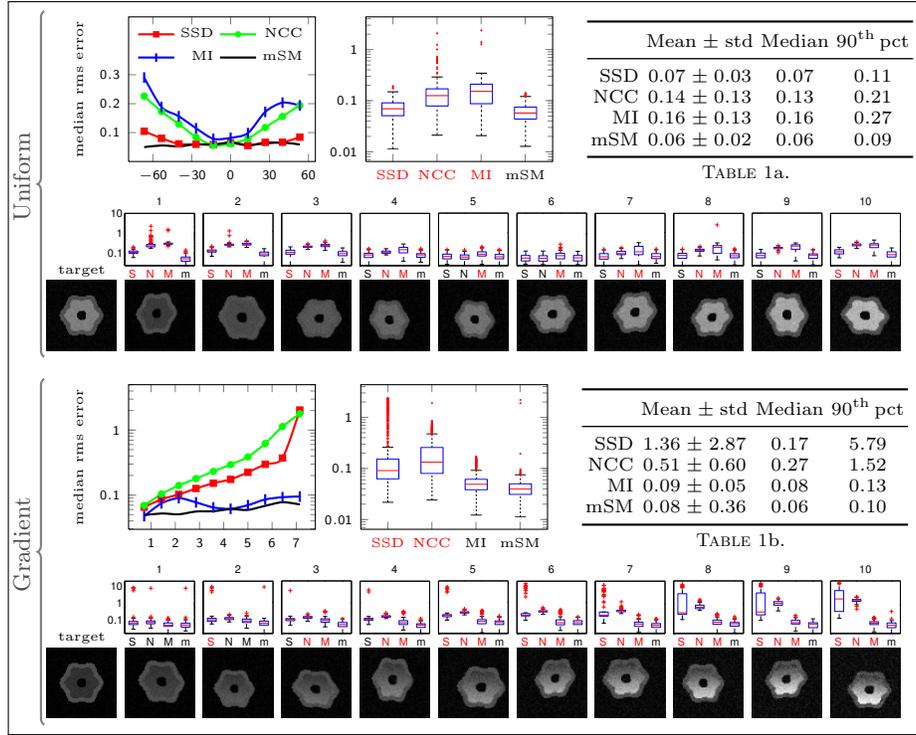


Fig. 5. Results for Experiment 1 with affine transformation, uniform and gradient white matter intensity change. For uniform white matter, the line plot shows the median RMS error vs. the white matter intensity difference between the source and the target images ($I_t^{\text{wm}} - I_0^{\text{wm}}$) for each time-point (0 means the images have the same contrast). For the gradient white matter, the x-axis of the line plot is the magnitude of the gradient. The boxplots and the tables summarize the aggregate results over all time-points (the box is the 25th and 75th percentile, the red line is the medium, the whiskers are $1.5 \times$ interquartile range, and the red marks are outliers). The small boxplots show results for each time-point (S, N, M, and m are SSD, NCC, MI, and mSM respectively). For each boxplot, the x-label is highlighted in red if mSM performed significantly better than that particular measure. The row of images shows the target and source images for a single trial. Note that all box plots and the bottom line plot have log y scales.

Deformable registration. Similarly to the affine experiment, Fig. 6 shows the error plots for deformable registrations in the presence of white matter intensity change. For uniform white matter, SSD again produced small registration errors when the contrast difference was small, but fared worse than MI and mSM in the presence of large intensity differences between the target and the source images. mSM performed slightly better than MI for all time-points.

The setup with deformable registration and white matter gradient resembles the real problem closely and therefore is the most relevant. Here, SSD and NCC introduced considerable registration errors with increasing gradient magnitude. The registration error of MI remained under 2 voxels (mean = 1.62 ± 0.45), while mSM led to significantly less error (mean = 1.25 ± 0.35) for all time-points.

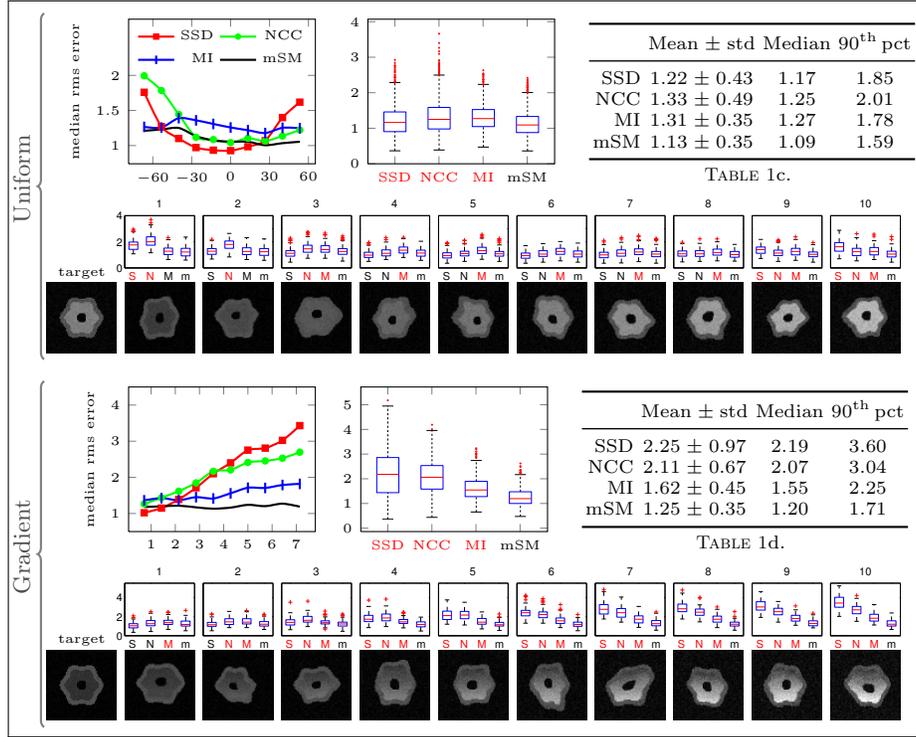


Fig. 6. Experiment 1 results with deformable transformation. The graphs are set up similarly as in Fig. 5 except all plots have linear y scales. The last setup with deformable transformation model and gradient white matter intensity is the most challenging and relevant to the real world problem.

3.3 Experiment 2: Simulated Brain Data

The next set of experiments used simulated brain images with white matter intensity distributions based on the monkey data. Four time-points I_0, \dots, I_3 were generated corresponding to the four time-points of the monkey data. At each time-point, the spatial white

matter intensity distribution of the simulated image was obtained by a random perturbation of the mean monkey white matter distribution for that particular time-point (see Fig. 3). The local magnitude of the perturbation was proportional to the local variation of monkey white matter data, therefore the generated curves had similar variation to the real data. The first time-point was designated as the target image. The other three time-points were deformed by a random deformation. The source images generated this way, similarly to Experiment 1,

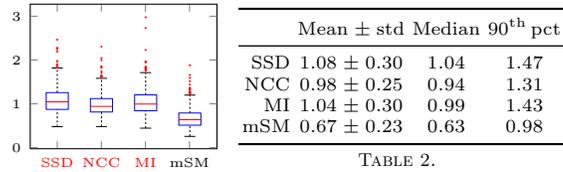


Fig. 7. Experiment 2 results. mSM lead to significantly better alignments than the global measures.

were then registered to I_0 with the four similarity measures. The experiment was repeated 200 times, each time with different random white matter intensity profiles and different random deformations. The boxplot and Tab. 2 in Fig. 7 show the aggregate results for all the time-points. mSM performed significantly better than SSD, NCC, and MI.

4 Conclusions

We presented a temporally-dependent model-based similarity measure and compared it to three of commonly used measures. mSM performed significantly better in the majority of experiments than the other measures, especially in the presence of considerable intensity gradients. These experiments provide strong evidence for the usefulness of a model based approach. Considerable improvement might be achieved by better model selection, spatial regularization of the model parameters, and improved model of intensity variation (instead of the quadratic model). These improvements and validation on real 3D data will be part of future work.

Acknowledgments. This work was supported by NSF EECS-1148870, NSF EECS-0925875, NIH NIHM 5R01MH091645-02, NIH NIBIB 5P41EB002025-28, U54 EB005149.

References

1. [Barkovich, A.J., Kjos, B.O., Jackson, D.E., Norman, D.: Normal maturation of the neonatal and infant brain: MR imaging at 1.5T. *Radiology* 166, 173–180 \(1988\)](#)
2. [Durrleman, S., Pennec, X., Trounev, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. *Med Image Comput Comput Assist Interv* 12\(1\), 297–304 \(2009\)](#)
3. [Friston, K., Ashburner, J., Frith, C., Poline, J., Heather, J.D., Frackowiak, R.: Spatial registration and normalization of images. *Human Brain Mapping* 2, 165–189 \(1995\)](#)
4. [Kinney, H.C., Karthigasan, J., Borenshteyn, N.I., Flax, J.D., Kirschner, D.A.: Myelination in the developing human brain: biochemical correlates. *Neurochem Res* 19\(8\), 983–996 \(1994\)](#)
5. [Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. *IEEE Transactions on Medical Imaging* 29\(1\), 19–29 \(2010\)](#)
6. [Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. *Med Image Comput Comput Assist Interv* 14\(2\), 655–662 \(2011\)](#)
7. [Roche, A., Guimond, A., Ayache, N., Meunier, J.: Multimodal elastic matching of brain images. In: *Proceedings of the 6th European Conference on Computer Vision-Part II*. pp. 511–527. *ECCV '00*, Springer-Verlag, London, UK \(2000\)](#)
8. [Sampaio, R.C., Truwit, C.L.: Myelination in the developing brain. In: *Handbook of developmental cognitive neuroscience*, pp. 35–44. MIT Press \(2001\)](#)
9. [Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Transactions on Medical Imaging* 25\(5\), 626–639 \(2006\)](#)
10. [Viola, P., Wells, W. M., I.: Alignment by maximization of mutual information. In: *Proc. Conf. Fifth Int Computer Vision*. pp. 16–23 \(1995\)](#)