Comments and Controversies

Avoiding asymmetry-induced bias in longitudinal image processing

Martin Reuter *, Bruce Fischl

Martinos Center for Biomedical Imaging, Charlestown, MA, USA
Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA
MIT Computer Science and AI Lab, Cambridge, MA, USA

ARTICLE INFO

ABSTRACT

Longitudinal image processing procedures frequently transfer or pool information across time within subject, with the dual goals of reducing the variability and increasing the accuracy of the derived measures. In this note, we discuss common difficulties in longitudinal image processing, focusing on the introduction of bias, and describe the approaches we have taken to avoid them in the FreeSurfer longitudinal processing stream.

© 2011 Elsevier Inc. All rights reserved.

Over the past decade an increasing amount of longitudinal image data has become available, where two or more scans are collected on the same subject over a period of time. Compared with cross-sectional studies, a longitudinal design can significantly reduce the confounding effects of inter-individual morphological variability by using each subject as his or her own control. As a result, longitudinal imaging studies are increasing in popularity in both basic neuroscience and clinical studies. In vivo cortical and subcortical measures are useful as biomarkers of the evolution of many neurodegenerative diseases, and are thus of great potential utility in evaluating the efficacy of disease-modifying therapies. For these reasons, it is critical to obtain robust and reliable morphological measurements by incorporating additional temporal information within a longitudinal processing stream. The expected reduction in variability of automatic measurements allows studies with smaller sample sizes to detect effects with the same power and significance level or provides increased sensitivity, necessary to detect small effects in drug trials.

A common issue with longitudinal image processing is the introduction of bias, leading to incorrect results and a potentially flawed interpretation of outcome measures. The recent critique by Wesley K. Thompson and Dominic Holland (2011), for example, points out a bias potentially due to inconsistent image registration with baseline images in the work published by Hua et al. (2010) hinting at severe underestimation of sample sizes. Bias can be introduced due to several different reasons, some of which can be controlled and are inherent to the processing stream, while others are introduced earlier, e.g., in the acquisition when scanning different time points on different scanners, software versions or with different scanning parameters. Salat et al. (2009) and Westlye et al. (2009) describe another source of potential bias of more general nature: the

intrinsic magnetic properties of the tissue (e.g., T1, T2*) can change over time in aging and also neurodegenerative disease, possibly introducing bias in measures of thickness or volumes in both longitudinal or independent (cross sectional) processing. In the following paragraphs we will focus on three common situations of bias introduced by different treatment of the input time points (usually the baseline image) and describe the steps taken to avoid them in the current FreeSurfer longitudinal processing stream (http://www.ssurfer.nmr.mgh.harvard.edu/fswiki/LongitudinalProcessing, Reuter et al., 2010). A detailed description of the longitudinal stream in FreeSurfer will be subject of a different paper.

Biased registration (lack of inverse consistency)

In order to study differences across time within the same subject, images are frequently registered using either a rigid or affine registration or higher dimensional non-linear warps. Typically, a first step is to remove pose differences due to varying head positions and orientations in the scanner. For this, a rigid registration is performed, to correct for translational and rotational differences (in some cases 9 or 12 degrees of freedom or more complex combinations are used to allow for additional scaling to account for factors such as differential scanner calibration, e.g., Smith et al. 2001).

Many contemporary registration algorithms may introduce a bias as they are not designed to be inverse consistent. Here, inverse consistency means that one expects to obtain the inverse transform when registering B to A as opposed to A to B. Several inverse consistent approaches exist for nonlinear warps. Often, both forward and backward warps are jointly estimated, e.g., (Christensen and Johnson, 2001; Zeng and Chen, 2008), while others match at the midpoint (Beg and Kahn, 2007). Symmetric pairwise registration is, for example, used in Avants et al. (2007) and Nakamura et al. (2011). Avants et al. (2007) use a non-linear registration (SyN) to construct a spatiotemporal parametrization; however, for longitudinal processing, the baseline is treated as a reference frame, differently from follow-up images, and the 1 year image is sampled from

* Corresponding author at: Martinos Center for Biomedical Imaging Charlestown, MA 02129, USA.
E-mail addresses: mreuter@nmr.mgh.harvard.edu (M. Reuter), fischl@nmr.mgh.harvard.edu (B. Fischl).

the spatiotemporal diffeomorphism. Nakamura et al. (2011) combine forward and inverse linear registrations to construct symmetric pairwise registrations, but register all follow-up images to baseline. We will discuss the possibilities of introducing bias via interpolation asymmetry or different treatment of specific time points in the sections below.

In the rather simple rigid registration case, inverse consistency is often tacitly assumed, but as shown in Reuter et al. (2010), it is not necessarily guaranteed. In order to avoid the introduction of this type of bias we developed a robust rigid registration method, as described in that paper, in which both the algorithm and the mathematical model are designed to be symmetric by mapping source and target images to a half way space. In several rigid registration tools, one image (source or movable) is resampled internally to the space of the target image, which destroys the symmetry. Our approach avoids the construction of both forward and inverse transforms. Furthermore, based on robust statistics (Nestares and Heeger, 2000), our algorithm increases the accuracy of the registration by reducing the influence of outlier regions (different jaw, neck, eye positions, motion artifacts, morphological change such as atrophy or tumor growths, or differences induced by the imaging technologies, such as gradient non-linearities). This is achieved by iteratively detecting and ignoring real image differences that cannot be accounted for by the registration.

Biased resampling space (interpolation asymmetry)

Longitudinal studies frequently acquire more than two images for each subject. In order to reduce variability and detect true change in structure or function, it is important to be able to transfer and share information across time. A standard way of constructing spatial correspondence is to register all follow-up images to the baseline scan. The images are then either resampled at that location or the inverse maps are used to transport information into native voxel spaces. Either way, even when using an inverse consistent pairwise registration tool, the baseline scan is used as a reference frame and therefore is treated differently from the other time points, introducing a potential bias. For example, resampling other time points to the baseline will smooth these images and of course alter any downstream measurements. This typically results in different (usually steeper) slopes in the rate of change between baseline and time two, than between the follow up images. The effect of this bias can easily be verified by choosing a different time point as the registration target and observing the corresponding change in slopes. Yushkevich et al. (2010) also documented this type of bias, likely due to the asymmetric interpolation in the linear registration pre-processing step.

In order to avoid this type of effect, it is essential to treat all time points identically and ensure they undergo the same degree of smoothing due to image interpolation. In FreeSurfer, for example, we resolve this problem by creating an unbiased within-subject template space for the common registration of all input images of the given subject. Resampling all inputs to the same voxel space further reduces variability of measurements. Note that for the specific case of two time points Smith et al. (2001) already avoided this type of bias by resampling both input images at the midpoint (half way space) after the (not necessarily inverse consistent) registration procedure. Also, see Joshi et al. (2004) and Avants and Gee (2004) for a similar approach in unbiased non-linear atlas creation, warping several images to a mean shape.

Biased information transfer

As described above, information is often transferred across time within each subject. For example, a common skull strip or shared Talairach transform can significantly reduce variability. Cortical and subcortical segmentation and parcellation procedures involve solving many complex nonlinear optimization problems, such as

deformable surface reconstruction, nonlinear atlas-image registration, and nonlinear spherical surface registration. These optimization problems are typically solved with iterative methods, and the final results are known to be sensitive to the selection of a particular starting point. Therefore, initializing later time points with earlier results (e.g., typically the baseline) will certainly improve the consistency, but at the cost of introducing a bias, as again the baseline image is treated in a fundamentally different manner than subsequent time points.

Possible solutions to remove this bias exist. One is to design the algorithms to optimize all time points simultaneously. Xue et al. (2006) aim in this direction and jointly segment the 4D volume within subject, but they also treat the baseline scan differently from follow-up images by using it as the reference frame for their registration. Unbiased simultaneous processing often involves a complete redesign and can be quite time consuming. Furthermore, memory usage is scaled by the number of time points, which implies that hardware requirements may not be met by standard desktop computers. A different approach is to create an unbiased subject template to describe the average subject anatomy across time. This template can be fully processed and many of the results, e.g., surface locations, can be used to initialize all the time points independently. For this purpose, FreeSurfer uses a robust median image of the co-registered inputs mapped to the unbiased template space.

Additional comments

Any of these biases can, of course, be removed by independently processing the inputs at the cost of increased variability. As soon as longitudinal information is incorporated as "prior knowledge," bias is introduced, for example due to temporal smoothing, and accuracy may suffer particularly when measuring large longitudinal change. Therefore, it is theoretically possible that changes of greater magnitude are underestimated by initializing each time point with common information from the template as done in FreeSurfer. While a more conservative estimate of change is often preferable in a power analysis, than an overestimation, we aim for accurate and unbiased results. The longitudinal stream in FreeSurfer therefore allows for more flexibility by using a probabilistic voting scheme of independently processed label maps from all time points, to determine the probability of a specific voxel having a specific label by weighting labels across time according to their intensity similarity. This fused segmentation is usually very similar to the time point's independent segmentation (slightly temporally smoothed). It is not the final solution, but then used to initialize the segmentation algorithm for each time point instead of the fixed segmentation of the subject template to allow for larger departures from the subject average, evident, for example, in several years of neurodegeneration.

Also note that, as soon as longitudinal data is employed, one needs to delay processing until all time points are available to remain unbiased. This is often not feasible and it is of course possible to add time points later and process them with the "old" template, created from the initial subset of time points. While this is clearly introducing bias, it is unclear how large the effect will be and likely depends strongly on the specific situation (e.g., how many time points were used for the template creation and how many were added, etc.). Reprocessing everything with a new template can change earlier results as the new template might be shifted towards a more diseased state. However, we believe it is favorable to have the template somewhere in the middle of the time series than closer towards the front, in order not to be biased towards a more healthy state.

The above discussion highlights several challenges of longitudinal image processing and underlines the importance of selecting methods carefully to avoid introducing a bias by treating individual inputs differently, which can be prevented, or by biasing towards no change

when encouraging reliability too intensely. For the second case, a good trade-off needs to be aimed for.

References

Avants, B., Gee, J.C., 2004. Geodesic estimation for large deformation anatomical shape averaging and interpolation. Neuroimage 23 (1), 139–150.

Avants, B., Anderson, C., Grossman, M., Gee, J.C., 2007. Spatiotemporal normalization for longitudinal analysis of gray matter atrophy in frontotemporal dementia. Proceedings Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007: Lecture Notes in Computer Science, 4792, pp. 303–310.

Beg, M.F., Kahn, A., 2007. Symmetric data attachment terms for large deformation image registration. IEEE Trans. Med. Imaging 26 (9), 1179–1189.

Christensen, G.E., Johnson, H.J., 2001. Consistent image registration. IEEE Trans. Med. Imaging 20 (7), 568–582.

Hua, X., Lee, S., Hibar, D.P., Yanovsky, I., Leow, A.D., Toga, A.W., Jack Jr., C.R., Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.W., Alzheimer's Disease Neuroimaging Initiative, 2010. Mapping Alzheimer's disease progression in 1309 MRI scans: power estimates for different inter-scan intervals. Neuroimage 51, 63–75.

Joshi, S., Davis, B., Jomier, B.M., Gerig, G., 2004. Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage 23, 151–160.

Nakamura, K., Fox, R., Fisher, E., 2011. CLADA: cortical longitudinal atrophy detection algorithm. Neuroimage 54 (1), 278–289.

Nestares, O., Heeger, D.J., 2000. Robust multiresolution alignment of MRI brain volumes. Magn. Reson. Med. 43 (5), 705–715.

Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. Neuroimage 53 (4), 1181–1196.

Salat, D., Lee, S.Y., van der Kouwe, A., Greve, D.N., Fischl, B., Rosas, H.D., 2009. Age-associated alterations in cortical gray and white matter signal intensity and gray to white matter contrast. In NeuroImage 48 (1), 21–28.

Smith, S.M., De Stefano, N., Jenkinson, M., Matthews, P.M., 2001. Normalized accurate measurement of longitudinal brain change. Comput. Assist. Tomogr. 25 (3), 466–475.

Thompson, W.K., Holland, D., 2011. Bias in tensor based morphometry Stat-ROI measures may result in unrealistic power estimates. NeuroImage 57, 1–4.

Westlye, L.T., Walhovd, K.B., Dale, A.M., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Greve, D.N., Fischl, B., Fjell, A.M., 2009. Increased sensitivity to effects of normal aging and Alzheimer's disease on cortical thickness by adjustment for local variability in gray/white contrast: a multi-sample MRI study. Neuroimage 47 (4), 1545–1557.

Xue, Z., Shen, D., Davatzikos, C., 2006. CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. Neuroimage 30, 388–399.

Yushkevich, P.A., Avants, B.B., Das, S.R., Pluta, J., Altinay, M., Craige, C., ADNI, 2010. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3 tesla MRI data. Neuroimage 50 (2), 434–445.

Zeng, Q., Chen, Y., 2008. Accurate Inverse Consistent Non-rigid Image Registration and Its Application on Automatic Re-contouring. Bioinformatics Research and Applications, ISBRA'08, Lecture Notes in Computer Science 4983, 293–304.