Experimental Plan for Assessment of Non-Rigid Registration in 3-D

Abstract

This collection of notes will be treated as an experimental plan. For context, relevant background is prepended as well. Herein it is assumed that 3-D data, which has been non-rigidly registered in advance, is available for assessment of the quality of registration. Granted, 4 different registrations are put to the test and a distinguishable difference – in terms of quality – exists among them. The document describes steps which need to be taken in order to evaluate and compare these 4 registrations of one 3-D brain dataset. Each registration makes use of a different algorithm, the intent being to conduct a benchmark-type study. In order to ensure that this laborious process is fruitful, it is vital to run evaluations in parallel, assessing the results as they emerge and get gradually refined.

1. Background

Non-rigid registration (NRR) of a group of images [6, 4] establishes correspondence among the set. That correspondence can also be assessed using measures of overlap between labels in the images [2] or by creating degraded test sets [5] and attempting to recover a correct solution. A new method, which has already been validated in 2-D, is proposed as the means for assessing 3-D registration of images of the brain. Ultimately, different NRR algorithms can be compared and ranked based on their performance.

2. Data and NRR Experiments

In this particular set of experiments we deal with the IBIM dataset, which comprises 38 brain volumes (with labels), acquired from different subjects. The data is being registered onto a common frame of reference, using 4[1] different NRR algorithms, and its overlap [2] is being computed to verify that the algorithm leads to distinguishable results. A model can then be built from the registrations. NRR establishes a correspondence from which models of deformation can be built [3]. In this particular case, 4 appearance models of shape and intensity [1] are being built from the resultant correspondence. By measuring the quality of the models, the quality of the registration can be studied as well.

[1] It may be safer to attempt this with 5 NRR algorithms, then choosing just 4 that are most distinguishable, potentially by hand-tweaking various free parameters. Without significant difference in quality, it would be hard to back the results using ground-truth label overlap.

3. Assessment of NRR

To study the quality of the model (from which the quality of NRR can be directly inferred), 1000 synthetic volumes are derived from that model. Using MATLAB (or potentially C++/VXL), each of these synthetic images is compared against the original (training) dataset – that which was registered and built the model. The shuffle distance between images in the training and synthetic sets can be aggregated or averaged to give a figure of merit. Such figures can be derived which indicate how good or bad the registration was.

In shuffle distance estimates, it is preferable to use a radius rather than consider just a cube (or box) of pixel as a neighbourhood. A radius of 2.5 voxels is said to be preferable (corresponds roughly to a box of 5x5x5 voxels). There emerges another issue: the thickness of the slices differs from the resolution in pixels, for any given slice. This means that a better solution should involve an elliptical neighbourhood (i.e. 2 radii). Another possibility is to do a plain-type comparison, comparing one slice against another corresponding slice. This can be done in a similar fashion to 2-D experiments, treating a volume as a set of plains.

4. Technical Details

The size of the synthetic set (namely 1000 high-resolution volumes) is a significant peril, due to scale (complexity), as well as runtime (duration). There are ways of addressing these issues. One is to distribute the work among several workstations. The other is to run assessments of different models *in parallel*, accumulating results and refining them progressively, rather than handle the assessments one at the time, i.e. serially. This has the advantage that results emerge rather early and give insight into whether the experiments are worthwhile or not.

The computer clusters will be running Fedora Core 5 with MATLAB 6.5. The time estimated for one assessment to be completed depends on the number of workstations used, but it most probably lies between the range of weeks and one month.

References

[1] T. F. Cootes, G.J. Edwards and C.J.Taylor. Active appearance models. In *European Conference on Computer Vision*, 2:484-498, 1998.

[2] W. R. Crum, O. Camara, D. Rueckert, K. Bhatia, M. Jenkinson, and D. L. G. Hill. Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 3749:99-106, 2005.

[3] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8)1014-1025, 2003.

[4] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, D. J. Hawkes. Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712-721, 1999.

[5] J. A. Schnabel, C. Tanner, A. Castellano-Smith, M. O. Leach, C. Hayes, A. Degenhard, R Hose, D. L. G. Hill, and D. J. Hawkes. Validation of non-rigid registration using finite element methods. In *Information Processing in Medical Imaging,* 2082:344-357, 2001.

[6] B. Zitova and J. Flusser. Image registration methods: a survey. *Image Vision Computing*, 21:977-1000, 2003.