Automated high-throughput registration for localizing 3D mouse brain gene expression using ITK

L. Ng[1], M. Hawrylycz[1], and D. Haynor[2]

[1] Allen Institute for Brain Science, Seattle, Washington,
lydian@alleninstitute.org
[2] Department of Radiology, University of Washington.

Abstract. The Allen Brain Atlas (ABA) project aims to create a cellular-resolution, genome-wide map of gene expression in the adult mouse brain. The resulting in situ hybridization (ISH) image data will be available free-of-charge to the public. Additionally, we are developing an informatics pipeline to support searching of the data by anatomic region and expression level and/or pattern. This paper describes a robust, high-throughput registration scheme to automatically annotate hierarchical brain structures in the ISH imagery.

1 Introduction

The Allen Institute for Brain Science (AIBS) has developed a high-throughput *in situ hybridization* (ISH) platform [1] to generate gene expression maps of the mouse brain. The goal of our inaugural project, the Allen Brain Atlas (ABA) is to survey the entire mouse transcriptome of C57/BL6 within three years and to provide an online searchable database of 3D anatomically mapped expression data (www.brain-map.org). An automated informatics pipeline is being developed to process the image data [2]. The computational challenges are compounded by the scale and throughput of the data, which is based on high-resolution (0.95 micron/pixel) large scale imagery. The algorithmic challenges include inter-modality image registration from ISH to a Nissl reference atlas, detection of the ISH signal and classification of the gene expression patterns.

This paper focuses on the registration component of the pipeline and describes a solution employing ITK (www.itk.org), an open source toolkit consisting of advanced image segmentation and registration algorithms.

2 Automated Anatomy Annotation

One of the essential goals of the ABA project is to enable quantifiable search over expression measurements in anatomic regions. This goal requires the target brain structures to be annotated on each ISH image. There are 20 sagittal or 56 coronal images per gene. As the mouse genome contains around 24,000 genes, achieving this goal requires an efficient, robust, unsupervised, automated process.

Examples of ISH imagery for different genes are shown in Figure 1. Investigation has indicated that it is unlikely a single segmentation technique will be suitable for delineating all the required structures as the boundary features differ significantly from one structure to another. In many cases, the boundaries betweens structures are ill-defined and hence automatic segmentation is a difficult problem requiring some *a prior* information. This might take the form of landmark fiducials or other initializations and in further work we plan to explore refinements using these techniques. As can be seen in Figure 1, the problem is complicated by the primary expression data itself which is regionally variable and of varying intensity.

One approach to this problem is atlas-based segmentation in which manual segmentation is done only once on an atlas image. The regional delineation of a second (subject) image is accomplished by a non-rigid registration of the atlas to the subject image. Registration results in a deformation field which best maps the atlas onto to the subject image. If the registration result is good enough, the deformation field can be used to warp the atlas annotation image thus automatically creating annotations for the subject.

For segmenting ISH imagery, the strategy in [3, 4] was to fit a mesh model using boundaries and landmarks. For our ABA project, we have investigated an hierarchical approach employing inter-modality registration techniques. Our proposed method is described in the next section.

3 Proposed Method

We are developing a high-throughput automated process for annotating ISH gene expression maps. For the first release of the application, we aim to automatically annotate 20 of the largest hierarchical structures in the mouse brain. We begin our description of the process with the annotated template, the Allen Reference Atlas, followed by specifics of the hierarchical inter-modality registration scheme.

3.1 Allen Reference Atlas

The Allen Reference Atlas (ARA) consists of 528, 25 micron thick, coronal Nissl stained sections of an unfixed, frozen mouse brain. In the first version, every 4th section was annotated for 40 of largest hierarchical brain structures on one hemisphere of the brain.

Differences in brain morphology and slicing orientation between the ARA and the specimens used for ISH necessitate a three-dimensional approach to registration. The first step towards 3D registration is to reconstruct a 3D volume from the 2D section image series. Figure 3 shows the result of a "rigid" reconstruction of the reference atlas [5] where each section is rigidly re-orientated to form a self-consistent 3D volume. Further, a low-resolution 3D averaged MRI volume was used as a template to ensure that the reconstruction resulted in a realistic volume. The reconstruction work was performed by our collaborators at the Gee Lab, Department of Radiology, University of Pennsylvania. Finally, a

corresponding dense annotation volume is constructed from the 1:4 annotations by using shape-based interpolation [6] to "fill-in" the un-annotated sections.

3.2 3D Reconstruction and Registration

For efficiency, each brain is used to survey several genes. The standard ABA protocol is to divide the sections into eight interleaved sets. Hence, for each gene, the resulting 2D ISH image series consists of sections which are 200 microns apart. In the present work, we process each gene series independently.

For each gene series, there are two registration problems: (1) 3D volume reconstruction, and (2) subsequent registration of the reference atlas to the reconstructed volume. Perhaps the most difficult challenge in the volume reconstruction problem is the large inter-section spacing; there is insufficient similarity between neighboring sections to align the images without the use of some external a prior constraints. For the 3D atlas to ISH registration task the challenges include:

(a) the relationship between the Nissl reference atlas image intensity and ISH image intensity is complex and different for each gene;
(b) independent section distortions result in jaggedness in the reconstructed volumes; and
(c) potentially large morphological variability between the reference brain and the subject brain primarily due to differential bending before the tissue is frozen.

For either task, robustness against independent section tissue distortions, missing sections and the spectrum of gene expression patterns is also extremely important for success.

Reference atlas to ISH registration is an inter-modality problem. In the literature, the use of mutual information has been popular for solving this type of registration problems [7]. In our preliminary studies on Nissl-ISH registration using mutual information, we observed that success is highly dependent on a good starting position. This is most likely due the complexity of the image intensity relationship which manifests as a cost function with many local maxima.

We determined that preliminary registration is needed to get the alignment close enough before we can apply mutual information. For this purpose, we have selected the tissue to slide background boundary to drive an initial segmentation-based registration process. This boundary was chosen as it is a distinct feature on most ISH images and can be robustly identified.

To simultaneously solve both the registration and reconstruction problems, we have developed a scheme that takes into account many of the limitations and issues discussed above. The scheme interleaves ISH volume reconstruction with atlas to ISH registration. The idea is to use the reference atlas to constrain the volume reconstruction. We start with a coarse reconstruction and registration and iteratively refine both in stages. The various steps are detailed below.

3.3 Algorithm Steps

I. **Coarse ISH reconstruction:** The 2D center of gravity is obtained for each ISH from the tissue mask computed in preprocessing. A coarse 3D ISH volume is constructed by aligning the center of gravity.
II. **Coarse Atlas to ISH registration:** The atlas is linearly registered to the coarse ISH volume by minimizing the mean square difference between the 3D tissue mask of the atlas and the ISH volume.
III. **ISH reconstruction refinement:** This step uses the reference atlas to help improve the ISH volume reconstruction. Using the mapping obtained from the last step, the atlas is resampled into ISH volume coordinates. Each 2D ISH image is then rigidly registered to the corresponding virtual 2D atlas slice mask by minimizing the mean square difference between the 2D tissue masks.
IV. **Deformable registration I:** The alignment is refined by using a low dimensional global deformable registration to warp the 3D atlas mask to the ISH volume mask. For this step, the deformation field is modeled using a B-spline grid with 3mm spacing.
V. **Deformable registration II:** In this step further refinement is obtained by performing regional deformable registration over three anatomic areas (front, center and back of the brain) using a weighted metric: mean square difference of tissue mask and mutual information of the image intensity. Focusing on each region in turn simplifies the underlying joint histogram of the mutual information metric. In this step, a finer 2mm B-spline grid was used to model the deformation.
VI. **Deformable registration III:** Finally, we refine the registration on a per section basis where the registration only considers one ISH section in turn using the same weighted metric as before. This step potentially handles independent section cutting/stretching distortions. In this step the B-spline is further refined to 1.6mm.

The interleaved, iterative refinement approach proposed here has found to be useful and necessary primarily due to the complex inter-modality datasets, high-throughput constraints without user guided input, and comparatively noisy and low sampling density of the data.

3.4 Preliminary Results

We performed an initial experimental run using 20 different genes. Even though the 3D reference atlas is based on coronal images, taking a 3D approach to registration enables both coronal and sagittal ISH images to be processed. A sagittal series takes on average 50 minutes to process while a coronal series requires 3 hours (2.8GHz, 2GRAM, Linux). Typical automatic annotation results are shown in Figure 2. From the results, we observed that the proposed scheme resulted in good gross delineation of approximately 20 of the largest objects in the brain structure hierarchy over various gene expression patterns and differential bending and positioning of the brain tissue.

The most problematic region was found to be the transition between the cortex and mid-brain (green/pink boundary in image (a)). One possible cause is that this region is susceptible to a large amount of differential bending between the specimens requiring more parameters to model the mapping. Further, in the sagittal plane, the two structures are delineated by a thin line. As can be seen in Figure 2(c), this fine-resolution feature has been smeared out in the 3D reconstruction and thus there is insufficient information to lock in the registration.

In further work we plan to investigate algorithmic improvements and conduct a quantitative performance study comparing automated annotation with manual annotation.

4 ITK Implementation

The registration scheme described above was implemented in ITK where registration is performed within a framework of interchangeable components [8]. This flexibility allows a combinatorial variety of registration methods to be easily created allowing the user to mix and match the right tools for a particular application. There are two main advantage of using ITK: (1) many of the popular registration components have already been implemented in ITK and (2) the generic framework allows incremental component-wise development of new techniques.

Steps I,II, and III use methods available in ITK without modification. Steps IV-VI use a modified mean squares metric which is more efficient then the exisiting one when used in combination with a B-spline deformable transform. Additionally a new "weighted metric" class has been developed which behaves like the weighted sum of the underlying metrics and can be used directly within the registration framework. These two new classes with example code accompany this paper in the supplementary materials.

5 Summary

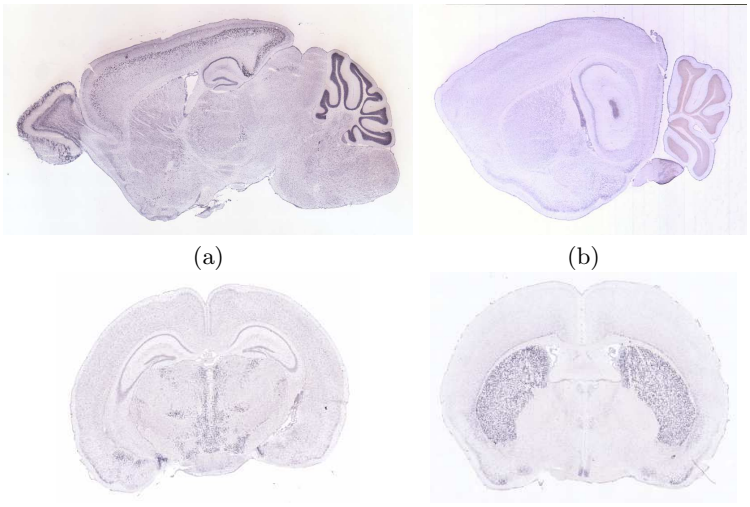
We have developed a high-throughput process for automatically annotate 20 of the largest hierarchical mouse brain structures from ISH imagery. The process relies on inter-modality registration to an annotated Nissl stained reference atlas. A hierarchical registration approach was required to take into account many of the limitations and issues posed by the histological ISH gene expression data. The software solution was developed and implemented in production by using and extending the flexible generic registration framework of ITK.

References

1. Sunkin, S., Boe, A., Brockway, K., Chen, L., Chin, M., Chong, J., Dang, C., Dee, N., Desaki, A., Desta, T., Dong, H., Ebbert, A., Frensley, C., Halverson, K., Hart, M., Hawrylycz, M., Hill, E., Ishihara, M., Johnson, R., Kawal, R., Kidney, J., Knapik,

R., Kuan, L., Larsen, K., Lein, E., Luong, L., Michaels, J., Mosqueda, N., Mortrud, M., Ng, L., Orta, G., Pathak, S., Pearson, O., Puchalski, R., Pak, T., Riley, Z., Rowland, S., Royall, J., Sarno, N., Shapovalova, N., Sherwood, A., Smith, B., Smith, K., Smith, S., Sodt, A., Slaughterbeck, C., Stewart, N., Stumpf, K.R., Sutram, M., Teemer, C., Thaller, C., Thompson, C., Wasell, K., Whitlock, R., Wohnoutka, P., Youngstrom, B., Yuan, X., Jones, A.: High throughput gene expression mapping in the mouse brain. In: The Biology of Genome, Cold Spring Harbor Laboratory (2005)

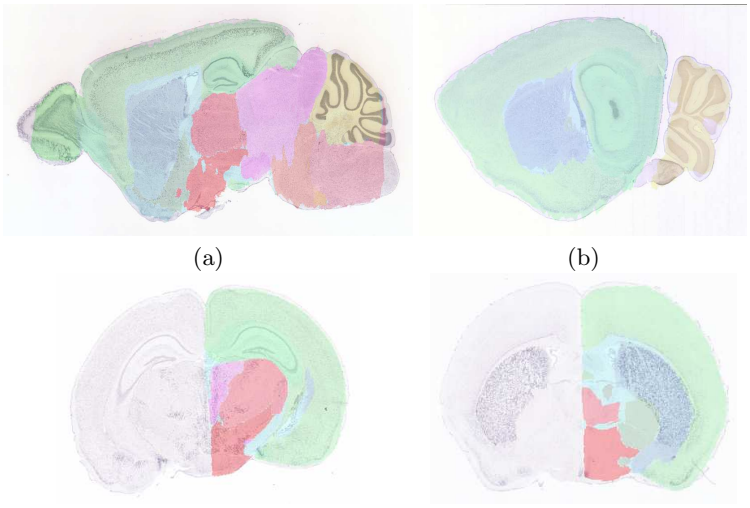
2. Chen, L., Dang, C., Dong, H.W., Hawrylycz, M., Jones, A., Kawal, R., Kuan, L., K.Larsen, Lau, C., Lein, E., Ng, L., Pathak, S., Teemer, C., Thompson, C., Sodt., A., Sunkin, S., Wohnoutka, P.: An automated informatics pipeline for computing 3d gene expression atlases. In: Society for Neuroscience, 35th Annual Meeting, Washington, D.C. (2005)
3. Kakadiaris, I., Bello, M., Arunachalam, S., Kang, W., Ju, T., Warren, J., Carson, J., Chiu, W., Thaller, C., Eichele, G.: Landmark-driven, atlas-based segmentation of mouse brain tissue images containing gene expression data. In: Proc. 7th International Conference on Medical Image Computing and Computer-Assisted Intervention, Rennes, France. (2004) 192–199
4. Ju, T., Warren, J., Eichele, G., Thaller, C., Chiu, W., Carson, J.: A geometric database for gene expression data. In: Proc. Eurographics Symposium on Geometry Processing, Aachen, Germany. (2003) 166–176
5. Yushkevich, P., Avants, B., Zhang, H., Burstein, P., Ng, L., Hawrylycz, M., Gee, J.: Using MRI to build a 3D reference atlas of the mouse brain from histology images. In: Proc. Intl. Soc. Magn. Res. Med. Volume 13. (2005) 2809
6. Raya, S.P., Udupa, J.K.: Shape-based interpolation of multidimensional objects. IEEE Trans. on Medical Imaging 9 (1990) 32–42
7. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. IEEE Trans. on Medical Imaging 22 (2003) 986– 1004
8. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. Kitware, Inc. ISBN 1-930934-10-6, http://www.itk.org/ItkSoftwareGuide.pdf. (2003)



(a) (b)

(c) (d)

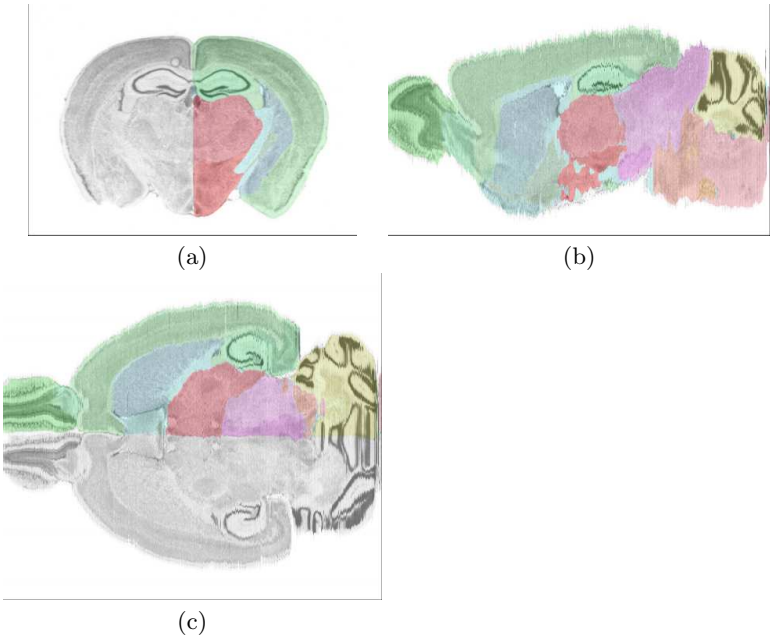
Fig. 1. Examples of ISH images showing different genes, slicing orientation and position in the mouse brain: (a) Etv1, (b) Tmem16b, (c) Calb2 and (d) Drd1a.



(a) (b)

(c) (d)

Fig. 2. ISH images from Figure 1 overlaid with hierarchical brain structure annotation obtained from atlas-based segmentation.



(a) (b)

(c)

Fig. 3. Three views of the 3D reconstructed Nissl volume overlaid with matching hierarchical brain structure annotation: (a) coronal, (b) sagittal and (c) axial