Landmark-driven, Atlas-based Segmentation of Mouse Brain Tissue Images Containing Gene Expression Data

Ioannis A. Kakadiaris[1], Musodiq Bello[1], Shiva Arunachalam[1], Wei Kang[1], Tao Ju[2],
Joe Warren[2], James Carson[3], Wah Chiu[3], Christina Thaller[3], and Gregor Eichele[4]

[1] Visual Computing Lab, Dept. of Computer Science, Univ. of Houston, Houston TX, USA
[2] Dept. of Computer Science, Rice University, Houston TX, USA
[3] Verna and Marrs McLean Dept. of Biochemistry, Baylor College of Medicine,
Houston TX, USA
[4] Max Planck Institute of Experimental Endocrinology, Hannover, Germany

Abstract. To better understand the development and function of the mammalian brain, researchers have begun to systematically collect a large number of gene expression patterns throughout the mouse brain using technology recently developed for this task. Associating specific gene activity with specific functional locations in the brain anatomy results in a greater understanding of the role of the gene's products. To perform such an association for a large amount of data, reliable automated methods that characterize the distribution of gene expression in relation to a standard anatomical model are required. In this paper, we present an anatomical landmark detection method that has been incorporated into an atlas-based segmentation. The addition of this technique significantly increases the accuracy of automated atlas-deformation. The resulting large-scale annotation will help scientists interpret gene expression patterns more rapidly and accurately.

1 Introduction

With the successful efforts in mammalian genome sequencing, biologists are increasingly interested in understanding the function of different gene products at the molecular level. One strategy towards this end is high-throughput *in situ* hybridization (HITISH) [1]. Using HITISH, the spatial distribution of gene expression patterns (where genes are actively transcribed) can be generated throughout an entire organ or organism for thousands of genes each year. By relating a gene's activity to specific functional populations of cells, one gains insight into the function of that gene's products. In addition, genes that express in the same set of cell populations are likely to be related in function. For these reasons, accurately characterizing the location of gene expression in relation to the underlying anatomical morphology is prerequisite to the interpretation of HITISH results. Such a characterization must become both rapid and accurate in order to handle the vast amount of data generated by HITISH. A major computational challenge is to develop efficient algorithms that can accomplish this task with little or no human intervention.

HITISH is currently being applied to both the mouse brain and the mouse embryo [2, 3]. For our algorithm development, we initially focus on the postnatal mouse brain. The postnatal mouse brain is small, accessible, and contains hundreds of functional

subregions, making it a particularly good specimen choice for HITISH. To date, several steps in the computational process of gene expression localization have already been developed: (1) an image filter that identifies the location of cells and the amount of gene expression in each cell [4]; (2) a deformable atlas of the mouse brain anatomy [5] built using subdivision meshes [6] that, when deformed to accurately overlay the anatomy of the HITISH image data collected, segments the major anatomical regions of the image and serves as a standard coordinate system onto which cellular gene expression information is attached; and (3) global deformation of the atlas using affine transformation to account for rotation and translation, and local deformation based on iterated least-squares to account for shape variation [5]. However, the accuracy of the local fitting, and thus the segmentation and interior coordinate system, is limited by its reliance on only tissue boundary detection. Thus, manual deformation of the internal regions of the atlas must still be performed (Fig. 1(a)). In this paper, we present a method that detects anatomical landmarks in tissue images using texture features. This method has been incorporated into the atlas deformation procedure and substantially improves the segmentation and fitting of the atlas' internal boundaries.

The challenges in segmentation are due to the nature of the images. Each individual mouse brain is inherently slightly different from every other one in structure, although the main anatomical features are the same. In spite of the extreme precision employed in carrying out the HITISH process, sectioning induces local distortion upon tissue slices. Also, due to the nature of gene expression patterns, different genes are not expressed equally in different anatomical regions, making it difficult to construct a consistent model of each landmark in all images.

The rest of the paper is organized as follows: Section 2 presents our method in detail including acquisition of the training data for our experiments, locating each landmark in an image, and utilizing the information about the location of the landmark to improve the mesh fitting. Section 3 presents our results and the validation techniques used.

2   Methods

2.1   Experimental Data

In situ hybridization to label cells expressing a gene of interest was performed on 20μm-thick serially sliced cross-sections of the postnatal day 7 (P7) C57BL/6 mouse brain, which were then imaged with a light microscope at 3.3μm per pixel resolution [2]. This resulted in approximately 400 2D images per brain, with each image being 2000 x 3500 pixels in size. For each gene, 11 images were selected that best matched a set of 11 previously defined standard sagittal sections (i.e., levels), to provide a consistent and global representation of the expression patterns. For this work, we selected 100 images of level 9 sagittal sections expressing different genes.[1] We marked the ground truth coordinates for seven selected anatomical landmarks (Fig. 1b) in each of the 100 images. The images were sorted into four categories (Q_1, Q_2, Q_3, Q_4) based on the average signal strength of gene expression in the images.

---

[1] The complete data set is available at www.geneatlas.org and currently consists of 1207 images across 11 sagittal levels.
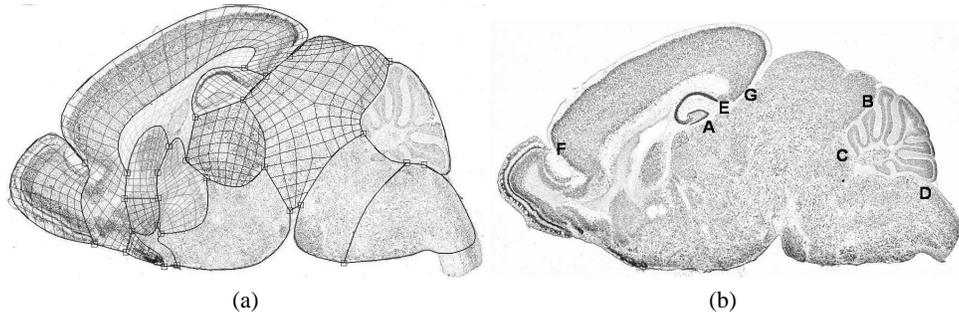
(a)   (b)

Fig. 1. (a) *Sca1* gene expression image overlaid with a subdivision mesh manually fitted by neuroanatomists. (b) A typical mouse brain image showing the expression of the *Camk2g* gene as revealed by *in situ* hybridization with the seven landmarks under consideration marked as follows: A - Dentate gyrus; B - Intersection of midbrain and cerebellum; C - Intersection of pons and cerebellum; D - Intersection of medulla and cerebellum on the outer boundary; E - Tip of the corpus callosum; F - Intersection of the cortex and olfactory bulb on the outer boundary; and G - Intersection of the cortex and the midbrain on the outer boundary.

2.2 Landmark Localization

To locate the desired landmarks in a given gene expression image, first we designed a template for the landmarks and selected their neighborhoods. We then extracted several features before selecting an optimal set of features. Based on a selected training set of images and the selected set of features, classifiers were constructed to distinguish each landmark from its neighborhood. These steps are further elaborated below.

Template Design: We designed a template that sufficiently includes the features of interest in a landmark, and that can be used to distinguish it from its neighbors (Fig. 2(a)). We allowed the rectangular template to have multiple regions so that different textures may be extracted from different areas around the landmark. The dimension (21x11 pixels) was chosen based on examining the area around the landmarks in several images.

Neighborhood Definition: For each landmark, we selected adjacent windows that do not include the landmark and that are different enough to be considered as typical examples of what the landmark is not (i.e., 'out-class'). For example, for landmark A we selected same-sized windows for the out-class, to the immediate right, above, and below the in-class window (Figs. 2(b-i)).

Feature Extraction: To be able to classify each landmark, features based on the statistics of local histograms were extracted as follows:

1. Convolve the original image (I_0) with two Gaussian kernels of width 5 (with $\sigma = 1, 2$) resulting in two smoothed images (I_1, I_2).
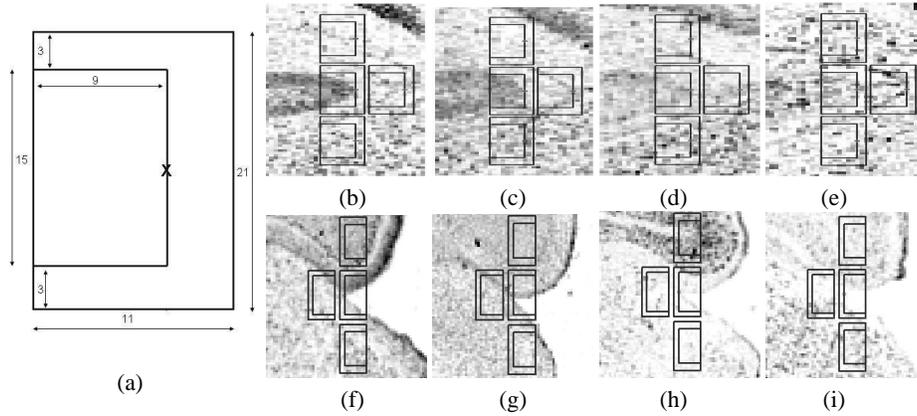
Fig. 2. (a) Landmark template. Selected neighborhood for landmark A in (b) Q_1 (c) Q_2 (d) Q_3, and (e) Q_4 image categories. (f-i) Selected neighborhoods for landmark D.

2. Convolve the three images (I_0, I_1, I_2) with two appropriate edge detector filters (e.g., 45° & 135° for landmark A, 135° & 180° for landmark E). This results in six filtered images. The choice of kernel is guided by the landmark edge orientation.
3. When the template is appropriately placed on the image, the image region under the inner part of the template is divided into overlapping 5x5 windows two pixels apart, and then into overlapping 9x9 windows. The outer part of the template is divided into three windows. The local histogram for each sub-window is obtained and the mean, standard deviation, and skewness are computed. The resulting feature vector with 414 elements (three features in 23 windows in six images) is labeled '1' ('in-class'). Similar features are extracted from the 'out-class' neighbors and labeled '-1'.

Selecting the Optimal Feature Set: It is expected that not all the features extracted will contribute equally to the classification. Using discriminant analysis with F-Test and Wilk's lambda, we ranked the features in order of discriminant power and performed a forward selection process [7]. For example, for landmark A, zero misclassification was obtained consistently with 164 features (out of 414) and these were selected as the optimal feature set.

Selecting the Training Set: Based on methods suggested in [8], we selected an optimal training set of images as follows (let N be the total number of images):

1. Select the first image and extract the optimal features for the in- and out-classes based on the ground truth.
2. Classify the ground truth in the remaining $N-1$ images based on the single-image training set.
3. Repeat for all N images. The image that gives the smallest number of misclassifications when used as training to classify all the other $N-1$ images is selected as the first training image.

4. For the remaining $N - 1$ images, form a new training set by adding each in turn to the first image. The two are used to classify the remaining $N - 2$ images and the pair that gives the least misclassification becomes the best two images.

5. Repeat the steps above adding one image in turn until zero misclassification is achieved (or until misclassification reduces below a threshold).

A classifier is needed to distinguish features of the landmark from that of neighboring regions. We used the k-Nearest Neighbor (KNN) [9] classifier in selecting the best training set because of its ability to classify even with a one-member training set. Figure 3(a) depicts the number of misclassifications as the number of images in the training set increases. The categories with low gene expression signals required more images to be included in the training set to reduce the misclassification to zero. However, in all categories, the misclassification is sufficiently low when half of the images are included. The best result was obtained by combining the optimal training images from categories Q_1, Q_2, and Q_3. Out of 100 images, 54 images formed this optimal set.

Estimating Classifier Parameters: We considered various classifiers including KNN [9], Support Vector Machines (SVM) [10], and Adaboost [11]. The best results were obtained for SVM. The Radial Basis Function (RBF) kernel [12] was used in the SVM and optimal values for the kernel parameters (C and γ) were obtained by cross validation. The training data was divided into four sets, with one set used as training data to classify the remaining three, in turn, using values of C between 2^{-5} and 2^{25} and γ between 2^{-15} and 2^3 as suggested in [13]. The results were sorted in descending order of misclassifications and the C and γ pair that gave the least misclassification across the four sets was chosen as the optimal. For the 24 images in the Q_1 category, the optimal values are $C = 2$ and $\gamma = 0.0078125$ for landmark A.

Locating Landmark Position: Given an approximate location of the desired landmark in the image (obtained after affine and local fitting as described in Sec. 1) and a given radius, a search is conducted in a square region around the approximate location. Rather than extracting features for all the pixels in the search area, features are extracted for every third pixel initially to expedite the process without making the search too coarse. A pixel-by-pixel search is then conducted around the area with the highest decision values (from SVM). This is possible because the decision values were found to be monotonically increasing towards the expected ground truth in all the images tested.

2.3 Integration into Mesh Fitting Process

To incorporate the knowledge of the landmarks into the mesh fitting framework [5], we set up a quadratic energy function $E^k(x)$ of the form: $E^k(x) = E_f^k(x) + E_d^k(x)$, where $E_f^k(x)$ measures the fit of the mesh at subdivision level k to the image and $E_d^k(x)$ measures the energy used in deforming the mesh. The fitting term $E_f^k(x)$ is formulated as: $E_f^k(x) = \alpha B^k(x) + \beta L^k(x)$, where $B^k(x)$ is the fitting error of the boundary of the mesh to the outer boundary of the image. $L^k(x)$ is the fitting error of the mesh to the landmarks defined as: $\sum_j (l_j - p_j(x))^2$, where $p_j(x)$ is the vertex of the mesh associated

with landmark l_j. The formulation of $B^k(x)$ and the deformation energy term $E_d^k(x)$ are the same as in [5]. The minimizer of $E^k(x)$ is then computed using a linear solver such as conjugate gradients.

3 Results

After incorporating the feature-driven landmark localization algorithm into the mesh-fitting framework, the fitting of internal region boundaries was substantially improved. Figures 3(b,c) depict the performance for different image categories. For each landmark, the distance to the ground truth before and after incorporating the feature-driven algorithm was compared (Fig. 4). It was observed that the distance from the ground truth is reduced in 89%, 68%, and 76% of the images for landmarks A, E, and F respectively. In cases where the initial fitting is very far from the ground truth, the result of fitting is similarly poor.
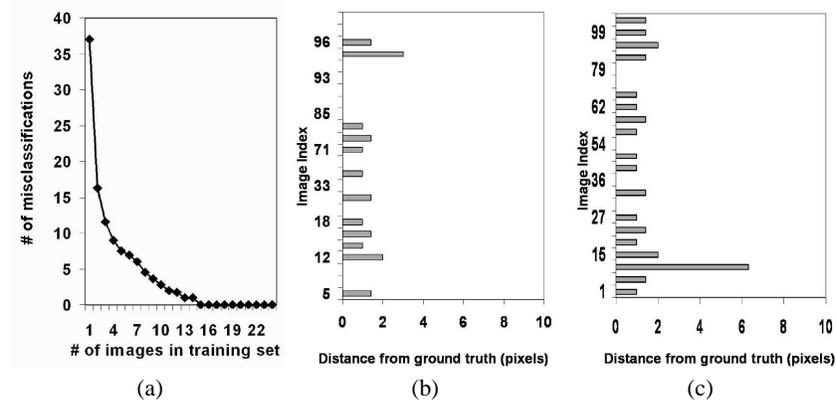
Fig. 3. (a) Number of misclassifications of landmark A and its neighbors as the number of training images increases for image category Q_1. Localization error for landmark A in image categories (b) Q_1, and (c) Q_2.

Different landmarks are more accurately detected in some images than others. As summarized in Table 1, at least three landmarks were found within a 4x4 neighborhood of the ground truth in 87% of the images. In 58% of the images, at least four landmarks were detected within a 3x3 neighborhood of the actual ground truth, and the number increases to 76% in a 5x5 neighborhood. Considering the inter-observer variability in locating the ground truth for the landmarks, detection errors within a 3x3 neighborhood are deemed acceptable. Figure 5 illustrates typical improvements after incorporating the feature-driven algorithm.
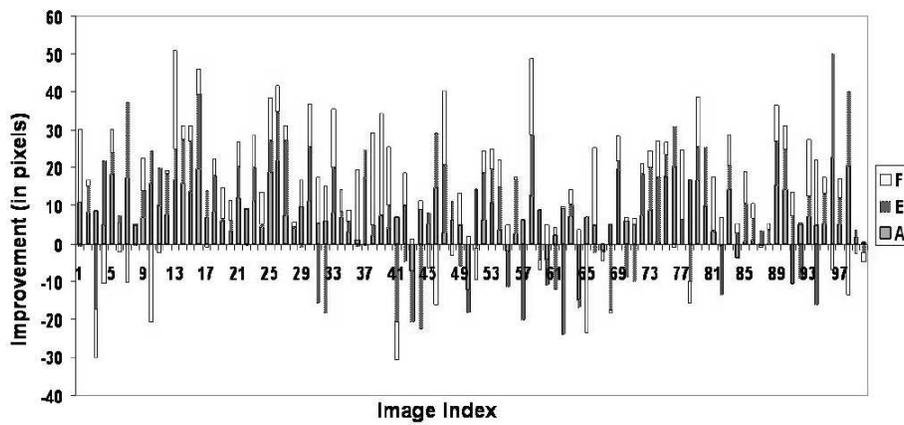
Fig. 4. Difference in distance to the ground truth in landmarks A, E and F after incorporating the feature-driven algorithm. A positive value indicates improvement. The cases where no improvement is observed are mainly due to the fact that the initial fitting would place the search window too far from the location of the landmark.

Table 1. Summary of results when fitting all the landmarks simultaneously.

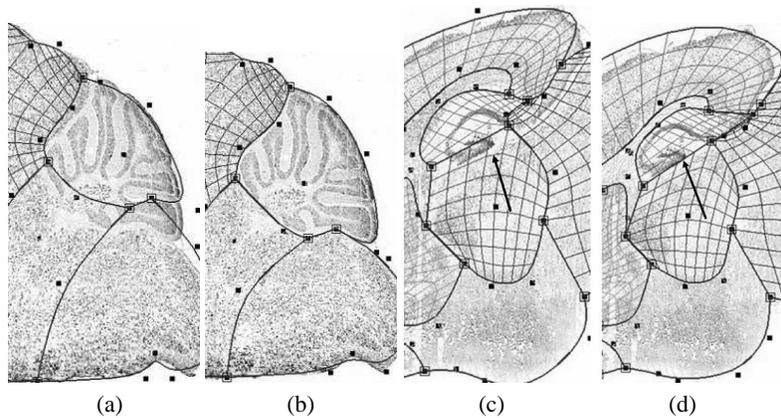| No. of Landmarks	1x1 window	2x2 window	3x3 window	4x4 window	5x5 window
1	93	100	100	100	100
2	62	91	94	96	97
3	27	59	78	87	89
4	9	33	58	68	76
5	1	10	28	47	62
6	0	3	9	16	25
7	0	0	0	0	3

(a)   (b)   (c)   (d)

Fig. 5. The cerebellum area of the *Ntrk3* gene expression image overlaid with the mesh fitted (a) without, and (b) with the feature-driven algorithm. The area around the dentate gyrus for the *Calb1* image fitted (c) without, and (d) with our algorithm.

4 Conclusion

In fitting a deformable atlas to mouse brain images, the accuracy is greatly improved when internal region boundaries are used for aiding the fitting process. In biological images such as sagittal sections of the mouse brain, there is a lot of variation in shape and appearance making it difficult to use traditional methods (e.g., snakes, active shape and appearance models) for segmentation of internal region boundaries. We have selected a feature-based approach that extracts information about the texture of specific landmarks (distinguishing anatomical features) in the image and locates those landmarks in any new image. The locations of the detected landmarks are then incorporated into an energy-minimization framework to fit the standard mesh to the given image. Incorporating the internal landmarks has substantially improved the segmentation results.

References

1. Herzig, U., Cadenas, C., Sieckmann, F., Sierralta, W., Thaller, C., Visel, A., Eichele, G.: Development of high-throughput tools to unravel the complexity of gene expression patterns in the mammalian brain. In Bock, G., Goode, J., eds.: Complexity in Biological Information Processing. Novartis Foundation Symposium 239. John Wiley & Sons, Chicester (2001) 129–149
2. Carson, J., Thaller, C., Eichele, G.: A transcriptome atlas of the mouse brain at cellular resolution. Current Opinion in Neurobiology **12** (2002) 562–5
3. Visel, A., Thaller, C., Eichele, G.: Genepaint.org: an atlas of gene expression patterns in the mouse embryo. Nucleic Acids Research **32** (2004) D552–556
4. Carson, J.: Quantitative annotation and analysis of gene expression patterns with an atlas of the mouse brain. PhD thesis, Baylor College of Medicine (2004)
5. Ju, T., Warren, J., Eichele, G., Thaller, C., Chiu, W., Carson, J.: A geometric database for gene expression data. In Kobbelt, L., Schröder, P., Hoppe, H., eds.: Eurographics Symposium on Geometry Processing, Aachen, Germany (2003) 166–176
6. Warren, J., Weimer, H.: Subdivision Methods for Geometric Design: A Constructive Approach. Morgan Kaufmann Publishers, San Francisco, CA (2002)
7. Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 4–37
8. Plutowski, M., White, H.: Selecting concise training sets from clean data. IEEE Transactions on Neural Networks **4** (1993) 305–318
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons Inc. (2001)
10. Vapnik, V.N.: The Nature of Statistical Learning. Springer (1998)
11. Freund, Y., Schapire, R.E.: An introduction to boosting. Journal of Japanese Society for Artificial Intelligence **14** (1999) 771–780
12. Schölkopf, B., Smola, A.: Learning with Kernels. MIT Press, Cambridge, MA (2002)
13. Hsu, C., Chang, C., Lin, C.: A practical guide to support vector classification. Technical report, National Taiwan University, Taipei 106, Taiwan (2003)