PDM-ENLOR: Learning Ensemble of Local PDM-based Regressions

Y.H. Le, U. Kurkure and I.A. Kakadiaris

Computational Biomedicine Lab, University of Houston, Houston, TX, USA

http://www.cbl.uh.edu/

Abstract

Statistical shape models, such as Active Shape Models (ASMs), suffer from their inability to represent a large range of variations of a complex shape and to account for the large errors in detection of model points. We propose a novel method (dubbed PDM-ENLOR) that overcomes these limitations by locating each shape model point individually using an ensemble of local regression models and appearance cues from selected model points. Our method first detects a set of reference points which were selected based on their saliency during training. For each model point, an ensemble of regressors is built. From the locations of the detected reference points, each regressor infers a candidate location for that model point using local geometric constraints, encoded by a point distribution model (PDM). The final location of that point is determined as a weighted linear combination, whose coefficients are learnt from the training data, of candidates proposed from its ensemble's component regressors. We use different subsets of reference points as explanatory variables for the component regressors to provide varying degrees of locality for the models in each ensemble. This helps our ensemble model to capture a larger range of shape variations as compared to a single PDM. We demonstrate the advantages of our method on the challenging problem of segmenting gene expression images of mouse brain.

1. Introduction

Statistical shape models are widely used for automated image segmentation ([9, 11, 10, 31, 1, 2, 29]) and landmark localization ([32, 25]). They overcome some of the limitations of the deformable shape models by imposing geometric constraints that are learned from a training dataset. The active shape model (ASM [7]) is one of the most popular statistical shape models that restricts the shape space to limit the range of possible shapes the model can form. However, one of their major limitations lies in their ability to represent the variations of a complex shape model, especially when the number of training samples is much smaller

than the dimensions of the shape model. Thus, they provide a restricted subspace of shapes that can be insufficient to represent the shape variations. Moreover, it is difficult to account for the large errors in the detection of the model points due to the noisy, incomplete or complex appearance information in the image. Such erroneous detections of a large number of model points can drive the fitting to an incorrect solution.

In this paper, we propose a new approach for statistical model fitting that also provides solutions to the problems of the model flexibility and the model point detection errors. Our PDM-ENLOR locates each shape model point individually using an ensemble of regression models built for that specific point. Specifically, a set of salient reference points are first selected to be used as explanatory variables of the regression models. These reference points are detected using our PASM-CTX algorithm. Then, each component regression model regresses the location of a model point of interest from the detected locations of its explanatory variables by fitting a point distribution model (PDM) [8], which is built to encode the spatial relationship between the dependent and the explanatory variables. In order to provide increased flexibility to the shape model and to handle the non-robust detection of the regression explanatory variables, the models are built with increasing degrees of locality based on the increasing number of reference points used. The final location of the point is obtained by combining the output of all models using non-negative coefficients determined from the training set. Note that the set of selected salient reference points is automatically selected during the training phase and may include points that do not belong to the boundaries. We evaluate our method on mouse brain gene expression images to segment sagittal sections from a mouse brain into 14 anatomical regions. The main challenges of this problem are the lack of visible edge cues of the regional boundaries and the shape variation of anatomical regions across images [16].

The contributions of our work are as follows. First, we propose a novel method to construct an ensemble of multiple regression models to impose shape constraints of varying degrees of locality from local-to-global to increase the

flexibility of the shape model. Our method automatically constructs an ensemble of local-to-global PDMs and therefore, there is no need to determine the suitable degree of locality. Second, we propose a method to learn the optimal coefficients for combining regressors in an ensemble using linear least squares with non-negative constraints. This data-driven approach is generic and can be applied to similar problems. Third, our PASM-CTX method to detect the reference points is the first work that uses similarity-based features instead of a local gray level model to detect the local best matches in ASM search. This modification makes ASM applicable to gene expression image data whose regional boundaries are indistinct. In addition, since our method incorporates appearance guidance from only the points that are likely to be detected correctly, our method does not require any post-processing step to minimize the errors in fitting due to unreliable model points.

The rest of the paper is organized as follows. In Sec. 2, we discuss the related work in literature. In Sec. 3, we present our method in detail. In Sec. 4, we evaluate the proposed method on 2D biological image segmentation and discuss the results. Finally, we conclude our work in Sec. 5.

2. Related work

Many efforts have focused on improving the flexibility of the point distribution models. One direction is to allow deviations that are not explained in the training data [9, 22], but it can result in an invalid shape. Another direction is to partition the shape model into local sub-shapes models and fit them individually [11, 10, 31, 24, 3]. However, determining an appropriate partition of the shape model is a non-trivial problem. Additionally, the appearance prior in a local region may not be strong enough to drive the fitting of a sub-shape model to any acceptable solution. Instead of partitioning the shapes, few recent methods [2, 18] fit a global shape model for each model point individually using local weights. The local weights for each model point control the neighborhood size for fitting that point and they are computed based on the distances between model points. While this approach avoids the partitioning of the shape model, determining the neighborhood size that controls the degree of locality is nontrivial. Moreover, it is very sensitive to the incorrect detection of the neighboring points.

One of the most common approaches to account for the non-Gaussian errors in the detection of points is to weigh each point based on the reconstruction residuals [13, 27, 1]. The performance of these methods degrades rapidly with the increase in the number of incorrectly detected points and the extent of error in detection. Lekadir *et al.* [21] proposed to identify the outlier points (and correct them) with incorrect detection using a local shape dissimilarity measure. Similar to other previous methods, this method also identifies the detection errors as a post processing step after

detection of all feature points. This post processing can be avoided by explicitly enforcing sparseness condition on the residuals [30] but only if the number of non-robust points is small. Yan *et al.* [29] proposed to detect the salient points based on prior knowledge about the contrast of the contour and reconstruct the full shape from the detection of salient points.

Our methods are closely related to the work of Amberg *et al.* [2] and of Yan *et al.* [29]. Similar to [2], PDM-ENLOR also locates each model point independently. However, instead of using a single global model, our method explicitly builds different regression models with different degrees of locality for each point to increase their flexibility. Moreover, our method automatically determines the suitable levels of local shape constraints for each point separately. Similar to [29], the PASM-CTX and PDM-ENLOR detect the salient points and reconstructs the shape based on the guidance of the salient points to account for the large errors in detection. However, our methods learn the set of salient points from training data in advance and exploits information from additional supporting salient points, which may not belong to the boundaries. In addition, PDM-ENLOR uses salient points selectively in the ensemble of multiple models to provide further flexibility at local level. Our work and [16, 20] use data from the same application domain. However, our approach explicitly incorporates geometric prior in the form of a statistical shape.

3. Method

3.1. Overview

In this section, we briefly present our method for fitting a shape model to an image. The shape is explicitly represented by an ordered set of points. A set of reference points, which were selected during training phase based on a saliency criteria, are detected using PASM-CTX. Then, each point of the boundary shape model is localized independently using an ensemble of regression models. Each regression model is obtained by fitting a PDM, which is specifically built to represent the spatial relationship of the model point of interest and a subset of the reference points. The final location of the model point is determined as a non-negative linear combination of the candidates proposed by the regression models for that point. The non-negative coefficients are learned from the training data. Specifically, the final location of a model point $\mathbf{p}_i$ in the shape model of interest, is given by:

$$\mathbf{p}_i = \sum_{j=1}^{k} c_j^i f_j^i(R_j^i), \qquad (1)$$

where k is the number of regression functions built for $\mathbf{p}_i$, R_j^i is a set of reference points used in the function f_j^i to infer $\mathbf{p}_i$ and c_j^i is the ensemble coefficient for the regression

function f_j^i. Note that, multiple shape models are used in our method. The shape model of interest that contains sampled points on the boundaries of the object is referred as the *boundary shape model*. When the set of reference points is not a subset of the set boundary points, the shape model that contains all the boundary points and the reference points is called the *extended shape model* while a shape model that contains a subset of points of the extended shape model is called the *local extended shape model* (Fig. 1).

3.2. Reference point selection

Only the points that can be reliably detected should be used as a reference to guide the inference for the location of the model points. The reliable points are likely to have high consistency in appearance across images. The mutual-saliency concept of two arbitrary points in two images, introduced by Ou *et al.* [26], was modified to similarity-saliency concept [20] to evaluate the degree of saliency in appearance for the corresponding anatomical point in multiple images. We adopt the similarity-saliency to select the reference point set. The similarity-saliency score of a point $\mathbf{u}$ with respect to a reference image T computed for a set of training images $\mathcal{I} = \{I_i\}_{i=1}^n$ is defined as:

$$\gamma(\mathbf{u}, T, \mathcal{I}) = \frac{\sum_{i=1}^n \gamma(\mathbf{u}, T, I_i)}{n} \qquad (2)$$

and for a single image I, it is defined as:

$$\gamma(\mathbf{u}, T, I) = \frac{\frac{1}{|\mathcal{N}_C(\mathbf{u}, I)|} \sum_{\mathbf{v} \in \mathcal{N}_C(\mathbf{u}, I)} \psi(T(\mathbf{u}), I(\mathbf{v}))}{\frac{1}{|\mathcal{N}_P(\mathbf{u}, I)|} \sum_{\mathbf{v} \in \mathcal{N}_P(\mathbf{u}, I)} \psi(T(\mathbf{u}), I(\mathbf{v}))}, \qquad (3)$$

where the terms $\mathcal{N}_C$ and $\mathcal{N}_P$ denote the core neighborhood region (within a small radius around $\mathbf{u}$) and the peripheral neighborhood region (an annulus at larger distance from $\mathbf{u}$) in image I, and $\psi(T(\mathbf{u}), I(\mathbf{v}))$ is the similarity metric computed for two image patches (centered at $\mathbf{u}$ in image T and centered at $\mathbf{v}$ in image I). The similarity-saliency score of a salient point should be high (the higher score, the more salient) and greater than 1.0 (i.e., the average similarity in $\mathcal{N}_C$ is higher than that in $\mathcal{N}_P$).

Let L be the set of points associated with an object in an image. The position of each point in L in every training image is known (e.g., they are obtained using an annotation process). A set of salient points are selected using the following steps. First, the similarity-saliency score of each point with respect to the reference image T over the set of training images $\mathcal{I}$ (Eq. 2) is computed. Then, a set of reference points whose similarity-saliency score is higher than a threshold t are selected. The threshold t is determined by using the histogram of the similarity-saliency scores higher or equal 1.0. Specifically, the threshold $t > 1.0$ is selected such that it separates the points in the first bin from the other points. We use Sturge's method [28] to estimate the

number of bins for computing histogram of ϕ samples as: $1 + 3.3 log_{10}(\phi)$. The motivation for removing the first bin is to obtain more reliable points. However, the similarity-saliency is an unnormalized score (the ratio of two similarity scores). Therefore, further investigation is needed to determine suitable techniques to determine the threshold for selecting the reference points.

Mouse brain gene expression images: For mouse brain images, L contains all 1,245 vertices of a subdivision mesh, a geometric model specifically constructed for mouse brain gene expression images by Ju *et al.* [15], at subdivision level 2. The boundary shape model contains all 394 vertices on the boundaries. For the 10-fold experiments, 183 to 196 reference points were selected and depicted as solid blue circles in Fig. 1. In the mouse brain gene expression images, the intensity pattern of each anatomical region may vary significantly from image to image as each image expresses a different gene. Therefore, a special image called Nissl-stained image (NSI), which was constructed using a universal gene probe and has maximum similarity to other gene expression images, is used as a reference image ([16, 20]).

3.3. Regression model definition

In this section, we present how to encode different degrees of locality in our ensemble scheme by defining the explanatory variables R_j^i for regressors in Eq.1. The locality level j of a regression model $f_j^i(R_j^i)$ is based on the spatial relationship between $\mathbf{p}_i$ and the reference points in R_j^i. We assume that the neighboring points provide simi-
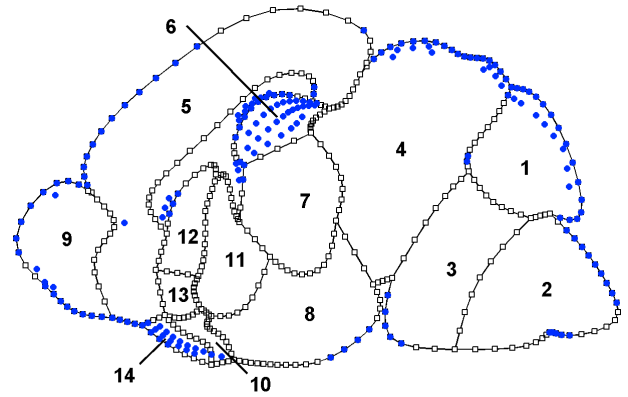

Figure 1: Illustration of the shape models in mouse brain gene expression image segmentation. The squares depict the non-reference boundary points and the solid blue circles depict the reference points. The boundary shape model contains all sampled points on the regional boundaries of 14 anatomical regions. The extended shape model contains all the boundary points and the reference points (i.e., it contains all the points shown in this figure).
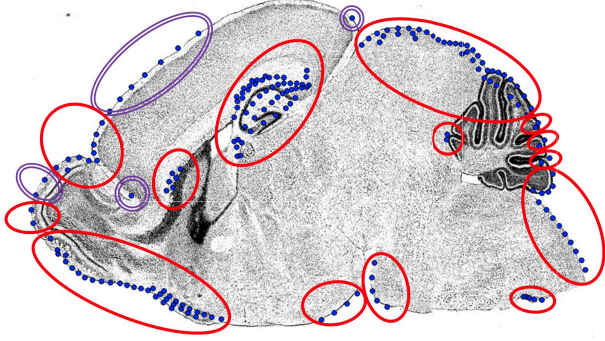
Figure 2: Example of the clusters obtained during training (10-fold cross validation). The clusters depicted by single-line ellipses are obtained from the first step of clustering. The clusters depicted by two ellipses are obtained from the second step of clustering.
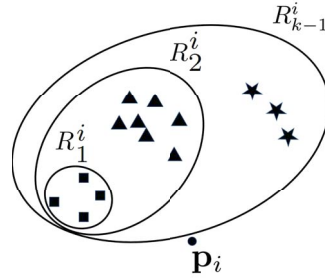

Figure 3: Illustration of the explanatory variables for regression functions. Points denoted by the same symbol belong to the same cluster. Multiple f_j^i models infer $\mathbf{p}_i$ from different sets R_j^i using corresponding local to global extended shape models.

lar geometric constraints to infer a model point. Therefore, we perform density-based clustering of the reference points. These clusters of the reference points are then used to construct the regression models for inference of the position of a target model point.

Clustering of the reference points: This step is performed during the training phase. The clustering should let the neighboring points belong to the same cluster. Therefore, in our experiments, we use the algorithm DBSCAN [12]. This algorithm has two parameters η (the number of neighbors) and ϵ (the neighborhood radius). The points that have less than η neighbors in their ϵ -neighborhood are the outliers. To cluster q points in d-dimensional space, we use the analytical estimation for the neighborhood radius ϵ ([5]):

$$\epsilon = \sqrt[d]{V \frac{\eta \Gamma(d/2 + 1)}{q\sqrt{\pi^d}}}, \qquad (4)$$

where V is the volume of the smallest cuboid that contains all the points that need to be clustered and Γ is the well-known Gamma function. Because there can be some reference points that are isolated from the others and should not be merged with their neighbors that are far away, we allow the clusters to have size of 1. However, we also want to group the points that are dense neighbors to reduce the computation cost. Therefore, the clustering is performed in two steps. First, we set $\eta = 1$ so the clustering requires that each cluster has at least two points. Then, the outliers from the first step ares clustered again with $\eta = 0$ allowing 1-point clusters. Therefore, every reference point belongs to one of the clusters. The output of this step is $k - 1$ non-overlapping clusters $Q_1, Q_2, ..., Q_{k-1}$. Figure 2 illustrates the clusters obtained by our method.

Construction of the regression models:

Let $h(i, j)$ be the index of the cluster that is the jth closest to $\mathbf{p}_i$ (based on the distance from the center of each clus-

ter to $\mathbf{p}_i$). We build $\{R_j^i\}_{j=1}^k$ in a nested-manner (Fig. 3): $R_1^i = Q_{h(i,1)}$; $R_k^i = R_{k-1}^i \cup Q_{h(i,k)}$, for $k : 2 \to (k - 1)$; and $R_k^i = R_{k-1}^i$. The smaller value of the locality level j of the regression model f_j^i corresponds to a more local model. Note that, R_k^i and R_{k-1}^i are the set of all reference points.

3.4. Shape model point regression

Given an input image, our method first detects the reference points and then infers the position of each model point using the constructed regression models. For clarity of presentation, first we present how to determine candidate locations using a PDM-based regression function assuming that the positions of the reference points are already available. Later, we present the method to detect the reference points.

3.4.1 PDM-based model point regression

In this section, we present how our method infers the position of a model point $\mathbf{p}_i$ based on reference points in R_j^i using a PDM. In PDMs, each shape is represented as the concatenation of coordinates of all the points $\mathbf{x} = [x_1, y_1, x_2, y_2, ..., x_m, y_m]^T$, where m is the number of model points. Given n training shapes, assume that they were rigidly aligned using Procrustes analysis [14], denote $\bar{\mathbf{x}}$ as the mean shape and $\mathbf{X}_{n \times 2m}$ as the data matrix where each row corresponds to one shape. Principal component analysis is applied on the covariance matrix of $\mathbf{X}$ to transform the aligned shapes into a new space, $\mathbf{P}$, whose axes are defined by the eigenvectors that are corresponding to the highest eigenvalues. Then, each shape $\mathbf{x}$ is represented as $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$, where $\mathbf{b}$ is the shape parameter. The fitting of the PDM $(\bar{\mathbf{x}}, \mathbf{P})$ to a shape $\mathbf{x}$ is given by:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} ||\mathbf{W}[\mathbf{x} - (\bar{\mathbf{x}} + \mathbf{P}\mathbf{b})]||_2^2, \qquad (5)$$

where the diagonal weight matrix $\mathbf{W}_{2m \times 2m}$ is introduced to emphasize the importance of the model points: $\mathbf{W}(2i - 1, 2i - 1) = \mathbf{W}(2i, 2i)$ is the weight of the ith model point. In standard PDM fitting, the weights of all model points are uniform. Generally, the shape is rigidly-aligned before fitting to remove global transformations by using generalized Procrustes [14].

For each of the first $k-1$ regression models, $f_j^i(R_j^i)$, $(j = 1, ..., k-1)$, a PDM $(\bar{\mathbf{x}}_j^i, \mathbf{P}_j^i)$ is learnt from training shapes where each shape contains $\mathbf{p}_i$ (always being the first point of the shape) and all points in R_j^i. To use a PDM to infer the position of an unknown point $\mathbf{p}_i$, we reconstruct the full shape from the known locations of points in R_j^i and retrieve the point of interest $\mathbf{p}_i$. Specifically,

$$\mathbf{p}_i = f_j^i(R_j^i) = [\mathbf{a}_j^{i*}(1), \mathbf{a}_j^{i*}(2)]^T, \qquad (6)$$

where $\mathbf{a}_j^{i*}$ is solution of Eq. 5 when using the model $(\bar{\mathbf{x}}_j^i, \mathbf{P}_j^i)$ and weight matrix $\mathbf{W}_j^i$ with weight equal to 0 for $\mathbf{p}_i$ and weight equal to 1 for the other points.

The inference of $\mathbf{p}_i$ in the first $(k-1)$ regression models employs only the geometric constraints between $\mathbf{p}_i$ and the reference points. To explicitly impose the geometric constraints on all the points in the boundary shape model, the extended shape model is used for training the PDM for the last regressor $f_k^i(R_k^i)$.

3.4.2   Reference point detection

ASM is a strategy for searching a shape that fits to the image evidence in an input image by using a PDM that was learnt from the training shapes. In the ASM approach, a shape iteratively evolves in two steps: (i) finding a new shape estimate (target shape) whose each model point is detected as the best match in the local neighborhood of that point from the previous iteration, and (ii) fitting the target shape by solving Eq. 5. Due to the indistinct anatomical boundaries in gene expression images, the local gray level model of the traditional ASM formulation is not suitable for the detection of the model points. Therefore, we introduce a similarity-based feature as a new feature detector for ASM to detect the reference points.

While we want to detect only the reference points, the extended shape model is used to train a PDM for maintaining the global shape constraint. Since the extended shape model contains context points which are not the points of interest, we refer our modified ASM method as PASM-CTX (Partial ASM with ConTeXt). At each iteration, we evolve the shape based on the guidance of the reference points only (i.e., 0 as weight for non-reference points and 1 as weight for reference points in the fitting process). At the final iteration, we trust the feature detector than the shape model to relax the geometric constraint. That means the fitting step at the last iteration is omitted and the best matches obtained from feature detectors at that iteration are the final estimates of the reference points.

Initialization: To obtain a robust initialization for the specific application, the shape parameters are first computed based on the points on the outer boundary of the brain (i.e., other model points have zero weights). The reason is

that the outer boundary in mouse brain can be easily obtained quite accurately and it can provide certain information about the global shape [15, 4, 17, 20]. Specifically, after binarizing the image, the contour of the outer shape is traced, followed by sampling the contour to obtain a set $\mathcal{C}$ of discrete points. At each iteration of ASM search, the target of an outer-boundary model point is the point in $\mathcal{C}$ closest to it.

Model point detector for ASM: For each model point, a local image patch around that point in the NSI is used as the 2D appearance model for detection of the points. The reason for using the NSI as a reference for appearance comparison has been discussed in Sec. 3.2. At each iteration, a local best match of $\mathbf{p}_i$ is selected from one of the neighbors that are sampled from a search window centered at that point. The neighboring point whose image patch around it in the test image is most similar to the image patch around $\mathbf{p}_i$ in the reference image (NSI) is selected as the best match for $\mathbf{p}_i$. The similarity between patches is evaluated using the similarity function $\psi(.,.)$ that was used in Eq. 3. A multi-resolution approach for sampling the neighbors is employed. The parameters were selected so that at the final step, the sampled neighbors include all the pixels in the search region.

3.5. Combining the models in ensemble

In this section, we provide the motivation for using an ensemble of models in Eq. 6 as well as the motivation for learning the ensemble weights. Finally, we present our method to learn the ensemble weights.

Using an ensemble of models: A local model can provide improved geometric constraints due to the locality and the simplicity of the local shape. However, it can suffer from non-robust detection of the local reference points. Using a global model in this case can help if the additional reference points can be detected more accurately. We assume that the true point can be estimated as a non-negative linear combination of candidate points.

Learning the ensemble weights: Two popular solutions to combine the models in ensembles are either to use the mean position of the candidates or to use the peak of distribution of the candidates. To select an appropriate voting rule, prior knowledge of the distribution of the candidates is needed. The former solution is preferable if the candidates are evenly scattered. The latter solution is preferable if the majority of the candidates are close to the true solution. However, density estimation is sensitive to parameter settings and different model points may need different parameter sets. Whereas, learning ensemble weights from data is a data-driven solution that can be generalized for similar applications. Therefore, we propose to learn the weight coefficients from the training data. This approach, however, depends on the assumption that the behavior of the multiple

regression functions for each model point is similar across images. An extreme example under this assumption is when some points always prefer the global model whereas some others always prefer the most local models. Therefore, we can use a universal coefficient set for each point in all images.

Learning the ensemble weights: For each model point $\mathbf{p}_i$, the coefficient vector $\mathbf{c}^i = [c_1^i, c_2^i, ..., c_k^i]$ is obtained by minimizing the error of the final regression results computed for $\mathbf{p}_i$ when compared to the true position over all the training images. The matrix $\mathbf{A}^i$ contains the coordinates of the results obtained from the regression models that infer $\mathbf{p}_i$ in all training images:

$$\mathbf{A}^i = \begin{bmatrix} x_i^{*(1,1)} & x_i^{*(2,1)} & \dots & x_i^{*(k,1)} \\ y_i^{*(1,1)} & y_i^{*(2,1)} & \dots & y_i^{*(k,1)} \\ \hline . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ \hline x_i^{*(1,n)} & x_i^{*(2,n)} & \dots & x_i^{*(k,n)} \\ y_i^{*(1,n)} & y_i^{*(2,n)} & \dots & y_i^{*(k,n)} \end{bmatrix}$$

where $(x_i^{*(j,l)}, y_i^{*(j,l)})$ is the output of f_j^i on training image I_l. The vector

$$\mathbf{g}^i = \left[x_i^{(g,1)}, y_i^{(g,1)}, x_i^{(g,2)}, y_i^{(g,2)}, \dots, x_i^{(g,n)}, y_i^{(g,n)} \right]^T$$

contains the coordinates of the annotation position $(x_i^{(g,l)}, y_i^{(g,l)})$ of $\mathbf{p}_i$ on training image I_l. Then, the ensemble coefficient vector for point $\mathbf{p}_i$ is computed as:

$$\mathbf{c}^{i*} = \underset{\mathbf{c}^i}{\operatorname{argmin}} ||\mathbf{A}^i \mathbf{c}^i - \mathbf{g}^i||_2^2, \text{ s.t. } c_j^i \geq 0 \qquad (7)$$

We solve Eq. 7 by using Lawson and Hanson's method for least squares with non-negative constraints [19].

4. Experiments and Results

Image data: We evaluated our method on 2D mouse brain gene expression images [4, 6]. The dataset contains 100 images depicting sagittal sections of postnatal day 7 mouse brains at standard section 9. Each of the images is rigidly-aligned to the template image (NSI). The annotated shape and the anatomical point set L were extracted from the manually annotated subdivision mesh at subdivision level 2 and were provided by [15, 4].

Experimental setup: We performed 10-fold cross-validation. We quantitatively compared the performance of the different methods using the Dice similarity coefficient (DSC) against the manual annotations for each of the 14 anatomical regions in the mouse brain. The statistics (mean, confidence-interval and statistical test) of the DSCs were computed on 100 resulting images from all 10 experiments. Although different models, built using different training data sets, were used to obtain the results, the evaluation is fair because we apply the same automatic framework (for training/testing) for all the experiments. To test if a method A outperforms a method B, we performed a non-parametric Wilcoxon Signed-Rank test using the DSC values in order to determine whether the relative median values are equal (i.e., null hypothesis) or A outperforms B (i.e., one sided alternative hypothesis). The Bonferonni correction was used to keep the overall significance level equal to 0.05. All the tests that reject the null hypothesis in our experiments have p-values smaller than $0.0015 < \frac{0.05}{14}$. We do not include in the text the comparison between two methods on a region if none of them outperforms the other.

For each PDM, we retained components that contributed at least 2% of the total variance. Thus, the number of components kept is in range 1-6, depending on the simplicity of the local shapes. The patch size to compute the similarity function is 41×41. The core neighborhood was defined within a radius of 7 pixels and the peripheral neighborhood annulus was defined by radii of 10 and 15 pixels. The sampled contour $\mathcal{C}$, used in the robust initialization part, contained 800 points. At each iteration of ASM, we sampled 11×11 neighbors. The search box size is initialized as 61×61 and is decreased 10% after each iteration till it reaches 11×11, after that it is fixed for further iterations. We used the *normalized mutual information* (NMI) [23] as the similarity metric. Same parameter settings were used in all the experiments in the ASM-based methods (Fig. 4).

Comparison with ASM-based approaches: We compared our proposed PDM-ENLOR with three ASM-based methods: ASM-SIM, Amberg-SIM and PASM-CTX. ASM-SIM and Amberg-SIM are modified versions of ASM and of Amberg *et al.* [2], respectively. Both methods were applied to the boundary shape model and used similarity-based feature for detecting local best matches in ASM search. Global fitting with uniform weights was used for ASM-SIM whereas local fitting [2] was used for Amberg-SIM. In PASM-CTX, partial ASM was used for the extended shape model. Thus, the reference points detected in Sec. 3.4.2 are obtained before global fitting step at the last iteration of PASM-CTX.

For Amberg-SIM, we ran multiple experiments with different values for parameter σ, which controls the locality degree of the fitting, and reported the best results (Fig. 4). Smaller value for σ corresponds to more local fitting. We observed that the performance of the method decreases as the values for σ become smaller. That can be explained by the non-robust detections of large number of model points due to the complex appearance of gene expression images. In the reported experiment, we set σ equal to the largest distance between any two points in the shape ($\sigma \approx 800$ in the context that the image size is 600×1000) so that, each point has an influence on the local fitting of all other points.

The fitting with such high value of σ is similar to the global model fitting. Thus, neither of Amberg-SIM and ASM-SIM outperforms the other in any of 14 regions.

The use of appearance guidance from only the reliable points helps PASM-CTX outperform ASM-SIM in nine regions (6-14) and outperform Amberg-SIM in eight regions (6-8, 10-14). Those regions have either mostly unreliable regional boundary points or many reliable interior points.

The improved results of PDM-ENLOR over PASM-CTX on all 14 regions indicate that the proposed ensemble scheme improves the flexibility of the models. PDM-ENLOR also outperforms ASM-SIM and Amberg-SIM in 14 regions. The improvement is more pronounced in the regions 6, 10, 12-14.
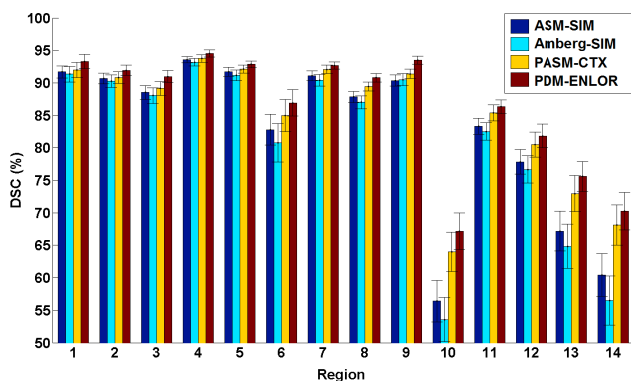

Figure 4: A comparison of the mean and confidence-interval of the DSC scores computed from the results of the ASM-based methods.

Comparison with previous works on the application: We also compared our method PDM-ENLOR with two state-of-the-art works on mouse brain gene expression image data of Kurkure *et al.* [16], which uses landmark-based registration, and Le *et al.* [20], which uses discrete Markov Random Fields to fit the subdivision mesh that models the shape. Our method outperforms [20] in all 14 regions and outperforms [16] in seven regions (2, 3, 8, 9, 11-13) which lack appearance cues. The improvement of our method is credited to the ability to incorporate shape prior as well as the appearance prior in a unified framework.

5. Conclusions

In this paper, we have presented a new approach to improve the model flexibility and to handle the detection errors for statistical shape fitting problem towards image segmentation. We proposed to locate each model point individually using an ensemble of PDM-based regression models which were constructed with increasing degree of locality to provide more flexibility. We presented automated methods to construct the local-to-global PDMs and to build an
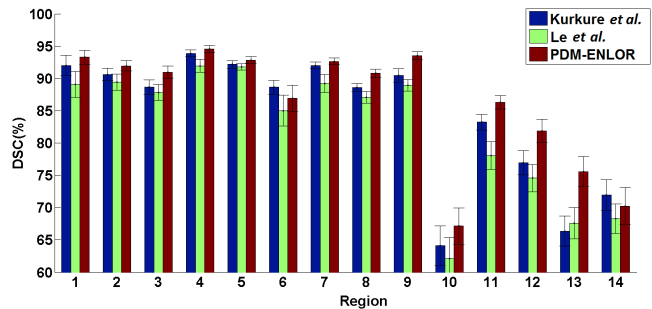

Figure 5: A comparison of the mean and confidence-interval of the DSC scores computed from the results of Kurkure *et al.* [16], Le *et al.* [20], and PDM-ENLOR.

ensemble by combining their output linearly using learned coefficients. Thus, the appropriate degree of locality is determined automatically from the training data. A set of selected salient reference points is used to construct the models to minimize the errors in fitting due to unreliable model points. Therefore, there is no need to perform any post-processing. We demonstrated that the use of appearance cues from selected model points can significantly improve the fitting results. Furthermore, our method outperforms the state-of-the-art methods on a challenging problem of multi-region segmentation of the mouse brain gene expression images.

Acknowledgments

References

[1] J. Abi-Nahed, M.-P. Jolly, and G.-Z. Yang. Robust active shape models: A robust, generic and simple automatic segmentation tool. In *Proc. 9th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II*, pages 1–8, Copenhagen, Denmark, Oct. 1-6 2006. 1, 2

[2] M. Amberg, M. Lüthi, and T. Vetter. Local regression based statistical model fitting. In *Proc. 32nd DAGM Conference on Pattern Recognition*, pages 452–461, Darmstadt, Germany, Sep. 22-24 2010. 1, 2, 6

[3] Y. Appia, B. Ganapathy, and A. Yezzi. Localized principal component analysis based curve evolution: A divide and conquer approach. In *Proc. 13th International Conference on Computer Vision*, pages 1981–1986, Barcelona, Spain, Nov. 6-13 2011. 2

[4] M. Bello, T. Ju, J. P. Carson, J. Warren, W. Chiu, and I. A. Kakadiaris. Learning-based segmentation framework for tis-

sue images containing gene expression data. *IEEE Trans. Med. Imag.*, 26:728–744, 2007. 5, 6

[5] S. Brecheisen, H. Kriegel, P. Krőger, M. Pfeifle, M. Schubert, and A. Zimek. Density-based data analysis and similarity search. In V. A. Petrushin and L. Khan, editors, *Multimedia Data Mining and Knowledge Discovery*, pages 94–115. Springer, 2007. 4

[6] J. P. Carson, T. Ju, M. Bello, C. Thaller, J. Warren, I. A. Kakadiaris, W. Chiu, and G. Eichele. Automated pipeline for atlas-based annotation of gene expression patterns: Application to postnatal day 7 mouse brain. *Methods*, 50:85–95, Aug. 2010. 6

[7] T. Cootes and C. Taylor. Active shape models: smart snakes. In *Proc. British Machine Vision Conference*, pages 266–275, Leeds, UK, Sep. 22-24 1992. 1

[8] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995. 1

[9] T. Cootes and C. J. Taylor. Combining point distribution models with shape models based on finite element analysis. *Image and Vision Computing*, 13(5):419–428, 1995. 1, 2

[10] C. Davatzikos, X. Tao, and D. Shen. Hierarchical active shape models, using the wavelet transform. *IEEE Trans. Med. Imag.*, 22(3):414 –423, 2003. 1, 2

[11] M. de Bruijne, B. Ginneken, M. Viergever, and W. Niessen. Adapting active shape models for 3D segmentation of tubular structures in medical images. In *Proc. 18th Information Processing in Medical Imaging*, pages 136–147, Ambleside, United Kingdom, Jul. 20-25 2003. 1, 2

[12] M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and U. Fayyad, editors, *Proc. 2nd International Conference on Knowledge Discovery and Data Mining*, pages 226–231, Portland, Oregon, Aug. 2-4 1996. 4

[13] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the Association for Computing Machinery*, 24(6):381–395, 1981. 2

[14] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, 53(2):285–339, 1991. 4

[15] T. Ju, J. Warren, G. Eichele, C. Thaller, W. Chiu, and J. Carson. A geometric database for gene expression data. In *Proc. Eurographics Symposium on Geometry Processing*, pages 166–176, Aachen, Germany, Jun. 22 - 25 2003. 3, 5, 6

[16] U. Kurkure, Y. H. Le, N. Paragios, J. P. Carson, T. Ju, and I. A. Kakadiaris. Landmark/image-based deformable registration of gene expression data. In *Proc. Computer Vision and Pattern Recognition Conference*, pages 1089–1096, Colorado Springs, CO, Jun. 21-23 2011. 1, 2, 3, 7

[17] U. Kurkure, Y. H. Le, N. Paragios, J. P. Carson, T. Ju, and I. A. Kakadiaris. Markov random field-based fitting of a subdivision-based geometric atlas. In *Proc. International Conference on Computer Vision*, pages 2540–2547, Barcelona, Spain, Nov. 6-13 2011. 5

[18] C. Last, S. Winkelbach, F. Wahl, K. Eichhorn, and F. Bootz. A locally deformable statistical shape model. *Machine Learning in Medical Imaging*, 7009:51–58, 2011. 2

[19] C. Lawson and R. Hanson. *Solving Least-Squares Problems*. Prentice-Hall, 1974. 6

[20] Y. Le, U. Kurkure, N. Paragios, T. Ju, J. Carson, and I. Kakadiaris. Similarity-based appearance prior for fitting a subdivision mesh in gene expression images. In *Proc. 15th International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 577–584, Nice, France, Oct. 1-5 2012. 2, 3, 5, 7

[21] K. Lekadir, R. D. Merrifield, and G.-Z. Yang. Outlier detection and handling for robust 3D active shape models search. *IEEE Trans. Med. Imag.*, 26(2):212–222, 2007. 2

[22] M. Loog. Localized maximum entropy shape modelling. In *Proc. 20th International Conference on Information Processing in Medical Imaging*, pages 619–629, Kerkrade, The Netherlands, Jul. 2-6 2007. 2

[23] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.*, 16(2):187–198, Apr. 1997. 6

[24] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Multiscale 3D shape representation and segmentation using spherical wavelets. *IEEE Trans. Med. Imag.*, 26(4):598–618, Apr. 2007. 2

[25] P. Nair and A. Cavallaro. 3D face detection, landmark localization, and registration using a point distribution model. *IEEE Trans. Multimedia*, 11(4):611–623, 2009. 1

[26] Y. Ou and C. Davatzikos. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. In *Proc. 21st International Conference on Information Processing in Medical Imaging*, pages 50–62, Williamsburg, VA, Jul. 5-10 2009. 3

[27] M. Rogers and J. Graham. Robust active shape model search. In *Proc. European Conference on Computer Vision*, pages 517–530, London, UK, May 27-31 2002. 2

[28] H. A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926. 3

[29] P. Yan, S. Xu, B. Turkbey, and J. Kruecker. Discrete deformable model guided by partial active shape model for trus image segmentation. *IEEE Trans. Biomed. Eng.*, 57(5):1158 –1166, 2010. 1, 2

[30] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: sparse shape composition. *Medical Image Analysis*, 16(1):265–277, 2012. 2

[31] Z. Zhao, S. R. Aylward, and E. K. Teoh. A novel 3d partitioned active shape model for segmentation of brain mr images. In *Proc. 8th International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 221–228, Palm Springs, CA, Oct. 26-30 2005. 1, 2

[32] D. Zhou, D. Petrovska-Delacretaz, and B. Dorizzi. Automatic landmark location with a combined active shape model. In *Proc. 3rd International Conference on Biometrics: Theory, Applications, and Systems, 2009.*, pages 1–7, Washington, DC, Sept. 28-30 2009. 1