Bayesian Model-Based Automatic Landmark Detection for Planar Curves

Justin Strait and Sebastian Kurtek
Department of Statistics
The Ohio State University
strait.50@osu.edu and kurtek.1@stat.osu.edu

Abstract

Identifying landmarks, points of interest on a shape, is crucial for many statistical shape analysis applications. Landmark-based methods dominate early literature; more recently, a method combining continuous shape outlines with landmark constraints was proposed. Unfortunately, methods requiring landmark specification depend on the number selected and their locations; such annotations are tedious for large datasets and subject to human interpretation. This work provides a Bayesian model-based method for automatic landmark selection, based on good approximations of landmark set interpolations. We outline an appropriate prior and likelihood, allowing for efficient posterior inference on landmark locations. The model allows for location uncertainty quantification, an important inferential procedure for further analysis. A method for selecting an appropriate number of landmarks is also discussed. Applications include a simulated example, shapes from the MPEG-7 dataset, and mice vertebrae.

1. Introduction

The comparison and analysis of *shape*, the representation of an object by its outline, lends itself useful in applications ranging from medical imaging to pattern recognition. *Statistical shape analysis* provides tools for summarization and inference based on shape data. Many of the techniques in literature resemble standard statistical methods for multivariate data; however, two complications arise from the analysis of shapes:

1. The lack of a unifying mathematical representation of shapes.
2. The nonlinearity of the shape space.

Both issues are caused by the mathematically complex nature of shapes. The definition of shape requires invariance to all shape-preserving transformations, including translation, rotation, and scaling, of the objects of interest. As a result, several different representations of these objects have been proposed in existing literature. Early methods (pioneered by Kendall in [7]) used the idea of *landmarks*, a finite set of important, labeled points on a curve which help represent the population of shapes. Using this representation allows one to apply multivariate statistical techniques (with some slight adjustments), like those discussed in Dryden and Mardia [4] and Small [15]. However, the landmark approach tends to provide a crude approximation of the objects of interest; much information about the geometry of the shape is lost, resulting in possibly biased statistical conclusions.

More modern approaches improve on landmark methods by formulating representations of shapes based on the parameterized curve defining the outline of an object ([19],[11],[8]). These representations are now infinite-dimensional, as opposed to their finite-dimensional landmark counterparts. Introducing frameworks with curve representations for shape analysis requires an additional invariance property to be added to the definition of shape: invariance to re-parameterization (dictates the rate at which the curve is traversed). *Elastic shape analysis* outlines methods to compare curves while incorporating this invariance, by matching features of curves through the search for an optimal re-parameterization. One such framework relies on the square-root velocity function (SRVF), developed in a series of papers ([16],[6],[9]). This framework is applicable to both open and closed Euclidean curves. However, recent work (see [17]) has shown that an improvement in shape alignment and comparison can be achieved by combining the SRVF framework with specified landmark points, essentially constraining curve parameterizations; this is known as landmark-constrained elastic shape analysis, and raises two important questions:

1. Where and how should the landmarks be selected on a shape?
2. How many landmarks should be selected?

For some shapes, it is immediately obvious where the important points lie; however, this is not true in most cases.

In this work, we develop a method which searches for landmark locations (and the number of landmarks), based on simple properties of the shape at hand. More importantly, the search is Bayesian model-based, which allows efficient statistical inference to be performed on the posterior landmark locations including uncertainty quantification. Finding the landmark locations then allows for landmark-constrained elastic shape analysis for shape comparison and statistical modeling.

The idea of automatic landmark detection is not novel by any means. Automatic methods for extracting landmark information have been discussed for many different classes of shapes; many of them tackle this problem from an image analysis point of view. However, much of the literature only focuses on specific objects; for instance, Chen *et al*. ([1]) discuss estimation of X-ray landmarks jointly through a voting scheme involving displacements of image patches. The tracking of facial expression landmarks over time using particle filters with kernel correlation techniques was proposed by Tie and Guan in [18]. Curvature-based methods have been discussed by Segundo *et al*. [14] and Gilani *et al*. [5], both of which are also regarding facial objects. Rueda *et al*. considered general shapes and employed variance equalization (in essence placing landmarks at locations with high variance) in [12]. They also considered local curvature in [13]. However, none of these methods are model-based. One Bayesian model-based approach was proposed by Domijan and Wilson in [3], using a segmented image idea with a model depending on pixel values.

In view of past literature on this topic, the main benefits of the proposed Bayesian method are three-fold:

1. The method is model-based, rather than relying on deterministic algorithms using features of the shape such as curvature or manual annotation of landmarks, which can be time consuming.

2. The Bayesian perspective allows for quantifying uncertainty in landmark locations based on posterior inference; this is very important for subsequent statistical shape analysis given that most objects do not have shapes for which the locations of landmarks are certain.

3. The model is applicable to general shapes of open or closed curves, and does not rely on intrinsic properties of the objects under study (e.g., the meaning of features on a face).

The rest of this paper is organized as follows. In Section 2, we provide a detailed description of the proposed Bayesian model for automatic landmark detection, which relies on linear interpolations of landmark sets. Section 3 presents a number of examples where the proposed method is applied to simulated and complex shapes. Finally, we provide a summary and some directions for future work in Section 4.

2. Statistical Framework for Landmark Selection

In this section, we present a Bayesian approach to determining landmark locations on planar curves. We begin with some background material related to elastic shape analysis, which is used for specification of the model; the specific prior and likelihood specifications follow. Once the prior and likelihood are determined, the posterior is derived (up to a proportionality constant). Finally, an importance sampling technique used for sampling from the posterior is discussed, along with a heuristic approach for selecting the number of landmarks.

2.1. Elastic Shape Analysis Preliminaries

Let $\beta : D \to \mathbb{R}^2$ denote an absolutely continuous, planar curve which represents the outline of a particular object of interest. Here D is the curve domain with $[0, 1]$ for open and $\mathbb{S}^1$ for closed curves. The goal is to automatically identify the locations of a fixed number of landmarks, say k, on β. In this framework, we assume k is relatively small, which is reasonable considering literature on landmark-based shape analysis. Choosing k can be a difficult task; one approach is discussed in Section 2.5. Notice that β is parameterized by $t \in D$. Thus, let $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(k)} \in D$ represent values of the parameter t, which correspond to landmark locations on β. Without loss of generality, assume $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(k)}$.

At this point, we must distinguish between open curves and closed curves. Closed curves require $\beta(0) = \beta(2\pi)$, whereas open curves do not. The importance in differentiating between these types of curves arises because while open curves have a natural starting and ending point, closed curves do not; any point on the curve can be treated as the starting and ending point. In practice, closed curves are provided sequentially, with the first point given equal to the last point. We desire our Bayesian method to be independent of the chosen starting point.

In order to compare two shapes β_1 and β_2, a metric must be specified on the shape space. This metric must be invariant to rigid motion, scaling, and re-parameterization. Unfortunately, as shown in several places including [16], the standard $\mathbb{L}^2$ metric is not parameterization invariant. To address this problem, Srivastava *et al*. ([16]) define the *square-root velocity function* (SRVF) as $q(t) = \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}}$, where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^2$. The original curve β can be recovered from q up to a translation; in other words, there is a bijective map between β and the pair $(\beta(0), q)$. The benefits of the SRVF representation include:

- The unit q vector at t represents the direction of the tangent vector to $\beta(t)$.
- The squared norm of $q(t)$ is the instantaneous speed of $\beta(t)$.
- Because the SRVF only depends on $\dot{\beta}$, invariance to translation is automatic.
- The representation is applicable to both open and closed curves.
- Finding distances between SRVFs via the $\mathbb{L}^2$ metric is equivalent to measuring differences between the corresponding curves using the elastic metric.

Invariance to scaling, rotations, and re-parameterizations can also be obtained through appropriate restrictions on the space of SRVFs and algebraic structures; details can be found in [16]. The importance of the SRVF for our purpose comes from the last two items. The ability to present a unified framework for both open and closed curves (with some modifications for closed curves) is rather beneficial. Additionally, the equivalence of the $\mathbb{L}^2$ metric for SRVFs to a special case of the elastic metric for curves allows a clear interpretation of the distance between SRVFs as the amount of bending and stretching required to transform one shape into the other (see [16] for more details). In a sense, this describes the length of the path of deformations from one shape to the other, and as will be seen later, is used in the specification of the likelihood in the proposed Bayesian model. A larger value of this metric indicates the need for more bending and stretching for shape deformations.

2.2. Prior Specification on Landmark Locations

The first task is to specify a prior distribution for $\boldsymbol{\theta} = (\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(k)})$. For simplicity, we describe this for the case of $D = [0, 1]$ and note that a simple extension can be used for $D = \mathbb{S}^1$. Since $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(k)} \in [0, 1]$ by definition, a naive prior specification would assume $\theta_1, \theta_2, \ldots, \theta_k \overset{iid}{\sim} U(0, 1)$; i.e., for $i = 1, \ldots, k$,

$$f(\theta_i) = I(0 < \theta_i < 1) \tag{1}$$

where I is the indicator function. However, this creates two issues: (1) landmark locations are certainly not independent, and (2) allowing each component to be independent and identically distributed from $U(0, 1)$ does not guarantee that $\theta_1 < \theta_2 < \ldots < \theta_k$ (which was imposed in Section 2.1). These problems can be resolved by noticing that $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(k)}$ are the order statistics of $\theta_1, \theta_2, \ldots, \theta_k$, since $\theta_{(1)} < \theta_{(2)} < \ldots < \theta_{(k)}$; the joint distribution of all order statistics can be derived with not much effort if the marginal distributions of each component are known and independent. Therefore, our noninformative prior distribution on the landmark locations is given by:

$$\pi(\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(k)}) = k!\, I(0 < \theta_1 < \theta_2 < \ldots < \theta_k < 1) \tag{2}$$

We also explored other prior distributions; however, many of these priors did not scale favorably with the number of landmarks in terms of sampling efficiency. As seen later, the proposed prior does not have this problem.

2.3. Likelihood Specification

Next, we choose a likelihood function for the data (β) given landmark locations $\boldsymbol{\theta}$. As a pre-processing step, we rescale β to have unit length (by rescaling its corresponding SRVF to have $\mathbb{L}^2$-norm equal to one); recall that this does not impact the object's shape. The motivation for the proposed likelihood comes from defining landmarks as point sets, which are important for representing the shape of interest. Suppose $\boldsymbol{\theta} = (\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(k)})$ are a sample of parameter values at which landmarks are located on β. Let $L_{\boldsymbol{\theta}}$ be the linear interpolation which passes through $\beta(\theta_{(1)}), \beta(\theta_{(2)}), \ldots, \beta(\theta_{(k)})$. For open curves, we additionally specify that $L_{\boldsymbol{\theta}}$ passes through the endpoints $\beta(0)$ and $\beta(1)$. Both β and $L_{\boldsymbol{\theta}}$ can be represented by their corresponding SRVFs, denoted q_β and $q_{L_{\boldsymbol{\theta}}}$, respectively. Then, the elastic distance between the two curves can be calculated using the $\mathbb{L}^2$ distance between their corresponding SRVFs:

$$d(\beta, L_{\boldsymbol{\theta}}) = \sqrt{\int_D |q_\beta(t) - q_{L_{\boldsymbol{\theta}}}(t)|^2 \, dt} \tag{3}$$

This distance calculation for closed curves requires an additional seed alignment, since it should not depend on the starting and ending point of the two curves. In practice, the starting point on the original curve β is shifted such that it coincides with the starting point of the linear interpolation, i.e., $\theta_{(1)}$. After posterior samples are drawn, $\boldsymbol{\theta}$ is then shifted to coincide with the first posterior sample (as a reference point for display purposes).

Calculating the distance in Equation 3 provides a measure of how well the sampled landmark locations represent the underlying curve β: a small value of d means that the linear interpolation through the sampled $\boldsymbol{\theta}$ represents the shape well, in relation to other possible sampled landmark locations. Figure 1(a) shows $\boldsymbol{\theta}$ which provide a good linear interpolation for the given curve shape (small value of d). Panel (b) shows an example of a poor linear interpolation (large value of d). It is clear that the likelihood should be much greater for the good linear interpolation, as the bad one features landmark points that are not very representative for the given curve.

After calculation of d, the likelihood can be expressed as

$$f(\beta|\boldsymbol{\theta}, c) \propto c^{-N} \exp\left(\frac{-1}{2c} d^2(\beta, L_{\boldsymbol{\theta}})\right) \tag{4}$$

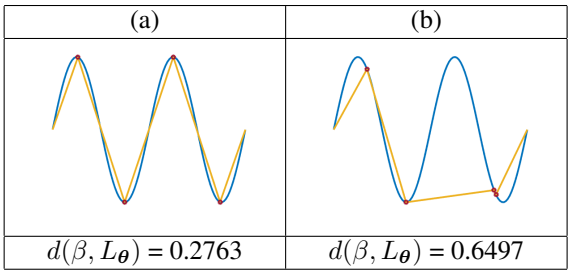


Figure 1. (a) Landmarks which produce a linear interpolation resulting in a small distance d (high likelihood). (b) Landmarks which produce a linear interpolation resulting in a large distance d (low likelihood). The original curve is shown in blue with sampled landmarks $\boldsymbol{\theta}$ in red. The linear interpolation is given in yellow. Distance between original curve and interpolations are provided.

where c represents the variance. The given curves are infinite-dimensional objects; however, in practice, we sample them using a finite number of points N. That is, we model the pointwise differences between the SRVFs of the original curve β and the linear interpolation using an isotropic multivariate Gaussian distribution with zero mean: $q_\beta - q_{L_{\boldsymbol{\theta}}} \sim N(0, cI_N)$, where I_N is an $N \times N$ identity matrix. Looking at the likelihood, we also require a prior on the variance parameter c, which is closely related to the scale of the distances d. Due to the imposed unit length constraint, this distance is bounded above by π (see [16] for details). We can therefore specify a noninformative prior on c as $\pi_c = U(0, \pi)$.

2.4. Posterior Distribution and Sampling

Now that the prior and likelihood distributions have been fully specified, the posterior can be obtained using Bayes' rule (up to a proportionality constant):

$$\pi(\boldsymbol{\theta}, c|\beta) \propto \pi(\boldsymbol{\theta})\pi_c(c)f(\beta|\boldsymbol{\theta}, c) \tag{5}$$

The posterior distribution provides information about the optimal landmark locations and likelihood variance, given the observed shape. In our setup, the variance c is a nuisance parameter, which we integrate out. Thus, the posterior we are interested in is given by:

$$\begin{aligned}\pi(\boldsymbol{\theta}|\beta) &\propto \pi(\boldsymbol{\theta}) \int_0^\pi \pi_c(c) f(\beta|\boldsymbol{\theta}, c) dc \\ &\propto \pi(\boldsymbol{\theta}) \int_0^\pi f(\beta|\boldsymbol{\theta}, c) dc = \pi(\boldsymbol{\theta}) f(\beta|\boldsymbol{\theta})\end{aligned} \tag{6}$$

Sampling from this posterior allows for meaningful posterior inference for landmark locations; however, the posterior as given is difficult to sample from. Thus, proceeding typically requires use of Markov chain Monte Carlo (MCMC) methods. However, since we specify the number of landmarks k to be fixed and relatively small, this low-dimensional posterior can be approximately sampled using the Sampling-Importance-Resampling (SIR) algorithm. To perform this, we first draw M samples $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$ from $\pi(\boldsymbol{\theta})$ (recall that each $\boldsymbol{\theta}$ is k-dimensional). In order to sample from the target distribution $\pi(\boldsymbol{\theta}|\beta) \propto \pi(\boldsymbol{\theta})f(\beta|\boldsymbol{\theta})$, importance sampling weights are calculated for each sample. In the present model, we take the importance function to be the same as the prior resulting in the following weights: $w(\boldsymbol{\theta}_i) = f(\beta|\boldsymbol{\theta}_i)$. Then, a sample of size $s << M$ is taken proportionally to the weights with replacement from the original sample $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$; in other words, for $i = 1, \ldots, M$, the probability of sampling $\boldsymbol{\theta}_i$ is given by $\frac{w(\boldsymbol{\theta}_i)}{\sum_{i=1}^M w(\boldsymbol{\theta}_i)}$. The resulting sample of size s is a sample from the posterior distribution $\pi(\boldsymbol{\theta}|\beta)$.

2.5. Selecting the Number of Landmarks

Choosing the number of landmarks k is not trivial. Some objects, such as the toy example in Figure 1, appear to have distinct landmark locations. Thus, selecting the number of landmarks for these cases is fairly straightforward. However, a general object from a given dataset may not share the luxury of having easy-to-specify landmarks. Thus, we also seek a method for determining k in the given framework. One of the main issues with selecting k relates to the classic overfitting problem in statistics. As one adds more and more landmarks, the linear interpolation approximation of β improves. But, adding too many landmarks complicates posterior inference and interpretation.

The problem of selecting k here resembles that of many other statistical problems where it is desired to select a relatively small number of components or groups to represent higher-dimensional data. For instance, principal component analysis (PCA) is a popular method for representing high-dimensional data by many fewer, but important, uncorrelated dimensions. Statisticians often perform PCA on data prior to conducting analyses; thus, the number of principal components to choose is a crucial step. Often, the number of components is selected by the percentage of variation that the components explain. Thus, as more components are included, the percentage of variation explained starts to level off, i.e., the marginal utility of including additional principal components decreases. This usually occurs at the "elbow" of a plot of percent variation explained versus number of principal components. A similar plot can be obtained to help choose the number of clusters k in k-means clustering, for example. We employ a similar method in our approach.

For a specified k, the posterior sample of $\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_s}$ can be used to form linear interpolations $L_{\boldsymbol{\theta_1}}, \ldots, L_{\boldsymbol{\theta_s}}$. Then, $d^2(\beta, L_{\boldsymbol{\theta_i}})$ can be computed for $i = 1, \ldots, s$. Finally, the average squared distances for all posterior samples are compared for different values of k in a plot similar to the one described in the previous paragraph. One expects a smaller squared distance as k increases (although this is not always true due to the complexity of shapes and sampling variabil-

ity). Based on this plot, k can be chosen as the "elbow", or the point in the plot for which adding more landmarks does not result in much decrease in the average squared distance. An example of this plot for the toy example, described in Figure 1, is given in the top left panel of Figure 2. It is clear that after reaching $k = 4$, the average squared distance begins to level off and any additional landmark does not significantly improve the interpolated curve representation.

3. Examples

In this section, we illustrate the use of this Bayesian model-based landmark detection method on some simulated and real data examples. Some of these examples have landmarks which are easily detectable; others are not quite as simple, and thus the Bayesian methodology proves very useful. For all closed curve examples, we chose $[0, 1]$ as the curve domain for simplicity.

3.1. Simulated Example

First, we consider the parameterized curve defined by the function $(x, f(x)) = (x, \sin(4\pi x))$, $0 \leq x \leq 1$. The number of landmarks ($k = 4$ as seen in Figure 2) and locations of the landmarks (at the "peaks" and "troughs" of the curve) for this shape appear to be fairly intuitive. For this example, $M = 10^6$ samples are initially drawn from the prior distribution. Using the corresponding likelihoods, we resample $s = 500$ sets of $\boldsymbol{\theta}$. Figure 2 shows the given shape with the landmark locations plotted in different colors. The posterior landmark locations are consistent with what is expected. The Bayesian methods used allow for visualization of the marginal histograms of the sampled posterior landmark locations to illustrate the uncertainty in these locations. For this example in particular, the locations are easily identifiable; the unimodality of the histograms reflects this, as well as the low spread in the sampled $\boldsymbol{\theta}$.

Interesting patterns emerge for this example when $k \neq 4$. Figure 3 shows sampled landmark locations for $k = 2, 3, 5$. For $k = 5$ (which is one more landmark than necessary), the model does not appear to know where to place the fifth landmark, thus impacting the locations of the other four. This is in contrast to $k = 2, 3$, which show multimodal locations because there are more important geometric features than landmarks specified. For each k, the sample standard deviation was also calculated for each component of $\boldsymbol{\theta}$, and the average was taken across components; a higher value would indicate more uncertainty overall in landmark placement. For the $k = 4$ case, this average was 0.0302. The figure indicates the average standard deviations were much greater for the other values of k selected, all of which are larger, even when $k = 5$.

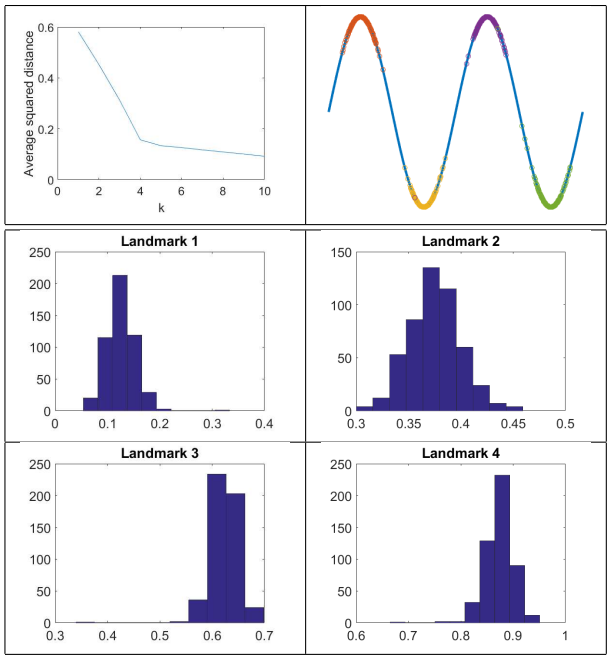


Figure 2. Top left: Plot of average squared distance vs. k. Top right: The curve of interest β with scatter points representing the posterior sample of landmark locations (red = landmark 1; yellow = landmark 2; purple = landmark 3; green = landmark 4). Bottom: Histograms of posterior samples for each of the four landmarks.

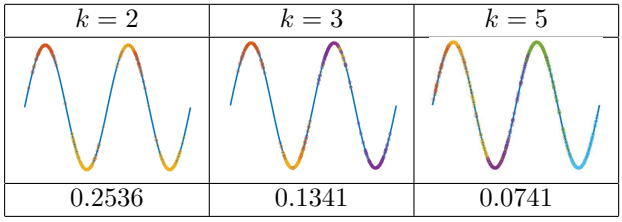


Figure 3. Sampled landmark locations for $k = 2, 3, 5$ for the simulated example. The average posterior sample standard deviation is given below each plot.

3.2. MPEG-7 Data Examples

Next, we consider several complex objects from the MPEG-7 dataset[1]. Figure 4 shows four example shapes from this data along with landmark locations samples from the posterior based on the proposed Bayesian models (with k chosen according to the proposed method). For all examples in this section, we use $M = 10^5$ and $s = 500$ (except the stingray, which uses $s = 250$).

Unfortunately, specifying the number of landmarks for these complex objects is a much more difficult task than for the toy example. In addition, landmark locations are not as clear. We examine the proposed method in detail for the half-circle shape and present the results in Figure 5. Our

[1] http://www.dabi.temple.edu/~shape/MPEG7/dataset.html

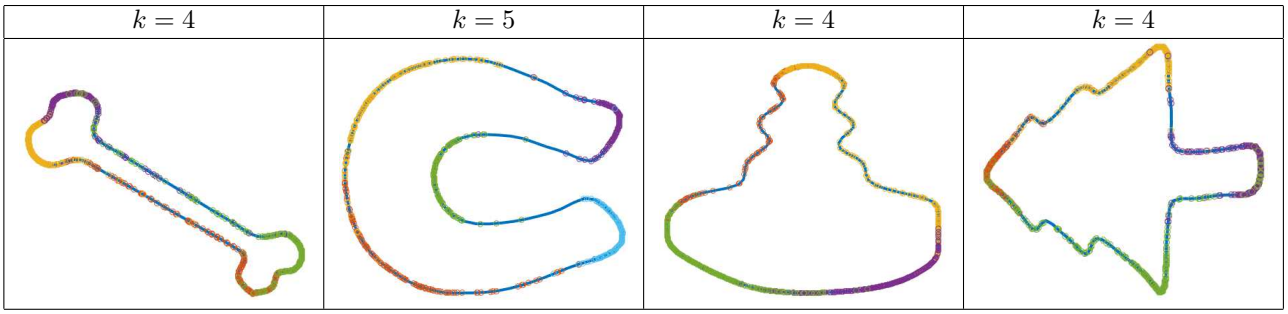


Figure 4. Four different shapes from the MPEG-7 dataset with their posterior landmark sample locations plotted in different colors.

landmark detection method is run for $k = 3, \ldots, 10$, and the average squared distance is computed for the posterior sampled $\boldsymbol{\theta}$. While not as pronounced, $k = 4$ appears to be the point beyond which further decreases to the average squared distance are inconsequential. Thus, we select $k = 4$ and proceed with posterior sampling. The posterior location of landmarks 3 and 4 (which are located close to the points of high curvature) feature small variances and unimodality, whereas landmarks 1 and 2 are much more variable. This again matches our intuition.

The MPEG-7 dataset features even more complex shapes, including the stingray displayed in Figure 6. The required number of landmarks in this example is not as clear. The top left panel of Figure 6 appears to indicate that selecting $k = 5$ is sufficient for the landmark set. The histograms display the posterior samples obtained from the SIR algorithm. Landmarks 1, 2, and 5 appear to be exactly where most would place them, i.e., at the two wings (yellow and blue landmark points) and the snout (red landmark points). Landmarks 3 and 4 are slightly less clear and exhibit much more variability. Again, the dependence in landmarks is evident here. There are some infrequent samples of landmark 3 at larger parameter values (e.g., the purple points that are mixed in with the green points along the stingray tail). In order to compensate for that different landmark location, the corresponding locations for landmark 4 also occur at larger parameter values relative to the rest of the histogram (the small mode around $t = 0.8$ corresponding to green points deviating from the tail). This dependence allows the linear interpolation to still be fairly close in distance to the original stingray outline yielding high likelihood values and corresponding posterior densities.

3.3. Mouse Vertebrae

The R package 'shapes'[2] contains a real dataset featuring the outlines of the second thoracic (T2) vertebrae of mice, some of which were genetically selected for either small or large weights, and the rest being from a control group which were not genetically selected. Further description of this data is provided in [4, 2]. We demonstrate the idea of selecting the number of landmarks k on the T2 vertebrae of a mouse from the large group as follows. First, posterior samples are drawn for landmark locations for $k = 3, \ldots, 10$ (with $M = 10^5$). For each k, the average distance of the linear interpolation of the posterior samples to the actual shape is computed and plotted in the top left panel of Figure 7. A large decrease in the average squared distance occurs from three to four landmarks; however, after $k = 4$, the decrease in the average squared distance is not as substantial, meaning that there is a lack of utility in adding additional landmarks.

Thus, we select $k = 4$ and apply the proposed Bayesian

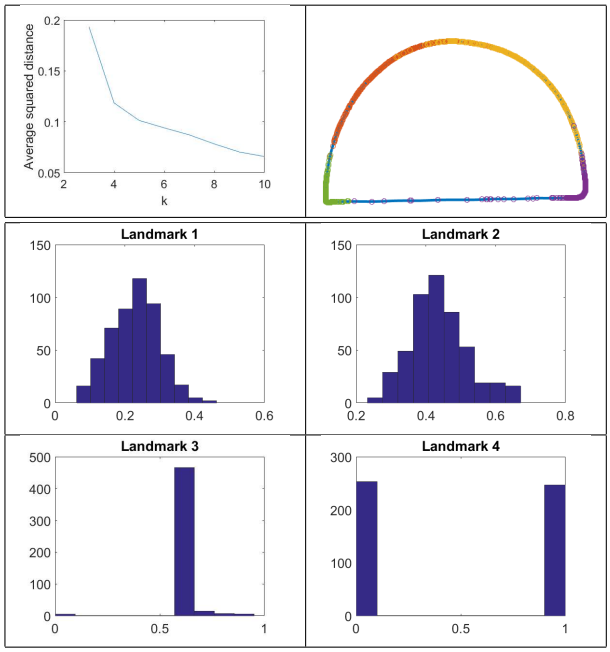


Figure 5. Top left: Plot of average squared distance vs. k. Top right: Half-circle β with scatter points representing the posterior sample of landmark locations (red = landmark 1; yellow = landmark 2; purple = landmark 3; green = landmark 4). Bottom: Histograms of posterior samples for each landmark.

[2] https://cran.r-project.org/web/packages/shapes/index.html

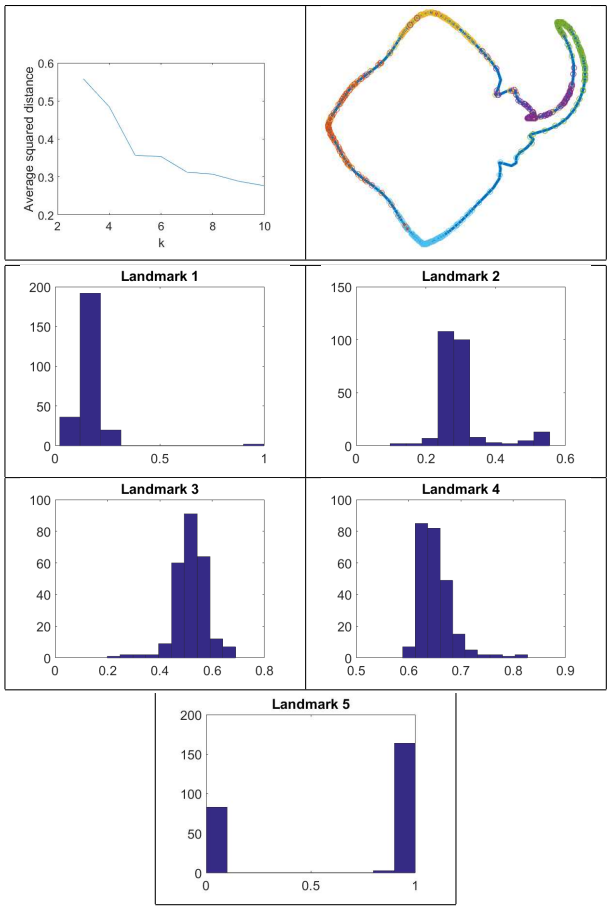


Figure 6. Top left: Plot of average squared distance vs. k. Top right: Stingray β with scatter points representing the posterior sample of landmark locations (red = landmark 1; yellow = landmark 2; purple = landmark 3; green = landmark 4; blue = landmark 5). Bottom: Histograms of posterior samples for each landmark.

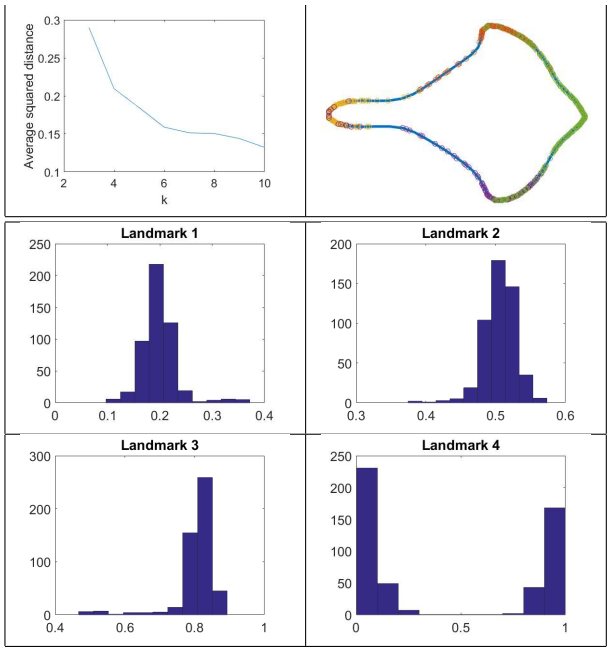


Figure 7. Top left: Plot of average squared distance vs. k. Top right: Mouse vertebrae β with scatter points representing the posterior sample of landmark locations (red = landmark 1; yellow = landmark 2; purple = landmark 3; green = landmark 4; blue = landmark 5). Bottom: Histograms of posterior samples for each landmark.

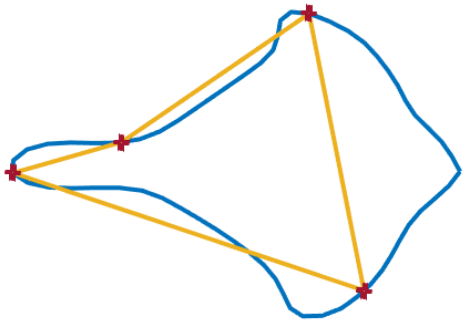


Figure 8. Example of sampled $\boldsymbol{\theta}$ which illustrates landmark dependence.

model for landmark detection. Figure 7 illustrates the sampled landmark locations along with the marginal posterior sample histograms. Again, all histograms have one distinct peak. However, we notice the slight multimodality of landmarks 1 and 3. These correspond to the upper and lower edges of the vertebrae. The stray peaks observed here illustrate the dependence of landmarks on each other. The high values of posterior sampled $\theta_{(1)}$ correspond to low values of posterior sampled $\theta_{(3)}$ in order to keep the distance between the linear interpolation and the vertebra relatively small (see Figure 8 for an example). Landmark 2 (yellow) displays the least amount of variation in location - this appears consistent with the idea that any linear interpolation going through that part of the vertebra must be relatively close to the point of maximum absolute curvature. Otherwise, the distance will grow very quickly.

3.4. Additional Posterior Summaries

In addition to graphical summaries of posterior samples of $\boldsymbol{\theta}$, numerical summaries can provide insight into landmark locations and quantification of uncertainty. Mean and median values of $\boldsymbol{\theta}$ can be calculated in the usual manner for open curves; however, some care must be taken when dealing with closed curves, as the domain is circular. In addition, the maximum a posteriori (MAP) estimate can be defined as

$$\boldsymbol{\theta}_{MAP} = \arg\max_{\boldsymbol{\theta}_i} w(\boldsymbol{\theta}_i) \tag{7}$$

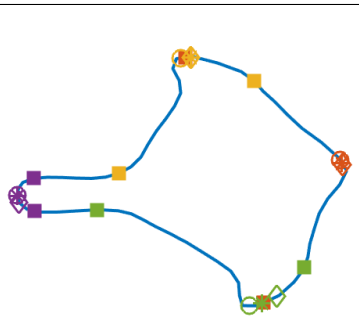
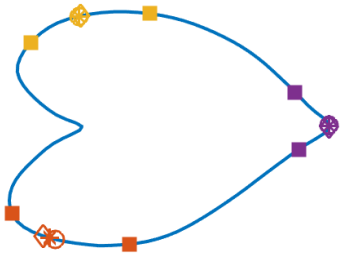
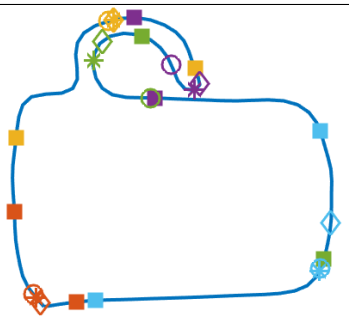
	(a)	(b)	(c)
			
Mean	$(0.013, 0.179, 0.500, 0.824)$	$(0.038, 0.440, 0.724)$	$(0.014, 0.280, 0.424, 0.565, 0.802)$
Median	$(0.010, 0.169, 0.503, 0.837)$	$(0.043, 0.440, 0.726)$	$(0.011, 0.284, 0.390, 0.505, 0.804)$
MAP	$(0.008, 0.169, 0.517, 0.848)$	$(0.049, 0.442, 0.728)$	$(0.005, 0.283, 0.379, 0.487, 0.763)$

Figure 9. Three examples, each with the posterior mean, median, and MAP of $\boldsymbol{\theta}$ (mean = circle, median = asterisk, MAP = diamond, 95% credible interval endpoints = square). The colors match for each component of $\boldsymbol{\theta}$ (red = 1, yellow = 2, purple = 3, green = 4, blue = 5).

where $w(\boldsymbol{\theta}_i)$ are the importance weights for $i = 1, \ldots, s$ (s being the number of posterior samples drawn from the SIR algorithm) described in Section 2.4. Variability can be addressed by constructing credible intervals using the percentiles of the posterior samples of $\boldsymbol{\theta}$ for each component. Figure 9 illustrates all of these summaries for three different examples (mouse vertebra in (a), heart in (b), cup in (c)); the curve outline is superimposed with markers for mean, median, and MAP locations, as well as endpoints for 95% credible intervals for each landmark. Note that on the plot, markers may overlap (especially for landmark locations with very low variability). This is evident, for instance, in the tail of the mouse vertebra; the landmark locations in purple appear to display low variability, and thus the mean, median, and MAP estimates all approximately coincide. The heart in Figure 9, with $k = 3$, is fairly easy to annotate landmarks on. The credible intervals reflect this, as the amount of variability in locations is small, and there is no overlap in intervals between landmarks. However, complex features on shapes may cause vastly different estimates in the mean, median, and MAP. The green landmarks in column (c), for example, are widely variable in location from the base of the cup to the handle itself; the MAP and median locations are fairly close together, but the mean location is further away and clearly impacted by some extreme samples drawn toward the cup's base.

4. Summary and Future Work

In this paper, we proposed a novel Bayesian model for automatic selection of landmark locations on planar curves. We also outline a simple heuristic method for choosing an adequate number of landmarks. A key component of our Bayesian model is the dependence of the likelihood on the distance between a linearly interpolated curve through sampled landmark locations and the original curve. We choose a noninformative prior on the landmark locations and capture dependence through order statistics. We perform efficient posterior sampling using the SIR algorithm with importance weights equal to the likelihood specified in the model.

There are several future directions for this method. One is related to the selection of the number of landmarks k. Instead of relying on a simple heuristic based on the "elbow" in the average squared distance plots, a more reasonable approach would be to put a prior on k and infer that quantity as well. Clearly, for some shapes, there could be large uncertainty in the number of landmarks to specify, especially if landmark locations are also very uncertain. This addition to the model would require a different way of sampling from the posterior, since the size of the parameter space changes dimension depending on k. Another direction is to consider the impacts of prior specification on $\boldsymbol{\theta}$ and c on posterior inference through sensitivity analysis. Finally, we will explore the idea of posterior inference on $\boldsymbol{\theta}$ given a sample of shapes (for instance many different T2 vertebrae of mice), rather than just one shape. Doing so will allow for automatic annotation of shapes to occur jointly (using much more information) instead of the current proposal of selecting landmarks on each shape individually (as a doctor may do with medical images, for example). A more ambitious goal is to extend this work to detect landmarks on surfaces, which has been of recent interest in graphics and medical imaging (see [10] and [20]).

References

[1] C. Chen, W. Xie, J. Franke, P. Grutzner, L. Nolte, and G. Zheng. Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements. *Medical Image Analysis*, 18(3):487–499, 2014.

[2] W. Cheng, I. L. Dryden, and X. Huang. Bayesian registration of functions and curves. *Bayesian Analysis*, 11(2):447–475, 2016.

[3] K. Domijan and S. P. Wilson. A Bayesian method for automatic landmark detection in segmented images. In *Proceedings of the 22nd International Conference on Machine Learning*, Bonn, Germany, 2005.

[4] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. Wiley, New York, 1998.

[5] S. Z. Gilani, F. Shafait, and A. Mian. Shape-based automatic detection of a large number of 3D facial landmarks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, 2015.

[6] S. H. Joshi, E. Klassen, A. Srivastava, and I. H. Jermyn. A novel representation for Riemannian analysis of elastic curves in $\mathbb{R}^n$. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007.

[7] D. G. Kendall. Shape manifolds, Procrustean metrics, and complex projective shapes. *Bulletin of London Mathematical Society*, 16:81–121, 1984.

[8] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(3):372–383, 2004.

[9] S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding. Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association*, 107(499):1152–1165, 2012.

[10] S. Kurtek, A. Srivastava, E. Klassen, and H. Laga. Landmark-guided elastic shape analysis of spherically-parameterized surfaces. *Computer Graphics Forum (Proceedings of Eurographics)*, 32(2):429–438, 2013.

[11] W. Mio, A. Srivastava, and S. Joshi. On shape of plane elastic curves. *International Journal of Computer Vision*, 73(3):307–324, 2007.

[12] S. Rueda, J. Udupa, and L. Bai. Landmark selection for shape model construction via equalization of variance. In *Proceedings of 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008*, Paris, 2008.

[13] S. Rueda, J. Udupa, and L. Bai. A new method of automatic landmark tagging for shape model construction via local curvature scale. In *Proceedings of SPIE, Medical Imaging, Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling*, 2008.

[14] M. P. Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo. Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 40(5):1319–1330, 2010.

[15] C. G. Small. *The Statistical Theory of Shape*. Springer, 1996.

[16] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:1415–1428, 2011.

[17] J. Strait, S. Kurtek, E. Bartha, and S. MacEachern. Landmark-constrained elastic shape analysis of planar curves. *Journal of the American Statistical Association*, In review, 2016.

[18] Y. Tie and L. Guan. Automatic landmark point detection and tracking for human facial expressions. *EURASIP Journal on Image and Video Processing*, 8, 2013.

[19] L. Younes. Computable elastic distance between shapes. *SIAM Journal of Applied Mathematics*, 58(2):565–586, 1998.

[20] J. Zaetz and S. Kurtek. A novel Riemannian framework for shape analysis of annotated surfaces. In *Proceedings of International Workshop on Differential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories*, pages 3.1–3.11, 2015.