

Robust Models of Object Geometry

Jared Glover, Daniela Rus and Nicholas Roy
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute Of Technology
Cambridge, MA 02139
Email: {jglov,rus,nickroy}@mit.edu

Geoff Gordon
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15289
Email: ggordon+@cs.cmu.edu

Abstract—Precise and accurate models of the world are critical to autonomous robot operation. Just as robot navigation typically requires an accurate map of the world, robot manipulation typically requires accurate models of the objects to be grasped. However, the statistical inference tools that enable robot mapping have not yet had the same impact in geometric object modelling. We describe an inference algorithm for learning statistical models of objects from image data. We describe a representation that allows us to compute a distribution over the complete geometry of different objects, and describe how a library of object geometry models can be learned. Finally, we describe how learned object models can both be used to recognize new instances of objects and to infer the geometry of occluded parts of objects.

I. INTRODUCTION

Precise and accurate models of the world are critical to autonomous robot operation. Just as robot navigation typically requires an accurate map of the world, robot manipulation typically requires accurate models of the objects to be grasped. However, the statistical inference tools that enable robot mapping have not yet had the same impact in geometric object modelling. One of the principal difficulties in learning object models is finding a good representation; frequently, the object representation that is most useful for one task is not particularly useful for another. For instance, there exist sparse feature-based object models such as SIFT features [12] that allow for very reliable object recognition and tracking but are completely inappropriate for motion planning in that the complete geometry of the object is not recovered. Techniques such as shape contexts [1] or spherical harmonics [9] allow general classes of objects to be learned over time, but again, the geometry of individual objects is not always preserved.

For autonomous robot operation, we would like to be able to infer the complete geometry of objects in a manner that is robust to variances in shape from object to object, and is robust to perceptual occlusions. First, our algorithm should learn a description of the complete geometry of the object. In order to allow a robot to carry out control tasks such as navigation, manipulation, grasping, etc. it is not sufficient to describe an object as a set of features; we will need the ability to describe the complete boundary of the object for computing potential collisions, grasp closures, etc. Secondly, our algorithm should allow us to infer the object’s geometry from a series of independent measurements. Just as in robot mapping, we would like to integrate a series of measurements in time, in order to learn the object model. This problem differs from the mapping problem in that we (usually) will not have to solve for the object description and sensor position simultaneously.



Figure 1. Contour extraction. Left: Raw image (poor white balance in the camera results in poor colour rendering). Middle: Contours extracted using pyramid segmentation. Right: Contours extracted using intensity thresholding. Our goal is to recognize different objects from the same class from the complete or partial outline of each object.

Finally, our algorithm should allow us to estimate the geometry of any occluded parts of the object from a partial view — we would like to have a robot that can recognize and pick up a tool without first building a complete and accurate model of the tool. In robot navigation, the map is usually assumed to be complete before any autonomous motion planning is attempted. In a populated, dynamic environment, the assumption of a complete description of the environment is clearly brittle as new objects will appear regularly that will require robot actions. Our representation must therefore be able to recognize objects it may not have seen before, and be able to make reasonable inferences about the parts of the object that are occluded.

This paper presents a unified probabilistic approach to modelling object geometry, in particular focusing on the problem of incomplete data. We draw upon a representation [10] for object geometry which is invariant to changes in position, scale, and orientation, and show how to learn object models directly from sensor data. This representation is robust to sensor noise and also allows us to capture the variation between shapes in the same object class in a meaningful way. By learning a probabilistic, generative model of the complete geometry of an object, we can also make predictions about sections of the object geometry that are not visible. We demonstrate this approach on some example images of everyday objects. The results presented in this paper are restricted to monocular video images; we have generalized this approach to three dimensions from data such as stereo or laser range data but do not include these results in this paper.

II. SHAPE SPACE

Let us represent a shape \mathbf{z} as a set of n points $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ in some Euclidean space. We will restrict ourselves to two-dimensional points (representing shapes in a plane) such that $\mathbf{z}_i = (x_i, y_i)$, although extensions to three dimensions are feasible. We will assume these points are ordered (so that \mathbf{z} can

be defined as a vector), and represent a closed contour (such as the letter ‘‘O’’, as opposed to the letter ‘‘V’’). In general, the shape model can be made *de facto* invariant to point ordering if we know correspondences between the boundary points of any two shapes we wish to compare. Even in the case of closed contours, however, care must still be taken to choose the correct starting point. In order to make our model invariant to changes in position and scale, we can normalize the shape so as to have unit length with centroid at the origin; that is,

$$\mathbf{z}' = \{\mathbf{z}'_i = (x_i - \bar{x}, y_i - \bar{y})\} \quad (1)$$

$$\tau = \frac{\mathbf{z}'}{|\mathbf{z}'|} \quad (2)$$

where $|\mathbf{z}'|$ is the L^2 -norm of \mathbf{z}' . We call τ the *pre-shape* of \mathbf{z} . Since τ is a unit vector, the space of all possible pre-shapes of n points is a unit hyper-sphere, \mathbb{S}_*^{2n-3} , called *pre-shape space*¹.

Any pre-shape is a point on the hypersphere, and all rotations of the shape lie on an orbit, $\mathcal{O}(\tau)$, of this hypersphere. If we wish to compare shapes using some distance metric between them, the spherical geometry of the shape space requires a geodesic distance rather than Euclidean distance. Additionally, in order to ensure this distance is invariant to rotation, we define the distance between two shapes τ_1 and τ_2 as the smallest distance between their orbits:

$$d_p[\tau_1, \tau_2] = \inf[d(\phi, \psi) : \phi \in \mathcal{O}(\tau_1), \psi \in \mathcal{O}(\tau_2)] \quad (3)$$

$$d(\phi, \psi) = \cos^{-1}(\phi \cdot \psi) \quad (4)$$

We call d_p the *Procrustean metric* [10] where $d(\phi, \psi)$ is the geodesic distance between ϕ and ψ . Since the inverse cosine function is monotonically decreasing over its domain, it is sufficient to maximize $\phi \cdot \psi$, which is equivalent to minimizing the sum of squared distances between corresponding points on ϕ and ψ (since ϕ and ψ are unit vectors). For every rotation of ϕ there exists a rotation of ψ which will find the global minimum geodesic distance. Thus, to find the minimum distance, we need only rotate one pre-shape while holding the other one fixed. We call the rotated ψ which achieves this optimum ($\theta_{\alpha^*}(\tau_2)$) the *orthogonal Procrustes fit* of τ_2 onto τ_1 , and the angle α^* is called the *Procrustes fit angle*.

Representing the points of τ_1 and τ_2 in complex coordinates, which naturally encode rotation in the plane by scalar complex multiplication, the Procrustes distance minimization can be solved:

$$d_p[\tau_1, \tau_2] = \cos^{-1} |\tau_2^H \tau_1| \quad (5)$$

$$\alpha^* = \arg(\tau_2^H \tau_1), \quad (6)$$

where τ_2^H is the *Hermitian*, or complex conjugate transpose of the complex vector τ_2 .

III. SHAPE INFERENCE

Given a set of measurements of an object, we would like to infer a distribution of possible object geometry. We will first derive an expression for the mean shape, and then derive the

¹Following [13], the star subscript is added to remind us that \mathbb{S}_*^{2p-3} is embedded in \mathbb{R}^{2p} , not the usual \mathbb{R}^{2p-2} .

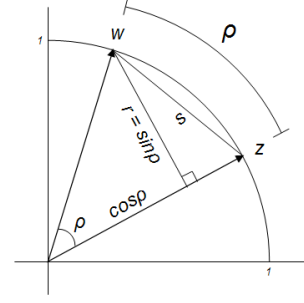


Figure 2. Distances on a hyper-sphere. r is a linear approximation to the geodesic distance ρ .

distribution covariance. Finally, we will show how to infer the full object shape using measurements of partial object geometry.

1) *The Shape Distribution Mean:* Let us assume that our data set consists of a set of measurements $\{\mathbf{m}_1, \mathbf{m}_2, \dots\}$ of the same object, where each measurement \mathbf{m}_i is a complete (but noisy) description of the object geometry, a vector of length p . We can normalize each measurement to be an independent pre-shape and use these pre-shapes to compute the mean, that is, the maximum likelihood object shape. If each measurement m_i is normalized to a pre-shape τ_i , then the mean is

$$\mu^* = \arg \inf_{\|\mu\|=1} \sum_i [d_p(\tau_i, \mu)]^2 \quad (7)$$

The pre-shape μ^* is called the *Fréchet mean*² of the samples τ_1, \dots, τ_n with respect to the distance measure ‘ d_p ’. Note that the mean shape is *not* trivially the arithmetic mean of all pre-shapes; d_p is non-Euclidean, and we wish to preserve the constraint that the mean shape has unit length.

Unfortunately, the non-linearity of the Procrustes distance in the $\cos^{-1}(\cdot)$ term leads to an intractable solution for the Fréchet mean minimization. Thus, we approximate the geodesic distance ρ between two pre-shape vectors, $\phi \in \mathcal{O}(\tau_1)$ and $\psi \in \mathcal{O}(\tau_2)$, as the projection distance, r ,

$$r = \sin \rho = \sqrt{1 - \cos^2 \rho}. \quad (8)$$

Figure 2 depicts this approximation to the geodesic distance graphically³.

Using this linear projection distance r in the mean shape minimization yields an expression for the mean shape μ^*

$$\mu^* = \arg \inf_{\|\mu\|=1} \sum_i (1 - |\tau_i^H \mu|^2) \quad (9)$$

$$= \arg \sup_{\|\mu\|=1} \sum_i (\tau_i^H \mu)^H (\tau_i^H \mu) \quad (10)$$

$$= \arg \sup_{\|\mu\|=1} \mu^H \left(\sum_i \tau_i \tau_i^H \right) \mu \quad (11)$$

$$= \arg \sup_{\|\mu\|=1} \mu^H S \mu, \quad (12)$$

thus, μ^* is the complex eigenvector corresponding to the largest eigenvalue of S [5].

²More precisely, μ^* is the Fréchet mean of the random variable ξ drawn from a distribution having uniform weight on each of the samples τ_1, \dots, τ_n .

³The straight-line Euclidean distance, s , is another possible approximation to the geodesic distance, but it will not enable us to easily solve the mean shape minimization in closed form.



Figure 3. Full-contour shape representation. Left to right: original image, an example extracted contour and the mean of the learned shape class. The top of the tool is slightly rounded off due both to the small number of points (50) sampled from the contour to form the preshape, as well as to the smoothing effect of computing the mean shape of the tool as it opens and closes.

Figure 3 shows an example of the mean shape learned for the tool class. On the left is an example raw image, in the middle is an extracted measurement contour, and on the right is the learned mean of the tool distribution.

2) *The Shape Distribution Covariance:* With our measurements $\{\mathbf{m}_1, \mathbf{m}_2, \dots\}$, we can now fit a probabilistic model of shape. In many applications, pre-shape data will be tightly localized around a mean shape, in such cases, the tangent space to the preshape hypersphere located at the mean shape will be a good approximation to the preshape space, as in figure 4. By linearizing our distribution in this manner, we take advantage of standard multivariate statistical analysis techniques, representing our shape distribution as a Gaussian. In cases where the data is very spread out, one can use a *complex Bingham distribution* [5].

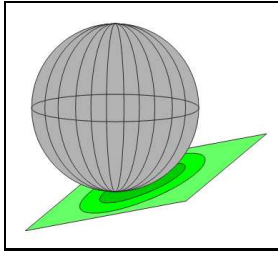


Figure 4. Tangent space distribution

In order to transform our set of pre-shapes into an appropriate tangent space, we first compute a mean shape, μ as above⁴. We then fit the observed pre-shapes to μ , and project each fitted pre-shape into the tangent space at μ . The tangent space coordinates for pre-shape τ_i are given by

$$v_i = (I - \mu\mu^H)e^{j\theta_i^*} \tau_i, \quad (13)$$

where $j^2 = -1$ and θ_i^* is the optimal Procrustes-matching rotation angle of τ_i onto μ . We can also now apply a dimensionality reduction technique such as PCA to the tangent space data v_1, \dots, v_n to get a compact representation of estimated shape distribution (Figure 4).

Figure 5 shows samples drawn from the learned generative model of the tool class along each eigenvector, in comparison to the mean shape. The first eigenshape largely captures how the shape deforms as the tool opens and closes, and the second and third eigenshapes largely capture how the shape deforms due to perspective geometry.

⁴We use μ to refer to the μ^* computed from the optimization in 7 for the remainder of the paper.

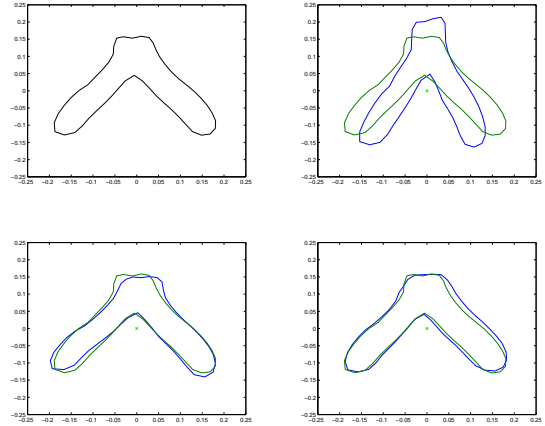


Figure 5. Shape model for a wire-cutter tool—the effects of the first three principle components (eigenvectors) on the mean shape. Top left: mean shape as in figure 3, top right: a sample shape from the first eigenshape, bottom left: a sample from the second eigenshape, bottom right: a sample from the third eigenshape. In each sampled shape, the mean shape is overlaid for comparison.

IV. SHAPE COMPLETION

We now turn to the problem of estimating the complete geometry of an object from an observation of part of its contour. We phrase this as a maximum likelihood estimation problem, estimating the missing points of a shape with respect to the Gaussian tangent space shape distribution.

Let us represent a shape as:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \quad (14)$$

where $\mathbf{z}_1 = \mathbf{m}$ contains the p points of our partial observation of the shape, and \mathbf{z}_2 contains the $n - p$ unknown points that complete the shape. Given a shape distribution D on n points with mean μ and covariance matrix Σ , and given \mathbf{z}_1 containing p measurements ($p < n$) of our shape, our task is to compute the last $n - p$ points which maximize the joint likelihood, $P_D(\mathbf{z})$. In contrast to previous sections, we will now work in real, *vectorized* coordinates ($\mathbf{z} = (x_1, y_1, \dots, x_n, y_n)^T$) rather than in the complex coordinates which were useful for encoding rotation.

In order for us to transform our completed vector, $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2)^T$, into a pre-shape, we must first normalize translation and scale. However, this cannot be done without knowing the last $n - p$ points. Furthermore, the Procrustes minimizing rotation from \mathbf{z} 's pre-shape to μ depends on the missing points, so any projection into the tangent space (and corresponding likelihood) will depend in a highly non-linear way on the location of the missing points. We can, however, compute the missing points \mathbf{z}_2 given an orientation and scale. This leads to an iterative algorithm that holds the orientation and scale fixed, computes \mathbf{z}_2 and then computes a new orientation and scale given the new \mathbf{z}_2 . The translation term can then be computed from the completed contour \mathbf{z} .

We derive \mathbf{z}_2 given a fixed orientation θ and scale α in the following manner. For a complete contour \mathbf{z} , we normalize

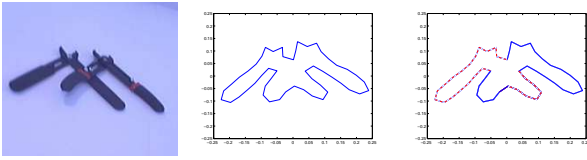


Figure 6. An example of occluded objects, where the right tool occludes the left tool. Left to right: The original image, the measured contour, the contour segments that must be completed.

for orientation and scale using

$$\mathbf{z}' = \frac{1}{\alpha} R_\theta \mathbf{z} \quad (15)$$

where R_θ is the rotation matrix of θ . To center \mathbf{z}' , we then subtract off the centroid:

$$\mathbf{w} = \mathbf{z}' - \frac{1}{n} C \mathbf{z}' \quad (16)$$

where C is the $2n \times 2n$ checkerboard matrix,

$$C = \begin{bmatrix} 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 0 & 1 \end{bmatrix}. \quad (17)$$

Thus \mathbf{w} is the centered pre-shape. Now let M be the tangent space projection matrix:

$$M = I - \mu \mu^T \quad (18)$$

Then the Mahalanobis distance with respect to D from $M\mathbf{w}$ to the origin in the tangent space is:

$$d_\Sigma = (M\mathbf{w})^T \Sigma^{-1} M\mathbf{w} \quad (19)$$

Minimizing d_Σ is equivalent to maximizing $P_D(\cdot)$, so we continue by setting $\frac{\partial d_\Sigma}{\partial \mathbf{z}_2}$ equal to zero, and letting

$$W_1 = M_1(I_1 - \frac{1}{n} C_1) \frac{1}{\alpha} R_\theta^1 \quad (20)$$

$$W_2 = M_2(I_2 - \frac{1}{n} C_2) \frac{1}{\alpha} R_\theta^2 \quad (21)$$

where the subscripts “1” and “2” indicate the left and right sub-matrices of M , I , and C that match the dimensions of \mathbf{z}_1 and \mathbf{z}_2 . This yields the following system of linear equations which can be solved for the missing data, \mathbf{z}_2 :

$$(W_1 \mathbf{z}_1 + W_2 \mathbf{z}_2)^T \Sigma^{-1} W_2 = 0 \quad (22)$$

As described above, equation 22 holds for a specific orientation and scale. We can then use the estimate of \mathbf{z}_2 to re-optimize θ and α and iterate. Alternatively, we can simply sample a number of candidate orientations and scales, complete the shape of each sample, and take the completion with highest likelihood (lowest d_Σ).

To design such a sampling algorithm, we must choose a distribution from which to sample orientations and scales. One idea is to match the partial shape, \mathbf{z}_1 , to the partial mean shape, μ_1 , by computing the pre-shapes of \mathbf{z}_1 and μ_1 and finding the Procrustes fitting rotation, θ^* , from the pre-shape of \mathbf{z}_1 onto the pre-shape of μ_1 . This angle can then be used as a mean for

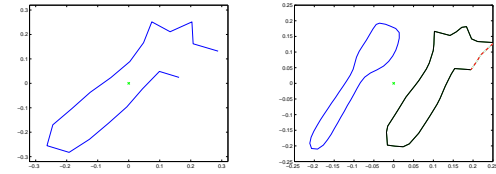
a von Mises distribution (the circular analog of a Gaussian) from which to sample orientations. Similarly, we can sample scales from a Gaussian with mean α_0 —the ratio of scales of the partial shapes \mathbf{z}_1 and μ_1 as in

$$\alpha_0 = \frac{\|\mathbf{z}_1 - \frac{1}{p} C_1 \mathbf{z}_1\|}{\|\mu_1 - \frac{1}{p} C_1 \mu_1\|}. \quad (23)$$

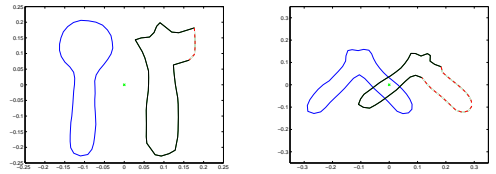
Any sampling method for shape completion will have a *scale bias*—completed shapes with smaller scales project to a point closer to the origin in tangent space, and thus have higher likelihood. One way to fix this problem is to solve for \mathbf{z}_2 by performing a constrained optimization on d_Σ where the scale of the centered, completed shape vector is constrained to have unit length:

$$\|x' - \frac{1}{n} C x'\| = 1. \quad (24)$$

This constrained optimization problem can be attacked with the method of Lagrange multipliers, and reduces to the problem of finding the zeros of a $(n-p)$ th order polynomial in one variable, for which numerical techniques are well-known. In preliminary experiments this scale bias has not appeared to provide any obvious errors in shape completion, although more rigorous testing and analysis are needed.



(a) Partial contour to be completed (b) Completion using fork model



(c) Completion using spoon model (d) Completion using tool model

Figure 7. Shape classification of partial contours. In the second, third and fourth image, the mean shape of the shape class is shown in blue on the left, and the approximate maximum likelihood completion to the partial contour is shown on the right.

V. SHAPE CLASSIFICATION

Given k previously learned shape classes C_1, \dots, C_k with shape means μ_1, \dots, μ_k and covariance matrices $\Sigma_1, \dots, \Sigma_k$, and given a measurement \mathbf{m} of an unknown object shape, we can now compute a distribution over shape classes for a measured object: $\{P(C_i | \mathbf{m}) : i = 1 \dots k\}$. The shape classification problem is to find the posterior mode of this distribution, \hat{C} ,

$$\hat{C} = \arg \max_{C_i} P(C_i | \mathbf{m}) \quad (25)$$

$$= \arg \max_{C_i} P(\mathbf{m} | C_i) P(C_i). \quad (26)$$

Assuming a uniform prior on C_i , we have the standard maximum likelihood (ML) estimator

$$\hat{C} = \hat{C}_{ML} = \arg \max_{C_i} P(\mathbf{m}|C_i). \quad (27)$$

A. Point Correspondences

The difficulty in proceeding with this ML problem lies in corresponding the measurement \mathbf{m} with each object model, that is, identifying which points in \mathbf{m} correspond to which points in model C_i . One potential drawback to contour-based shape analysis (compared to other representations such as point cloud) is that we require correspondences between all points of any two shapes we wish to compare using the Procrustean metric⁵. While this caveat is clearly the major source of power to our model, it can also be a major difficulty in real-world problems where correspondences can be difficult to identify.

The correspondence problem also requires determining whether \mathbf{m} is a complete, or partially occluded measurement. These problems are complicated even further when \mathbf{m} is allowed to be a measurement of more than one object (Figure 6). In this case, we have a segmentation problem as well as a correspondence problem. However, it is reasonable to assume that segmentation will be made easier by using stereo vision or range data, and can in many cases be handled in a pre-processing stage. A full treatment of data correspondence is beyond the scope of this paper; constraining ourselves to the case where \mathbf{m} is a measurement of a single object, we will address the following two cases:

- 1) \mathbf{m} is a complete measurement of a single object.
- 2) \mathbf{m} is a *simply*, partially occluded measurement of a single object.

By *simply* occluded we mean that \mathbf{m} consists of one, contiguous segment of a full (closed) contour.

For the 2-D contour case, one method for finding point correspondences is to identify local features of interest on each contour (for example, based on curvature or local changes in shape) and match the higher-level features to similar sets of features on other contours, interpolating correspondences for the points between features.

For our experiments, we implemented a feature correspondence algorithm for closed contours based on a hierarchical probabilistic model incorporating local feature match likelihoods together with the global feature shape match likelihood—this was done primarily to demonstrate that a purely geometric feature correspondence algorithm was tenable. However, in practice algorithms incorporating more robust features (such as SIFT features) are more desirable than purely geometric algorithms. The details of this feature correspondence algorithm are omitted due to space constraints.

B. Complete Shape Class Likelihood

Given that a measurement \mathbf{m} represents a complete contour, the problem of finding the maximum likelihood estimate of each shape class, \hat{C}_{ML} , is relatively straight-forward. First, calculate the pre-shape τ associated with \mathbf{m} . Then, for each

class $C_i : i = 1 \dots k$, find correspondences between τ and the mean shape μ_i of model C_i and normalize τ to have the same number of points as μ_i , with the points around τ 's contour corresponding as closely as possible to the points around the contour of μ_i . Next, re-normalize this feature-matched τ so that it lies in pre-shape space (i.e. center at the origin, length one), and call this re-normalized pre-shape τ' . Finally, compute the orthogonal Procrustes fit of τ' onto μ_i , giving orientation θ^* , and project into tangent space so that

$$\mathbf{z} = MR_{\theta^*}\tau', \quad (28)$$

where again R_{θ^*} is the rotation matrix of θ^* . The probability of this point x in tangent space is then given with the standard Gaussian likelihood,

$$P(\mathbf{m}|C_i) \approx P(\mathbf{z}|C_i) = \frac{1}{\sqrt{(2\pi)^{2n}|\Sigma_i|}} \exp\left(-\frac{1}{2}\mathbf{z}^T \Sigma_i^{-1} \mathbf{z}\right). \quad (29)$$

C. Partial Shape Class Likelihood

The most obvious approach to partial shape class likelihood is to simply complete the missing portion of the partial shape corresponding to \mathbf{m} with respect to each shape class, then classify the completed shape as above (Figure 7). To make this concrete, let $\mathbf{z} = \{\mathbf{z}_1, \mathbf{z}_2\}$ be the completed shape, where \mathbf{z}_1 is the partial shape corresponding to measurement \mathbf{m} , and \mathbf{z}_2 is unknown. Then

$$P(C_i|\mathbf{z}_1) = \frac{P(C_i, \mathbf{z}_1)}{P(\mathbf{z}_1)} \propto \int P(C_i, \mathbf{z}_1, \mathbf{z}_2) d\mathbf{z}_2 \quad (30)$$

Rather than marginalize over the hidden data, \mathbf{z}_2 , we can approximate with estimate $\hat{\mathbf{z}}_2$, the output of our shape classification algorithm, yielding:

$$P(C_i|\mathbf{z}_1) \approx \eta \cdot P(\mathbf{z}_1, \hat{\mathbf{z}}_2|C_i) \quad (31)$$

where η is a normalizing constant (and can be ignored during classification), and $P(\mathbf{z}_1, \hat{\mathbf{z}}_2|C_i)$ is the complete shape class likelihood of the completed shape.

There are several variables of the shape completion algorithm which may influence the accuracy of the completed shape, e.g. feature correspondences, sample rotations and sample scales. As there is some randomness associated with these variables, they should ideally be marginalized out from the partial shape class likelihood. Additionally, we may want to include a prior on the number of points to be completed as this number must be estimated with a partial shape feature correspondence algorithm. Such a prior would indicate the common-sense principle that shape completions with fewer unknown points are to be preferred over shape completions with a large number of unknown points.

VI. EXPERIMENTAL RESULTS

Videos of a controlled desktop environment were processed with routines from the Intel Open Source Video Project (OpenCV). Contours were extracted from the raw image frames using intensity thresholding and pyramid segmentation edges (as in figure 1). Shape models of three object classes—forks, spoons, and wire-cutter tools—were then generated from

⁵We also require that the number of points on each shape be the same.

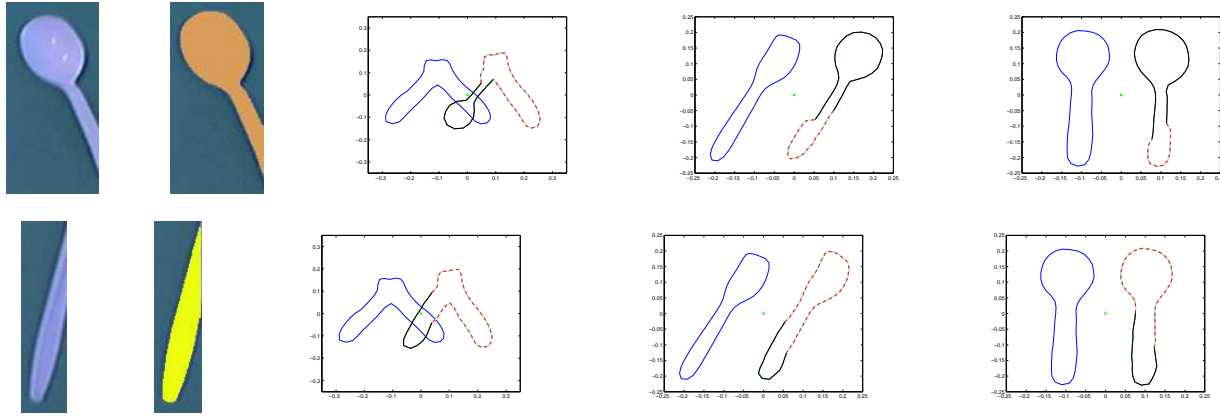


Figure 8. Partial contour shape classification. Top row: the extracted contour is correctly classified as a spoon. Left to right: original image, extracted contour, shape completion using tool, fork, and spoon models. Bottom row: extracted contour is misclassified as a spoon; however, partial contour is uninformative, even for humans. Left to right: original image, extracted contour, shape completion using tool, fork, and spoon models.

a manually-chosen subset of the extracted contours. An illustration of the tool shape model is shown in figure 5.

Complete contour shape classification was then tested on a subset of the training data and, as expected, the classification rate was 100% (Figure 3). Next, eight partial contours (three of tools, three of forks, and two of spoons) were extracted from video frames of occluded objects. Partial contours were segmented by hand (where needed), completed with respect to each shape model, and classified (Figure 8). This process was repeated while varying N , the number of principle components kept in each shape model, from one to ten. For $N < 5$, the classification rate was 5/8, while for $N = 5$ and above, the classification rate was 6/8. Incorporating a prior on the completion size (as in the previous section) increased the classification rate to 7/8 for $N > 7$. The one partial shape which was consistently misclassified was the end of a fork handle, impossible even for a human to classify correctly more than one-third of the time (figure 7, lower row).

VII. RELATED WORK

There is a great deal of work on statistical shape modelling, beginning with the work on landmark data by Kendall [10] and Bookstein [4] in the 1980's. In recent years, more complex statistical shape models have arisen, for example, in the active contours literature [3]. Procrustes analysis predates statistical shape theory by two decades; algorithms for finding Procrustean mean shapes [6], [2] were developed long before the topology of shape spaces were well-understood [11]. In terms of shape classification, shape contexts [1] and spin images [8] provide robust frameworks for estimating correspondences between shape features for recognition and modelling problems. An interesting take on shape completion using probable object symmetries has been done in [14].

VIII. CONCLUSION

We have presented an approach to geometric object modelling that unified the problems of modelling geometry, recognizing object shapes and inferring occluded portions of the object model. The algorithm depends on a technique known

as Procrustean shape analysis [10], and in particular we have derived an expression for the maximum likelihood object geometry given only a partial observation of the object. We have shown some preliminary results on everyday objects, but we plan to extend these results to more complex scenes and geometric inference problems.

The description of our algorithm given in this paper is restricted to inferring geometry from two-dimensional images. However, the technique extends to higher dimensions, and we plan to demonstrate the same approach to object modelling on three-dimensional object data, such as from a laser range finder or stereo camera.

REFERENCES

- [1] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(24):509–522, April 2002.
- [2] Jos M. F. Ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, June 1977.
- [3] A. Blake and M. Isard. *Active Contours*. Springer-Verlag, 1998.
- [4] F.L. Bookstein. A statistical method for biological shape comparisons. *Theoretical Biology*, 107:475–520, 1984.
- [5] I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1998.
- [6] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, March 1975.
- [7] John R. Hurlley and Raymond B. Cattell. The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7:258–261, 1962.
- [8] Andrew Johnson and Martial Hebert. Using spin images for efficient object recognition in cluttered 3-d scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):433 – 449, May 1999.
- [9] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3d shape descriptors, 2003.
- [10] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math Soc.*, 16:81–121, 1984.
- [11] D.G. Kendall, D. Barden, T.K. Carne, and H. Le. *Shape and Shape Theory*. John Wiley and Sons, 1999.
- [12] David G. Lowe. Object recognition from local scale-invariant features. In *Proc. of the International Conference on Computer Vision ICCV, Corfu*, pages 1150–1157, 1999.
- [13] C. G. Small. *The statistical theory of shape*. Springer, 1996.
- [14] S. Thrun and B. Wegbreit. Shape from symmetry. In *Proceedings of the International Conference on Computer Vision (ICCV)*, Beijing, China, 2005. IEEE.