What does the occluding contour tell us about quantitative shape?

Yiming Qian  
Institute of High Performance Computing (IHPC)  
Agency for Science, Technology and Research (A*STAR)  
Singapore  
qiany@ihpc.a-star.edu.sg

James H. Elder  
Centre for AI & Society  
Centre for Vision Research  
York University, Toronto, Canada  
jelder@yorku.ca

Abstract—The local shape of the occluding contour of an object is known to constrain the local shape of the object surface [1], however, these known constraints are qualitative. Here we posit that in addition to these qualitative constraints, typical regularities of common objects and rules of projection induce dependencies that can be used to derive statistical estimates of quantitative solid shape from the occluding contour. To explore this conjecture, we partition the problem into two parts: 1) Estimation of the 3D rim from the 2D occluding contour, and 2) Estimation of the visible surface shape from the estimated 3D rim. We train and evaluate a number of statistical models on two distinct 3D object datasets and demonstrate that capturing these statistical regularities leads to better estimates of 3D shape than existing methods.

Keywords—shape; single-view 3D; statistical models; auto-encoders

I. INTRODUCTION

Multi-view methods for 3D object shape estimation such as stereopsis and motion parallax provide direct depth cues via triangulation, however these methods have limitations. Their accuracy is inversely proportional to the square of the distance to the object, making them less effective for distant objects, and they can fail when surface texture is too faint to generate features that can be tracked reliably across frames.

These limitations highlight the value of single-view cues such as shape from shading and texture. However, these surface cues also have limitations for 3D surface reconstruction. While providing useful local information about qualitative surface shape, global shape estimation is subject to depth sign (convex/concave) and more general bas-relief ambiguities [2]–[5]. This raises the question of whether the shape of the boundary of the object can also inform the single-view estimation of 3D shape, complementing potentially weak surface cues.

In 1984, Koenderink wrote the seminal paper ‘What does the occluding contour tell us about solid shape?’, in which he pointed out important qualitative relationships between the local shape of the occluding contour in the image and the local shape of the object surface [1]. However, this paper does not speak to whether the occluding contour can tell us anything quantitative about solid shape. While strict quantitative constraints relating the occluding contour to solid shape are unlikely, we posit here that typical regularities of common objects and rules of projection induce dependencies that can be used to derive statistical estimates of quantitative solid shape from the occluding contour.

We follow Koenderink [1] and refer to the set of surface points grazed by the view vector as the 3D rim of the object, and the perspective projection of these points onto the image as the 2D occluding contour. Koenderink noted that the occluding contour provides strong local constraints on quantitative surface shape at the rim: convex points on the occluding contour must project from convex points on the 3D surface, while concave points on the occluding contour must project from saddle points [1].

Here we consider whether the 2D occluding contour can also provide useful quantitative information about the global shape and pose of an object. We are motivated by evidence that human judgement of surface shape is strongly influenced by the shape of the occluding contour [2], [5] - often the occluding contour alone is enough to provide a compelling sense of volumetric shape [6], [7] (Fig. 1(a)). Moreover, most natural and artificial objects possess symmetries [8], [9] that will be inherited to some degree by the 3D rim, and the distortions of those symmetries induced by perspective projection should provide cues to depth.

![Figure 1: (a) The occluding contour can evoke a strong sense of solid shape. (b) Puffball reconstruction [10], [11].](image)

The topologies of both the occluding contour and 3D rim can in general be quite complex; here we make two simplifications. First, we ignore self-occlusions, where the view vector both grazes and pierces the object, restricting our attention to the boundary of the object. One of the motivations for doing this is that self-occlusions can be trickier to detect in real images, since the figure and ground often have similar illumination, colour and texture. Second,
we ignore holes in the object projections, further focusing our attention on the outer boundary of the object. We leave analysis of self-occlusions and more complex topologies for future work.

With these simplifications, we partition the problem of estimating 3D shape from the occluding contour into two parts: 1) Estimation of the 3D rim from the 2D occluding contour, and 2) Estimation of the visible surface shape from the estimated 3D rim.

A. Estimating the 3D rim from the 2D occluding contour

In what follows we will first demonstrate a statistical link between the shape of the occluding contour and depth variation in the 3D rim through a simple intuitive model that links the depth of a point on the rim to the distance of the corresponding point on the occluding contour from the object’s centre of mass. We then explore multivariate normal and autoencoder models that more completely capture this statistical relationship. We show that a statistical model that links both the position and the tangent of the occluding contour to the depth of the rim yields superior results.

We note that the 3D rim may be useful in its own right for certain applications, including free space estimation in autonomous navigation and surface contact point selection for robotic grasp planning. We also note that human stereoscopic 3D perception of an object is strongly driven by disparity cues at the object boundary [12]. This suggests that accurate estimation of the 3D rim will be crucial for 2D to 3D film conversion.

B. Estimating surface shape from the 3D rim

Can the estimated 3D shape of the rim be used on its own, i.e., without direct surface cues, to deliver information about the 3D shape of the visible object surface? To explore this question, we develop and evaluate a generalization of the puffball approach [11]: a 3D object is completed by the union of maximal osculating spheres tangent to the estimated 3D rim.

C. Summary of contributions

In summary, we make five specific contributions:

1) We introduce two novel datasets consisting of 3D object rims and their 2D projections. We will make these datasets public to encourage continuing research on 3D shape from contour.
2) We demonstrate a statistical connection between the 3D shape of the object rim and the observable 2D shape of its occluding contour, and capture this relationship with a series of novel statistical models.
3) We show that these models can be used to make predictions of the depth variation in the 3D object rim from the 2D occluding contour alone.
4) We introduce a novel spherical completion approach for reconstructing the visible surface based solely on the estimated 3D rim.
5) We show that our approach yields more accurate estimates of 3D object shape and pose than competing approaches [11], [13].

II. RELATED WORK

Early computer vision algorithms for single-view 3D shape estimation exploited the occluding contour in conjunction with surface cues within an optimization framework. Typically, user interaction and/or some inflation term in the objective function was required to avoid a trivial (flat) solution. Users were required to specify the depth of some surface points [14], a fixed volume that the shape must fill [15], or an inflation term that indirectly determines the volume [16].

A very different strategy for single-view estimation of smooth solid shapes was introduced in the interactive sketching interface dubbed Teddy [10] and later studied by Twarog et al. [11] under the name Puffball. In this approach, the solid shape is defined as the union of spheres centred on the interior skeleton [17] of the shape in the image, and bi-tangent to the occluding contour (Fig. 1(b)). The method is simple and can produce surprisingly reasonable results in some cases. However, a major limitation of this approach is that the 3D rim of the object is assumed to be planar and fronto-parallel, which in general will not be the case.

More recent approaches to single-view 3D object reconstruction have managed to avoid user interaction or arbitrary inflation terms. Shape Collage [18] employed a non-parametric example-based approach for local surface patch estimation within an MRF framework and a thin-plate model to integrate the local patches into a global shape. They trained and evaluated their method on synthetic images of random 3D ‘blob’ objects. They reported results for shaded and textured versions of the objects, and also for line drawings that include the object boundary but also interior occluding and ‘suggestive’ contours, i.e., contours that follow points at which the view vector and surface normal are close to orthogonal (and so the view vector ‘almost’ grazes the surface). They reported a 30 deg RMS error in surface normal estimation. It is unclear to what degree performance depended upon internal contours vs the object boundary.

In their Shape, Illumination and Reflectance from Shading (SIRFS) approach [13], Barron & Malik adopted a probabilistic framework, employing priors over shape, reflectance and illumination together with surface constraints imposed by the boundary of the object. They trained and evaluated their approach on the MIT intrinsic images dataset of 20 3D objects [19], and reported results of their algorithm that used only the object boundary, reporting a 24 deg mean absolute error (MAE) in surface normal estimation.

It is difficult to directly compare the errors reported for Shape Collage and SIRFS, given the differences in the datasets and measures of error (RMS vs MAE). Here we
note that when errors are distributed normally, $E[MAE] = \sqrt{2/\pi}E[RMS]$, which is roughly the ratio of errors reported by these authors.

Karsch et al [20] extended the SIRFS probabilistic framework to include surface normal constraints at internal geometric contours (self-occlusions and folds). On the full 20-object MIT intrinsic image database, they reported a 29.9 deg MSE in surface normal estimation using the SIRFS object-boundary method, consistent with the results above. This error was found to drop by about 1 deg when either self-occlusions or folds were incorporated, and declined to 27.6 deg when both cues were employed. Interestingly, they found that reconstructions from the silhouette benefitted more from the addition of these internal contours than from smooth shading. In fact, once internal contours were incorporated, adding shading cues was found to lower performance. This highlights the importance of contour cues for single-view 3D reconstruction.

A distinct branch of research explores the problem of class-conditional single-view 3D object reconstruction, which allows strong within-category regularities to be exploited [21]. The more recent state of the art algorithms are deep auto-encoders. While typically trained and evaluated on a small number of object classes (e.g., chairs, cars, planes [22]), newer versions have been shown to generalize well to new object classes [23]–[25].

Here we are not trying to compete with these auto-encoder methods, which make use of many training instances of a small number of objects categories, and use all of the cues (e.g., shading, texture, self-occlusions, part structure) afforded by the colour imagery. Rather, we focus on the scientific question of whether the occluding contour carries quantitative information about 3D object shape, and if so, how that information can be harnessed. It is our hope that in the long run, a better understanding of the information afforded by the occluding contour will ultimately lead to better (and more explainable) multi-cue algorithms for single-view 3D object reconstruction.

Generally, prior work has focused on orthographic projection, which is unrealistic and also ignores important 3D information available in the distortions induced by perspective projection. A specific contribution of the present paper is to examine the 3D information afforded by the occluding contour when viewed in perspective.

**III. DATASETS**

We employ two datasets of solid 3D objects. The first dataset comprises 122 scanned objects originally employed by Mehrani et al [26] and obtained directly from the authors. We randomly split the Mehrani dataset into training and test sets of 61 objects each. For each FOV we generated 1,000 random image projections for each training object and 20 random projections for each test object.

The second dataset is the ShapeNet Core [27] dataset of 52,472 synthetic objects. We split the ShapeNet Core dataset objects into random training (60%), validation (20%) and 20% test (20%) partitions. For each FOV we generated 20 random image projections for each training object and 1 random projection for each test object. Samples from both datasets are shown in Fig. 2.

Perspective projections were formed using a virtual pinhole camera with unit focal length and field of view (FOV) $\in (2, 4, 8, 16, 32, 64)$ deg (Fig. 3). We employ a camera-centred coordinate frame with $Z$ representing distance from the lens plane along the optical axis. Objects were centred on the optic axis with centroids at a depth of $Z = 1$. (Note that under perspective projection, points on the rim will have average depth $Z < 1$.) The size of each object was adjusted so that the object was just contained within the field of view, granting the frustrum at at least one point. The rim was sampled at 32 points with equal arc-length separation, and these 32 points were projected to the image to form the occluding contour. Each 3D rim $\gamma_3(s) = (X(s), Y(s), Z(s))$ is thus a $32 \times 3$ matrix, and the corresponding 2D occluding contour $\gamma_2(s) = (x(s), y(s))$ is a $32 \times 2$ matrix.

To facilitate learning, we rotated all contours in the image.
plane to align the first principal component with the x-axis of the image, with the taller side on the left (i.e., points with \( x < 0 \) have greater y-variance then points with \( x > 0 \)).

IV. Estimating the 3D rim from the 2D occluding contour

A. Eccentricity model

Why do expect the occluding contour to carry information about depth variation in the 3D rim? In explaining perspective projection to a child, one might start with the fact that as objects get closer to the eye they get bigger in the image. Applying the inverse of this logic to the bounding contour, we might predict that more eccentric points on the occluding contour i.e., points that are further from the centre of mass project from points on the rim that are closer to the eye (Fig. 4 (a)).

![Figure 4: The eccentricity model for estimating distance \( Z(s) \) from the image plane. (a) The cue is the squared distance \( r(s)^2 \) of the occluding contour from the centre of mass of the contour in the image. (b-c) Histogram models for the Mehrani and ShapeNet datasets.](image)

To explore this idea, we examine the statistical relationship between eccentricity \( r(s) \) of points on the occluding contour and the depth \( (Z(s)) \) values of the corresponding points on the rim. Empirically, we find that the relationship between \( Z(s) \) and \( r(s)^2 \) is nearly linear, and so to form a model we bin the ground truth \( Z \) values from our training datasets as a function of \( r(s)^2 \). The bin width is selected to minimize leave-one-out cross-validation error over objects for the Mehrani dataset, and error on the validation partition for the ShapeNet Core dataset.

The systematic relationship between eccentricity and depth can be seen in the resulting histograms - Figs. 4(b-c) show the results for 64 deg FOV. This clearly demonstrates that the occluding contour carries information about depth variation in the 3D rim. We explore the performance of this simple model in our Evaluation section below, but intuitively it seems unlikely that the eccentricity model captures all of the statistical information that the occluding contour can provide. We therefore turn now to more general statistical models that we hope can more fully capture this relationship.

B. Normal models

When exploring a statistical relationship for the first time it is natural to consider a normal model, and here we consider two. In our base model, we assume that the occluding contour \( \alpha(s) = (x(s), y(s)) \) is jointly normal with the unknown depth coordinate \( Z(s): (\alpha(s), Z(s)) \sim \mathcal{N}(\alpha(s), Z(s); \mu_1, \Sigma_1) \). We model the covariance across all pairs of points \( (s_1, s_2) \) on the occluding contour and rim. As a result the model contains a 96-dimensional mean vector \( \mu_1 \) and a 96 × 96 covariance matrix \( \Sigma_1 \).

We also explore a second, augmented normal model. The prevalence of orientation regularities such as parallelism and rectilinearity, and the importance of linear perspective in human perception, suggests that the local orientation of the occluding contour may also be an important cue to depth. While the tangent vector \( t(s) \) of the occluding contour is implicitly defined by the contour \( \alpha(s) = (x(s), y(s)) \) itself, the linear nature of the normal model may limit its capacity to capture the influence of the tangent on the depth of the rim. In our second model, we therefore augment the occluding contour coordinates with the tangent vector \( t(s) = (t_x(s), t_y(s)): (\alpha(s), t(s), Z(s)) \sim \mathcal{N}(\alpha(s), t(s), Z(s); \mu_2, \Sigma_2) \). This model consists of a 160-dimensional mean vector \( \mu_2 \) and a 160 × 160 covariance matrix \( \Sigma_2 \). Maximum likelihood estimates of these parameters are estimated from the training data. The parameterized models can then be used for inference: Given a partially observed test vector \( \alpha(s) \) (base model) or \( (\alpha(s), t(s)) \) (augmented model), the expectation of the unobserved depth \( Z(s) \) can be estimated using the standard conditional expectation formula ( [28], Eqn. 2.81):

\[
\mu_{Z|u} = \mu_Z + \Sigma_{Zu} \Sigma_u^{-1} (u - \mu_u)
\]

where \( u \) represents \( \alpha(s) = (x(s), y(s)) \) for the base model and \( (\alpha(s), t(s)) \) for the augmented model.

C. Auto-encoder model

Given the success of auto-encoders for pixel-wise single-view 3D object reconstruction, it is natural to consider an auto-encoder for estimating depth variation in the 3D rim from the occluding contour. We trained the auto-encoder to minimize squared error in depth \( Z \). For both Mehrani and ShapeNet datasets we found empirically that a simple architecture with just one encoder layer and one hidden layer performs best: larger architectures tend to reduce performance on unseen data.

Given their difference in size, we optimize the number of units in each layer separately for the two datasets. We perform this optimization by maximizing the correlation between the ground truth and the estimated depth for a field of view of 64 deg, using leave-one-out cross-validation on the the Mehrani training dataset and the ShapeNet validation dataset. For both datasets, we found a 128-unit encoder layer to be optimal. For the Mehrani dataset, we found an 48-unit
hidden layer to be optimal, while for the ShapeNet dataset we found 16 hidden units to be optimal.

V. ESTIMATING SURFACE SHAPE FROM THE 3D RIM

Can we use an estimate of the 3D rim to generate an estimate of the visible object surface? In the puffball approach [10, 11], the solid shape is defined as the union of spheres centred on the interior skeleton [17] of the shape in the image, and bi-tangent to the occluding contour (Fig. 1(b)). A major limitation of this approach is that the 3D rim of the object is assumed to be planar and fronto-parallel, which in general will not be the case. Here we propose a generalization of the puffball method that can be applied to oblique non-planar 3D rims. The method produces a 3D volumetric estimate of the object (Fig. 5).

![Figure 5: Spherical surface completion.](image)

For each pixel \( i \) of the image interior to the occluding contour we identify the maximal inscribing circle centred on the pixel. This circle lies within the shape, is tangent to at least one point \( \gamma_2(s_i) \) on the occluding contour, and defines an elliptical cone with axis passing through the optical centre and the pixel \( i \). This cone is tangent to the rim at the corresponding rim point \( \gamma_3(s_i) \). We identify the unique maximal sphere that lies within this cone and is tangent to the cone at \( \gamma_3(s_i) \); note that the projection of this sphere is the inscribing circle. The union of these spheres over all pixels \( i \) in the interior of the 2D occluding contour defines our estimate of the solid 3D shape. The surface normal of the estimated shape will be orthogonal to the view vector at each rim point, as required for a smooth solid object.

In practice, the computation involves identifying the orthogonal projection of each of the \( k \) 3D rim points onto each of the \( n \) interior pixel rays to define \( kn \) tangent spheres. Then we must check the distance of all other \( k-1 \) rim points from the centre of each sphere, eliminating any spheres that subsume other rim points, and therefore identifying the unique maximal inscribing sphere for each pixel. The computation thus has complexity \( k^2 n \).

Since the 3D locations and radii of the inscribing spheres tend to vary smoothly in a local pixel neighbourhood, the spherical completion method generates smooth solid completions. This will work well for smooth objects, but less well for objects with sharp folds at the rim.

VI. EVALUATION

A. Estimating the 3D rim from the 2D occluding contour

We first evaluate our models for estimating the 3D rim from the occluding contour. As perspective projection cues to depth generally increase with FOV, we first train and evaluate the accuracy of these models as a function of FOV, on the Mehrani training and test sets, respectively. We vary the FOV over \([2, 4, 8, 16, 32, 64]\) deg. As a measure of performance we use the Pearson correlation between the ground truth depth values \( Z \) and the estimated depth values \( \hat{Z} \) over all points on the rim. Fig. 6 shows that correlation increases monotonically for all methods as a function of FOV. We find that the normal and auto-encoder models outperform the simple eccentricity model by a large margin. The advantage of the augmented normal model over the base normal model confirms our intuitions that the local orientation of the occluding contour is informative about the depth of the rim. Interestingly, the auto-encoder model was found to underperform the normal models.

![Figure 6: Pearson correlation between the ground truth depth values \( Z \) and the estimated depth values \( \hat{Z} \) over all points on the rim, averaged over objects in the Mehrani test dataset.](image)

To provide a qualitative feel the 3D rim estimates, Fig. 7 shows best, median and worst case estimates of the rim depth \( Z \) (as measured by correlation) for test shapes in the Mehrani dataset. In the best case the estimate is excellent and the correlation almost perfect, but in the worst case the algorithm flips the sign of the depth. The median case is most representative: the low frequency trend of the rim depth is captured, but finer details are lost and the amplitude of depth variation is attenuated.

To explore the generality of these approaches, we also train and evaluate on the ShapeNet dataset, focusing on a FOV of 64 deg. close to the FOV for the standard lens of a typical smart phone camera. Table I shows performance of all models on both Mehri and ShapeNet test sets, trained on their respective training sets. Interestingly, we see that the auto-encoder model performs better than the normal models on the Shapenet dataset. This may be due to the fact that the ShapeNet objects tend to be less smooth, resulting in highly non-linear statistical dependencies that are more easily captured by the multi-layer auto-encoder.
Table I: Within-dataset Pearson correlation between the ground truth depth values $\hat{Z}$ and the estimated depth values $\tilde{Z}$ over all points on the rim, mean±std. err. over test objects, for 64 deg FOV.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mehrani</th>
<th>ShapeNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eccentricity</td>
<td>0.12 ± 0.008</td>
<td>0.16 ± 0.003</td>
</tr>
<tr>
<td>Base normal</td>
<td>0.26 ± 0.015</td>
<td>0.31 ± 0.004</td>
</tr>
<tr>
<td>Aug. normal</td>
<td>0.31 ± 0.015</td>
<td>0.33 ± 0.004</td>
</tr>
<tr>
<td>Autoencoder</td>
<td>0.27 ± 0.015</td>
<td>0.36 ± 0.004</td>
</tr>
</tbody>
</table>

We also examine generalization across the datasets, training on one and testing on the other (Table II). While we see a drop in performance in both cases, the drop is less profound when generalizing from the ShapeNet to the Mehrani dataset. This is unsurprising, given that the ShapeNet training dataset is much larger than the Mehrani training set, one would expect better generalization. But this difference may also be due to the greater diversity of the ShapeNet dataset. Since it contains both smooth and less smooth objects it support inference for both, whereas the Mehrani dataset, containing primarily smooth objects, could be expected to fail when presented with less smooth objects from the ShapeNet dataset.

Finally, we make use of the ShapeNet category labels to examine how 3D rim estimation accuracy may depend upon the type of object being viewed, and whether there is generalization across categories. We focus here on the augmented normal model, and consider three levels of generalization: (1) Train on all training data, (2) Train individually on the training set for each category, (3) Train on all categories but the test category (leave one out).

Fig. 8 shows the results ranked by performance when trained on all training data. We see substantial variation in performance across categories. Performance is generally better for smooth, single-part objects (e.g., clock, bottle, basket) that are more volumetric, and worst for objects that contain multiple parts or are less volumetric or contain parts that are almost 2D (e.g., bag, earphone, knife). On the other hand, we see almost no drop in performance when the test category is not included in training, indicating that the model is learning more general geometric principles rather than memorizing categories.

B. Estimating the surface shape from the 3D rim

To evaluate the potential of using the estimated 3D rim for surface reconstruction, we apply our spherical completion
method to 3D rims estimated using the eccentricity, base normal, augmented normal and auto-encoder models. We compare the results against the shape-from-contour version of SIRFS [13] and the Puffball method [10], [11]. We also evaluate an idealized model that applies the spherical completion method to the ground truth 3D rim: this provides an indication of how improvements to 3D rim estimation could influence the accuracy of surface completion. We evaluate all methods in terms of the mean Pearson correlation and RMS error between ground truth and estimated surface depth over the pixels on the object.

Table III shows the results. We find the eccentricity model to be weak. It does produce lower RMS error than SIRFS and Puffball, but only because it generates more conservative (flatter) shape estimates: the correlations with ground truth are substantially lower than both SIRFS and Puffball.

Table III: Pearson correlation and RMS error between the ground truth surface depth values and the estimated depth values over all pixels of the shape, mean±std. err. over test objects, for 64 deg FOV.

<table>
<thead>
<tr>
<th>Model</th>
<th>Mehrani</th>
<th>ShapeNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIRFS</td>
<td>0.60</td>
<td>0.28</td>
</tr>
<tr>
<td>Puffball</td>
<td>0.58</td>
<td>0.20</td>
</tr>
<tr>
<td>Eccentricity</td>
<td>0.43</td>
<td>0.15</td>
</tr>
<tr>
<td>Base normal</td>
<td>0.58</td>
<td>0.16</td>
</tr>
<tr>
<td>Aug. normal</td>
<td>0.60</td>
<td>0.17</td>
</tr>
<tr>
<td>Auto-encoder</td>
<td>0.58</td>
<td>0.16</td>
</tr>
<tr>
<td>Ground truth</td>
<td>0.78</td>
<td>0.12</td>
</tr>
</tbody>
</table>

However, we find that the normal and auto-encoder models perform substantially better than both SIRFS and Puffball on both Mehrani and ShapeNet datasets. The augmented normal model performs best on the Mehrani dataset, increasing correlation with ground truth by 8% over SIRFS and Puffball, and reducing RMS error by 52% and 29% over SIRFS and Puffball, respectively.

For the ShapeNet dataset, we find that the normal models and auto-encoder perform comparably, increasing correlation with ground truth by about 19% over SIRFS and Puffball, and reducing RMS error by 28% and 3% over SIRFS and Puffball, respectively.

At the same time, we see that much higher accuracy is achieved by the spherical completion model if based upon the ground truth 3D rim. This underlines the value in continuing research on the inference of 3D shape from the bounding contour.

The spherical completion method is appropriate for smooth objects, but not all objects are smooth. Is it possible given only the occluding contour to predict whether spherical completion should be applied? To explore this question, we analyze the accuracy of surface completion as a function of the maximum turning angle of the occluding contour (Fig. 9). We see that for both datasets, correlation for all methods is relatively high for small maximum turning angles but drops substantially as maximum turning angle increases. This suggests that the turning angle statistics can be used to help guide the selection of surface completion methods.

![Figure 8: Categorical evaluation on Augmented normal. (a) Trained on all data. (b) Trained and evaluated on each category independently. (c) Leave one out evaluation on each category.](image)

![Figure 9: Correlation of estimated and ground-truth surface depth for (a) Mehrani and (b) ShapeNet datasets, as a function of the maximum turning angle of the occluding contour.](image)
spherical completion from the augmented normal 3D rim estimate for best, median and worst case estimates of the rim depth $Z$ (as measured by correlation) for test shapes in the Mehrani dataset.

VII. CONCLUSIONS

In this work, we have introduced the problem of estimating the quantitative 3D shape of the object rim from its 2D occluding contour. Our results show that the shape of the occluding contour and depth variation in the rim are statistically related, and that Gaussian and auto-encoder models can capture this relationship. We have also introduced a novel spherical completion algorithm that allows an estimate of the visible surface to be reconstructed from the estimated 3D rim, and shown that this produces more accurate surface estimates than prior state-of-the-art approaches. Finally, we have shown that the maximum turning angle of the occluding contour can be used to predict the accuracy of the spherical completion method, which we hope can be used in future work to guide model selection for surface completion.

ACKNOWLEDGMENT

This research was supported by the NSERC Discovery program.

REFERENCES