Judging Whether Multiple Silhouettes Can Come from the Same Object

David Jacobs[1], Peter Belhumeur[2], and Ian Jermyn[3]

[1] NEC Research Institute
[2] Yale University
[3] New York University

Abstract. We consider the problem of recognizing an object from its silhouette. We focus on the case in which the camera translates, and rotates about a known axis parallel to the image, such as when a mobile robot explores an environment. In this case we present an algorithm for determining whether a new silhouette could come from the same object that produced two previously seen silhouettes. In a basic case, when cross-sections of each silhouette are single line segments, we can check for consistency between three silhouettes using linear programming. This provides the basis for methods that handle more complex cases. We show many experiments that demonstrate the performance of these methods when there is noise, some deviation from the assumptions of the algorithms, and partial occlusion. Previous work has addressed the problem of precisely reconstructing an object using many silhouettes taken under controlled conditions. Our work shows that recognition can be performed without complete reconstruction, so that a small number of images can be used, with viewpoints that are only partly constrained.

1 Introduction

This paper shows how to tell whether a new silhouette could come from the same object as previously seen ones. We consider the case in which an object rotates about a single, known axis parallel to the viewing plane, and is viewed with scaled orthographic projection. This is an interesting subcase of general viewing conditions. It is what happens when a person or robot stands upright as it explores a scene, so that the eye or camera is directed parallel to the floor. It is also the case when an object rests on a rotating turntable, with the camera axis parallel to the turntable.

It is easy to show that given any two silhouettes, and any two viewpoints, there is always an object that could have produced both silhouettes. So we suppose that two silhouettes of an object have been obtained to model it, and ask whether a third silhouette is consistent with these two images. We first characterize the constraint that two silhouettes place on an object's shape, and show that even when the amount of rotation between the silhouettes is unknown, this constraint can be determined up to an affine transformation. Next we show that for silhouettes in which every horizontal cross-section is one line segment, the

C. Arcelli et al. (Eds.): IWVF4, LNCS 2059, pp. 532–541, 2001.

question of whether a new silhouette is consistent with these two can be reduced to a linear program. Linear programming can also be used to test a necessary, but not sufficient condition for arbitrary silhouettes. We provide additional algorithms for silhouettes with cross-sections consisting of multiple line segments. We describe a number of experiments with these algorithms.

Much prior work has focused on using silhouettes to determine the 3D structure of an object. Some work uses a single silhouette. Strong prior assumptions are needed to make reconstruction possible in this case (eg., [10,1,4,9]). A second approach is to collect a large number of silhouettes, from known viewpoints, and use them to reconstruct a 3D object using differential methods (eg., [5], [3], [12], [11]) or volume intersection (eg., [8], [2]). These methods can produce accurate approximations to 3D shape, although interestingly, Laurentini[7] shows that exact reconstruction of even very simple polyhedra may require an unbounded number of images. Our current work makes quite different assumptions. We consider using a small number of silhouettes obtained from unknown viewpoints and ask whether the set of prior images and the new image are consistent with a single 3D shape without reconstructing a specific shape.

2 Constraints from Two Silhouettes

Let p, q, r denote the boundaries of three silhouettes. Let P, Q, R denote the filled regions of the silhouettes. When rotation is about the y axis, there will be two 3D points that appear in every silhouette, the points with highest and lowest y values. Denote the image of these points on the three silhouettes as: $p_1, p_2, q_1, q_2, r_1, r_2$. Let M denote the actual 3D object.

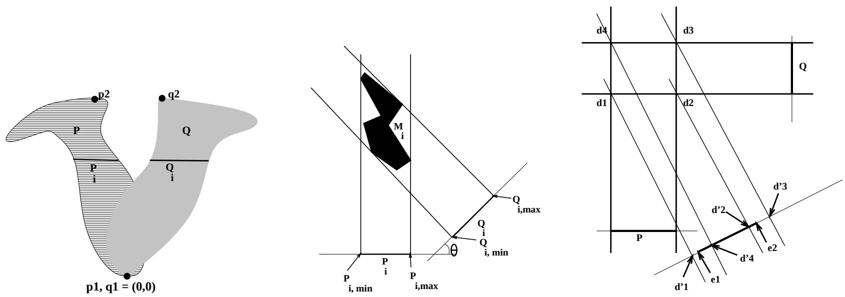

Fig. 1. Left: Two silhouettes with bottom points aligned. Middle: The $y = i$ plane. Right: Rectangular constraints project to a new image.

Given two silhouettes, p and q, we can always construct an object that can produce p and q, with a method based on volume intersection (eg., [8]). We may assume, without loss of generality (WLOG), that rotation is about the y-axis (when we consider three silhouettes this assumption results in a loss of generality). Also, WLOG assume that the silhouettes are transformed in the plane so that $p_1 = q_1 = (0,0)$ (see Figure 1, left), and the tangent to p and q at that point is the x axis. Assume the silhouettes are scaled so that p_2 and q_2 have the same y value. Assume also WLOG that M is positioned so that the point on M that projects to p_1 is placed at $(0,0,0)$. Moreover, we can assume that M is projected without scaling or translation. That is, in this setup, we can assume the object is projected orthographically to produce p, then rotated about the y axis, and projected orthographically to produce q.

If we cut through P and Q with a single horizontal line, $y = i$, we get two line segments, called P_i and Q_i. Denote the end points of P_i by $(P_{i,min}, i)$, $(P_{i,max}, i)$. These line segments are projections of a slice through M, where it intersects the plane $y = i$. Call this slice M_i. The segment P_i constrains M_i. In particular, in addition to lying in the $y = i$ plane, all points in M_i must have $P_{i,min} \leq x \leq P_{i,max}$, and there must be points on M_i for which $P_{i,min} = x$ and for which $x = P_{i,max}$ (and there must be points on M_i that take on every intermediate value of x). We get constraints of this form for every i. Any model that meets these constraints will produce a silhouette p.

Now, suppose that Q has been produced after rotating M by some angle, θ (see Figure 1, middle). The constraints that Q places on M have the same form. In particular, P_i and Q_i provide the only information that constrains M_i. However, the constraints Q_i places are rotated by an angle θ relative to the constraints of P_i. Therefore, together, they constrain M_i to lie inside a parallelogram, and to touch all of its sides. Therefore, we can create an object that produces both silhouettes simply by constructing these parallelograms, then constructing an object that satisfies the constraints they produce.

We denote the entire set of constraints that we get from these two images by C_θ. We now prove that it is not important to know θ, because the set of constraints that we derive by assuming different values of θ are closely related. Let $C_{\frac{\pi}{2}}$ denote the constraints we get by assuming that $\theta = \frac{\pi}{2}$. Then

Lemma 1. *$T_\theta C_\theta = C_{\frac{\pi}{2}}$ where: $T_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\cos\theta & 0 & \sin\theta \end{pmatrix}$*

That is, C_θ consists of a set of parallelograms, and applying T_θ to these produces the set of rectangles that make up $C_{\frac{\pi}{2}}$.

We omit the proof of this lemma for lack of space. This shows that C_θ and $C_{\frac{\pi}{2}}$ are related by an affine transformation. Since the affine transformations form a group, this implies that without knowing θ we determine the constraints up to an affine transformation. This is related to prior results showing that affine structure of point sets can be determined from two images ([6]). However, our results are quite different, since they refer to silhouettes in which different sets of points generate the silhouette in each image, and our proof is quite different.

3 Comparing Three Silhouettes

3.1 Silhouettes with Simple Cross-Sections

Now we assume that the third silhouette r is generated by again rotating M about the same axis. This problem is easier than the case of general rotations, because each of the parallelograms constraining M project to a line segment. For any other direction of rotation, they project to parallelograms. To simplify notation we refer to $C_{\frac{\pi}{2}}$ as C. These constraints are directly determined by assuming that p constrains the x coordinates of M, and q constrains the z coordinates. The true constraints, C_θ, depend on the true angle of rotation between the first two images, which is not known.

We can translate and scale r so that the y coordinates of the tops and bottoms of the silhouettes are aligned. This accounts for all scaling, and all translation in y. We may then write the transformation that generates r in the form:

$$T_\phi M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix} M + \begin{pmatrix} t \\ 0 \end{pmatrix}$$

which expresses x translation of the object, rotation about the y axis by ϕ, and then orthographic projection. We now examine how this transformation projects the vertices of the constraining parallelograms into the image. As we will see, the locations of these projected vertices constrain the new silhouette r. Since $C = T_\theta C_\theta$, we have $T_\theta^{-1} C = C_\theta$. Therefore, the projection of the true constraints, $T_\phi C_\theta = T_\phi T_\theta^{-1} C$, or:

$$T_\phi C_\theta = \begin{pmatrix} \cos\phi - \frac{\sin\phi\cos\theta}{\sin\theta} & 0 & \frac{\sin\phi}{\sin\theta} \\ 0 & 1 & 0 \end{pmatrix} C + \begin{pmatrix} t \\ 0 \end{pmatrix}$$

We will abbreviate this as: $\begin{pmatrix} a & 0 & b \\ 0 & 1 & 0 \end{pmatrix} C + \begin{pmatrix} t \\ 0 \end{pmatrix}$ with $a = \cos\phi - \frac{\sin\phi\cos\theta}{\sin\theta}$, $b = \frac{\sin\phi}{\sin\theta}$. a, b and t are unknowns, while the constraints, C and the new silhouette r are known. Our goal is to see if a, b and t exist that match the constraints and silhouette consistently.

Constraints on transformation parameters. We will be assisted by the fact that the projection of the constraints can be described by equations that are linear in a, b and t. However, because a and b are derived from trigonometric functions they cannot assume arbitrary values. So we first formulate constraints on these possible values.

We can show (derivation omitted) that a and b are constrained by:

$$-1 \leq |a| - |b| \leq 1; \qquad -(|a| + |b|)| \leq 1 \leq |a| + |b|$$

and that any a and b that meet these constraints lead to valid values for θ and ϕ. For any of the four possible choices of sign for a and b, these constraints are linear on a and b.

Constraints from new silhouette. Now consider again a cross-section of the constraints, C_i, and of the filled silhouette, R_i (see Figure 1). C_i is a rectangle, with known vertices $d_{i,1}, d_{i,2}, d_{i,3}, d_{i,4}$. Specifically: $d_{i,1} = (P_{i,min}, i, Q_{i,min})$ $d_{i,2} = (P_{i,max}, i, Q_{i,min})$ $d_{i,3} = (P_{i,min}, i, Q_{i,max})$ $d_{i,4} = (P_{i,max}, i, Q_{i,max})$.

Under projection, these vertices map to a horizontal line with $y = i$. We will consider constraints from just one $y = i$ plane, and drop i to simplify notation. Call the x coordinates of these projected points d'_1, d'_2, d'_3, d'_4. That is, $d'_j = (a, 0, b) \cdot d_j + t$. Notice that the sign of a and b determine which of these points have extremal values. For example, $a, b \geq 0 \Rightarrow d'_1 \leq d'_2, d'_4 \leq d'_3$. We continue with this example; the other three cases can be treated similarly.

R_i is a line segment, with end points whose x values we'll denote by e_1, e_2, with $e_1 \leq e_2$. Since M is constrained to lie inside C_i in the $y = i$ plane, we know that e_1 and e_2 must lie in between the two extremal points. That is: $d'_1 \leq e_1, \;\; e_2 \leq d'_3$. Furthermore, we know that M touches every side of C_i. This means that the projection of each side must include at least one point that is in R. This will be true if and only if: $e_1 \leq d'_2, \;\; e_1 \leq d'_4, \;\; d'_2 \leq e_2, \;\; d'_4 \leq e_2$.

These are necessary and sufficient constraints for r to be a possible silhouette of the shape that produced p and q. Finally, since $d'_j = (a, 0, b) \cdot d_j + t$ these constraints are linear in a, b, and t. As noted above, for $a, b \geq 0$ we also have linear constraints on a and b that express necessary and sufficient conditions for them to be derived from rotations. So we can check whether a new silhouette is consistent with two previous ones using linear programming.

Because of noise, the constraints might become slightly infeasible. It is therefore useful to specify a linear objective function that allows us to check how close we can come to meeting the constraints. We can write the constraints as, for example, $(a, 0, b)d_{i,1} + t \leq R_{i,min} - \lambda$. Then we run a linear program to satisfy these while maximizing λ. The constraints we have derived are all met if these constraints are met with $\lambda \geq 0$. If $\lambda < 0$ then λ provides a measure of the degree to which the constraints are violated.

3.2 Silhouettes with Complex Cross-Sections

Up to now, we have assumed that a horizontal cross-section of a silhouette consists of a single line segment. This will not generally be true for objects with multiple parts, holes, or even just concavities. These *multi-line* silhouettes complicate the relatively simple picture we have derived above. We wish to make several points about multi-line silhouettes. First, if we fill in all gaps between line segments we can derive the same straightforward constraints as above; these will be necessary, but not sufficient conditions for a new silhouette to match two previous ones. Second, if we merely require that the new silhouette have a number of lines that is consistent with the first two silhouettes, this constraint can be applied efficiently, although we omit details of this process for lack of space. Third, to exactly determine whether a new silhouette is consistent with previous ones becomes computationally more demanding, requiring consideration of either a huge number of possibilities, or an explicit search of the space of rotations.

But if we make a simple genericity assumption, the complexity can be reduced to a small size again.

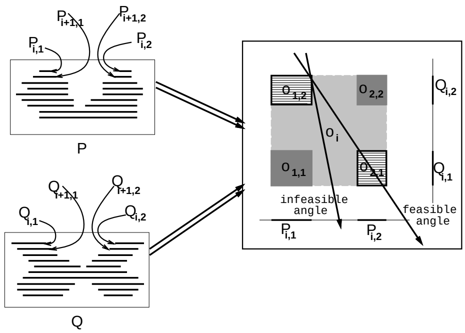

Fig. 2. Here we show two silhouettes, P and Q, that have cross-sections that consist of two line segments. On the right we show how the i cross-section leads to four parallelograms that must contain the object. Either the dark grey pair $(o_{1,1}, o_{2,2})$ or the patterned pair $(o_{1,2}, o_{2,1})$ must contain parts of the object. A third viewing angle, labeled "feasible angle" is shown for which this object may appear as a single line segment. An infeasible angle is also shown; from this viewpoint the object must produce two line segments in the image. Our system uses this constraint, though we omit details of how this is done. In some cases we make a continuity assumption across cross-sections. On the left, this means that, for example, if $P_{i,1}$ matches $Q_{i,1}$ ($o_{1,1}$ contains part of the object) then $P_{i+1,1}$ matches $Q_{i+1,1}$.

In the example shown in Figure 2, we can suppose either that $o_{1,1}$ and $o_{2,2}$ are occupied, or that $o_{1,2}$ and $o_{2,1}$ are occupied (there are other possibilities, which we can handle, but that we omit here for lack of space). When we assume, for example, that $o_{1,1}$ contains part of the object we can say that $P_{i,1}$ is matched to $Q_{i,1}$. Each of the two possible ways of matching $(P_{i,1}, P_{i,2})$ to $(Q_{i,1}, Q_{i,2})$ must be separately pursued, and gives rise to separate constraints that are more precise than the coarse ones we get by filling in the gaps in multi-line cross-sections. Suppose that for k consecutive cross-sections, the first two silhouettes each have two line segments. If we consider all possible combinations of correspondences across these cross-sections, we would have 2^k possibilities, a prohibitive number. But we can avoid this with a simple genericity assumption. We assume that in 3-D, it does not happen that one part of an object ends exactly at the same height that another part begins. This means that given a correspondence between line segments at one cross-section, we can typically infer the correspondence at the next cross-section.

3.3 Occlusion

The methods described above can also be applied to partially occluded silhouettes. To do this, something must be known about the location of the occlusion. For

Fig. 3. Seven objects used in experiments.

example, if a cross-section is known to be occluded in one silhouette, that cross-section can be discarded. If a cross-section is known to be partially occluded in the third silhouette, the visible portion can be required to lie inside the projection of the constraining parallelogram derived from the other two. Occlusion may not only make it impossible to derive constraints from occluded cross-sections, it may also create uncertainty in determining which cross-sections correspond to each other. For example, if the bottom of an object is blocked in a third view, we will not know how many cross-sections are occluded. We can solve this by searching through different scalings of the silhouette, which imply different possible ways of matching its cross-section to the first two silhouettes. We can then select the scale or scales that allow the resulting constraints to be met.

3.4 Experiments

We test these ideas using the objects shown in Figure 3. Our experimental system varies in which approach we use to handle multi-lines.

Experiment 1: First, we experiment with coarse constraints that fill in any gaps present in a silhouette cross-section. Also, we heuristically throw away some constraints that may be sensitive to small misalignments between different silhouettes. In this experiment we use five silhouettes taken from a figure of Snow White (Figure 4) photographed after rotations of 20°. First, all ten triplets of these silhouettes are compared to each other. In all cases they are judged consistent ($\lambda > 0$). Next, we compared each pair to 95 silhouettes, taken from the objects shown in Figure 3. About 6% of these other objects are also judged consistent with two Snow Whites (see Figure 4).

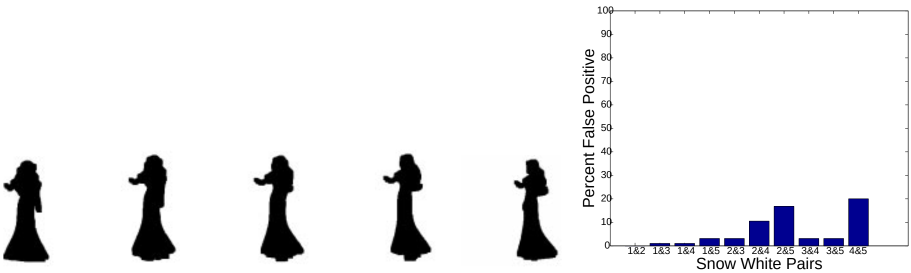

Fig. 4. Silhouettes of Snow White, numbers one to five from left to right. On the right, experimental results.

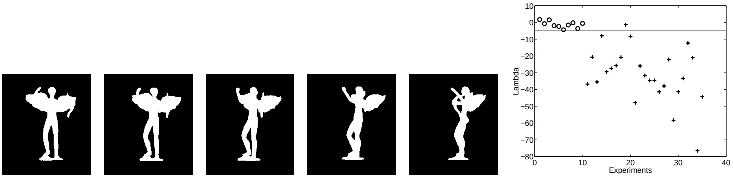

Fig. 5. On the left, silhouettes of Hermes. On the right, experiments comparing pairs of these to either a third silhouette of Hermes (first ten data points, shown as circles) or to silhouettes of other objects (shown as crosses). A horizontal line at $\lambda = -5$ separates correct answers with one false positive.

Experiment 2: Next, we performed a similar experiment using the silhouettes of Hermes in Figure 5. The axis of rotation was tilted slightly, so that the images do not exactly lie on a great circle on the viewing sphere. We heuristically compensate for this by searching for good in-plane rotations. For all 10 triples of silhouettes of Hermes, we obtain values of λ ranging from -4.4 to 1.7. However, when we compare randomly chosen pairs of Hermes silhouettes to randomly chosen silhouettes of the other five objects, we obtain only one case in twenty-five with λ larger than -4.4; other values are much smaller (see Figure 5).

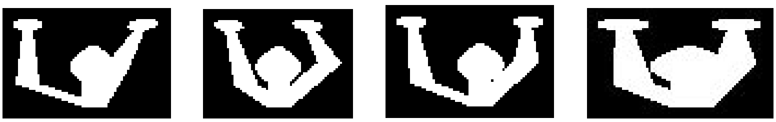

Fig. 6. The two figures on the left show the first and second silhouettes of the object used in experiment 4. The third silhouette shows this object from a new view. The fourth silhouette shows the same view with the object scaled so that its cross-sections are 1.8 times as big. This silhouette cannot be matched to the first two without greatly violating the constraints ($\lambda < -12$).

Experiment 3: We now show an experiment in which we search through possible correspondences between different object parts. We use a synthetic shape, vaguely like a human torso (Figure 6). Given three silhouettes, there are four possible correspondences between the "hands". We consider all four, then use the continuity constraint to determine correspondences at subsequent cross-sections. We compare two silhouettes to a third in which the shape has been "fattened", so that the cross-section of each part is scaled by a constant factor. When scale is 1, therefore, the third silhouette comes from the same object that produced the first two. In Figure 7 we show how λ varies with scale. We also show what happens if we do not hypothesize correspondences between the parts of the figure, but just fill in gaps in multiline cross-sections to derive a simple, conservative

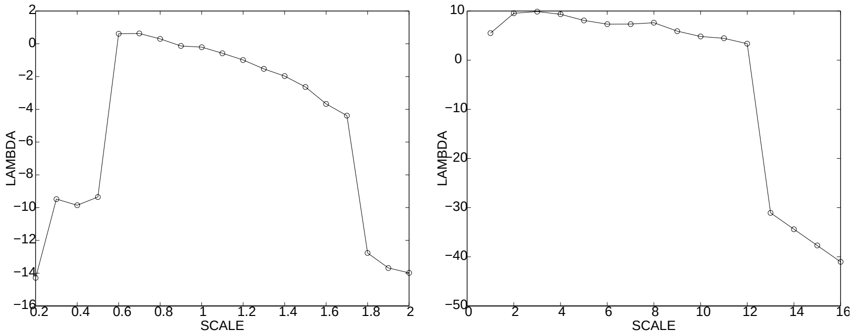

Fig. 7. On the left, we show how λ varies as the third silhouette scales. A scale of 1 means the third silhouette comes from the same object as the first two. On the right, we show this when individual parts are not matched, and only coarser constraints are derived by filling in all gaps in each cross-section of each silhouette. Note that the horizontal scale is approximately ten times larger on the right.

set of constraints. For an object like this, with small parts widely separated, this approach is much too conservative.

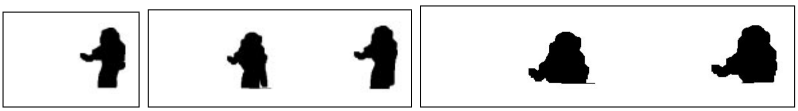

Fig. 8. On the left, the third Snow White silhouette, half occluded. The next two images show one hypothesized occlusion of the first two silhouettes that match this; the last two images show a second (false) match.

Experiment 4: Finally, we experiment with a case in which the first two Snow White silhouettes are matched to the third, but the bottom half of the third is occluded. In this case we try matching this half-silhouette to some top portion of the other two silhouettes, considering all possible top portions. We find two regions in the set of possible scales in which the third silhouette matches portions of the first two within one pixel of error; either when the first two are supposed about half occluded (the correct choice) or when they are supposed about 70% occluded (incorrect). Both are shown in Figure 8.

4 Conclusions

We have analyzed the problem of object recognition using silhouettes. We especially focus on the problem in which our knowledge of an object comes from seeing it from only a few viewpoints, under relatively unstructured viewing conditions, and in which we do not have a priori knowledge that restricts the model to belong to a special class of objects. This situation has not been much addressed, presumably because in this case it is not possible to derive a definite 3D model. However, we show that even though we may have considerable uncertainty about the 3D object's shape, there is still a lot of information that we can use to recognize the object from new viewpoints. This fits our general view that recognition can be done by comparing images, if our comparison method is based upon the knowledge that images are the 2D projections of the 3D world.

Our analysis has been restricted to the case where the objects or camera rotate about an axis that is parallel to the image plane. This is a significant restriction, but we feel that this case is worth analyzing for several reasons. First, it occurs in practical situations such as when a mobile robot navigates in the world, or in images generated to study human vision. Second, this analysis gives us insight into the more general problem, and provides a starting point for its analysis.

References

1. Brady, M. and Yuille, A., 1984, "An Extremum Principle for Shape from Contour," *IEEE Trans. PAMI* **6**(3):288-301.
2. Chien, C. and Aggarwal, J., 1989, "Model Construction and Shape Recognition from Occluding Contours," *IEEE Trans. PAMI* **11**(4):372–389.
3. Cipolla, R. and Blake, A., 1992, "Surface Shape from Deformation of Apparent Contours," *IJCV* **9**(2):83–112.
4. Forsyth, D., 1996. "Recognizing Algebraic Surfaces from their Outlines," *IJCV* **18**(1):21-40.
5. Giblin, P. and Weiss, R., 1987, "Reconstruction of Surfaces from Profiles," *IEEE Int. Conf. on Comp. Vis.*:136-144.
6. Koenderink, J. and van Doorn, A., 1991. "Affine structure from motion", *Journal of the Optical Society of America*, **8**(2): 377–385.
7. Laurentini, A., 1997, "How Many 2D Silhouettes Does it Take to Reconstruct a 3D Object," *Computer Vision and Image Understanding* **67**(1):81–87.
8. Martin, W. and Aggarwal, J., 1983, "Volumetric Descriptions of Objects from Multiple Views," *IEEE Trans. PAMI* **5**(2):150–158.
9. J. Ponce, D. Chelberg, W. B. Mann, 1989, "Invariant Properties of Straight Homogeneous Generalized Cylinders and Their Contours" *IEEE Trans. PAMI*, **11**(9) pp. 951-966.
10. Richards, W., Koenderink, J., and Hoffman, D., 1989, "Inferring 3D Shapes from 2D Silhouettes," in *Natural Computation*, edited by W. Richards.
11. Szeliski, R. and Weiss, R., 1998, "Robust Shape Recovery from Occluding Contours Using a Linear Smoother," *IJCV* **28**(1):27-44.
12. Vaillant, R. and Faugeras, O., 1992, "Using Extremal Boundaries for 3-D Object Modeling," *IEEE Trans. PAMI* **14**(2):157-173.