

Contour Migration: Solving Object Ambiguity with Shape-Space Visual Guidance

Wael Abd-Almageed

Christopher E. Smith

*Robotics, Artificial Intelligence and Vision Laboratory
Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131
{wamageed, chsmith@ece.unm.edu}*

Abstract

A fundamental problem in computer vision is the issue of shape ambiguity. Simply stated, a silhouette cannot uniquely identify an object or an object's classification since many unique objects can present identical occluding contours. This problem has no solution in the general case for a monocular vision system. This paper presents a method for disambiguating objects during silhouette matching using a visual servoing system. This method identifies the camera motion(s) that gives disambiguating views of the objects. These motions are identified through a new technique called contour migration. The occluding contour's shape is used to identify objects or object classes that are potential matches for that shape. A contour migration is then determined that disambiguates the possible matches by purposive viewpoint adjustment. The technique is demonstrated using an example set of objects.

1 Introduction

Silhouette matching is a technique that has its origins long before the advent of computers or computer vision. The technique has been used in computer vision for object classification/recognition tasks [1] [2]. However the technique has a fundamental problem; given a single occluding contour (silhouette), the object that produced that contour cannot be uniquely identified in the general case due to ambiguity. That is, the mapping from silhouette to object is a one-to-many mapping, with a single silhouette matching many objects. Strict silhouette-based recognition (or even classification) has no solution in the general case. One can augment silhouette matching with other techniques to attempt to reduce the number of potential matches to one.

Active vision systems (in particular, visual servoing systems) offer an advantage over static camera sys-

tems. The viewpoint of the camera can be changed with respect to the target for which a classification or recognition is desired. This allows multiple views that may disambiguate the target's actual identity from all of the possible matches. Given such a framework, the question now becomes: "What camera position excludes the largest number of members from the set of potential matches?". Furthermore, one might ask, what sequence of camera positions will allow the set of potential matches to be reduced to one member.

In this paper we present a technique that can provide optimal or near-optimal (in a statistical sense) camera motions for the purpose of disambiguating objects in a silhouette matching system. The technique, which we have termed *contour migration*, uses a database of object models that are stored as manifolds in a multi-dimensional shape space. Objects that can, dependent upon orientation, present the same occluding contour will have their shape-space manifolds co-tangent at the point corresponding to the shape model of the contour.

Working from this tangent point, a path through shape space can be identified that disambiguates the objects from one another by finding viewpoints where the objects will have different occluding contours. This path through shape space is called a contour migration, since the actual occluding contour is migrated across the target's surface and thus is also migrated through the shape space. It is this ability to identify contour migrations that allows us to solve a well-known problem in silhouette matching, namely object disambiguation.

This technique not only can be used with visual servoing, but also can be extended to static camera systems that are observing a target in motion. Rather than actively determining the viewpoint change that disambiguates objects, we can only observe the tar-

get and verify that the contour migration due to target motion is consistent with the various object hypotheses. When inconsistencies arise, those object hypotheses are removed from consideration until only one hypothesis remains.

Unfortunately, it is quite possible in many robotic domains to have a target that is not consistent with any object in the database. In this case, all of the potential silhouette matches will eventually be eliminated from the set of hypotheses. In this case, we could expand the database by adding a new manifold for the unknown object, without having to regenerate the manifolds for all other objects.

In the remainder of this paper, we will present a brief overview on some of the previous work related to active object recognition, introduce our contour migration technique, present experimental results, and discuss the implications and future work in the new area of contour migration.

2 Previous Work

In contrast to the long history of the silhouette classification problem in computer vision research, the problem of active object recognition is yet in its early stages. It is important here to differentiate between two different problems: viewpoint planning for a given 3-D object [3][4], and using visual servoing to resolve classification/recognition ambiguities. As observed by Calleri and Ferrie in [5], the idea of using the output of the classification/recognition system as a feedback signal to drive the visual servoing system has only been considered in a limited number of publications.

The similarity between all approaches in this area is finding a viewpoint that disambiguates a set of potential matches for an object undergoing recognition. All approaches differ with respect to the method of finding this viewpoint.

For instance, in [6] the authors introduced a new approach for driving the visual servoing system to a reference point that is associated with the set of objects that are known to be in the scene. That approach offered a solution to polyhedral object recognition, whereby the camera is moved to a canonical viewpoint of the object based on maximizing the projected lengths of two nonparallel edges in the image.

Hutchinson and Kak [7] introduced an approach for disambiguating objects from range images. In their approach, the authors evaluated a set of candidate sensing operations with respect to their effectiveness in minimizing ambiguity.

Another example is [8] in which Sven et al. presented an interesting approach for guiding a visual sensor to

a “better” viewpoint. They used an *aspect prediction graph* [9] to encode the object features at different viewpoints. An attention mechanism was then used to develop a strategy for moving the camera to the “better” viewpoint.

Maver and Bajcsy [10] presented an approach for selecting the next viewpoint in an occluded range image. They used the height information of the polygonally approximated border of the occluded region to plan a sequence of views.

In [5], the authors propose a system for “active object recognition in which a mobile agent traverses a static scene to determine the identity of objects within the scene”. They used a range image for probabilistic model matching on an incomplete data set.

3 Contour Migration

As previously stated, silhouette matching (and generally any object recognition technique) is under determined and leads to the problem that objects with similar occluding contours cannot be disambiguated. We propose a technique we call “contour migration” to assist solving the inherent ambiguity caused by silhouette matching. In concept, the technique is related to sliding contours [11]; however, the work by Kutulakos et al. involved acquiring object shape rather than performing classification or recognition.

To develop an active object recognition system, three main questions have to be addressed. First, how will the objects be encoded in what we call the shape space? The answer to this question is strongly related to the object classification/recognition technique that will be used. The second question is, what distance measure will be adopted to measure the distance between object encodings in the shape space? Thirdly, what is the camera movement strategy to be used? Answering these three questions establishes a framework for any active object recognition system. This section presents our proposed answers to these three questions.

3.1 Object Encoding and Classification Technique

Object Encoding in the Shape-Space. Several techniques have been used to represent the shape of a contour [2][12]. Our technique uses active deformable models (snakes) [13][14] to extract the occluding contour of objects in the workspace of a visually-guided robot. The object’s silhouette is then represented as a set of vectors $V = \{\bar{v}_1, \bar{v}_2, \dots, \bar{v}_L\}$ in the image space, where L is the silhouette (snake) length. The external angle between any two successive vectors, \bar{v}_i and \bar{v}_{i+1} , is defined by equation (1).

$$\theta_i = \cos^{-1} \frac{\bar{\mathbf{v}}_i \cdot \bar{\mathbf{v}}_{i+1}}{|\bar{\mathbf{v}}_i| |\bar{\mathbf{v}}_{i+1}|} \quad (1)$$

Computing the external angle between all pairs of successive vectors in V yields an observation sequence $\Theta = \theta_1 \theta_2 \dots \theta_L$ that represents the contour of the object. In the learning phase, this observation sequence is used to obtain a Hidden Markov Model (HMM) $\lambda_{i,j}$, as described by [15], that corresponds to object i at viewpoint j in the shape space, where:

$$1 \leq i \leq N \quad \text{and} \quad 1 \leq j \leq M \quad (2)$$

, N is the number of objects in the database and M is the number of discrete viewpoints in the shape space. In the testing phase, the observation sequence is used to test the HMMs of all objects. It is important to mention here that extracting a consistent starting point on the silhouette is not important because of using the formulation of [15].

The previous representation provides a method of modelling the contour of the object for one viewpoint. Migrating the viewpoint to every point in the 3-D space results in a new occluding contour (hence the name contour migration) at the new point. Creating an HMM for each point yields a manifold of HMMs representing this specific object. In practice, a manifold is created using a coarse discretization of the all possible viewpoints in the 3-D space, producing M HMMs for each object. Since the manifolds are smoothly changing, the surface can be interpolated between viewpoints. It is obvious that two or more objects can have similar occluding contours when looked at from one or more viewpoints, which is the main reason for ambiguous object classification. This causes the object manifolds to be co-tangent at certain viewpoints in the 3-D shape space. Figure 1 illustrates this idea.

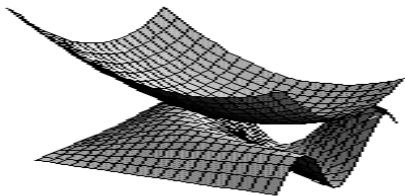


Figure 1: Two object manifolds co-tangent at different points.

Classification Technique. In most computer vision and pattern recognition applications that use HMMs as a classification mechanism, the classification decision is given by equation (3).

$$C = \arg \max_{1 \leq i \leq N} [P(\Theta_O | \lambda_i)] \quad (3)$$

where $P(\Theta_O | \lambda_i)$ is the probability generated by the HMM associated with object i if we are observing the sequence Θ_O and C is the classification of object O being tested.

This decision approach works well if there are no object similarities at certain viewpoints. For ambiguous objects, there is no guarantee that equation (3) will yield the correct classification. To obtain a set, S , of potential matches to the query object, we introduce equation (4) for decision making.

$$S = \{C_j : \begin{aligned} &|\max_{\forall i} [P(\Theta_O | \lambda_i)] - P(\Theta_O | \lambda_j)| \leq T, \\ &1 \leq i, j \leq N \end{aligned} \quad (4)$$

where T is a manually set threshold that determines the size of the set of potential matches. Setting $T = 0$ reduces the size of S to 1, which simplifies to equation (3).

3.2 Distance Measure for Object Encodings

The question we want to answer here is: given two HMMs representing two object encodings at a specific viewpoint on the object manifold, how similar are the two HMMs? This similarity measure represents the shape differences between the two objects in hand. For instance, the distance between two HMMs representing a ball and a circular disk should be minimum if looking at the two objects from above, while it should be maximum if looking at them from the side. In [16], Huang and Rabiner introduced a distance measure between two given HMMs as:

$$D(\lambda_i, \lambda_j) = \frac{1}{L} [\log P(\Theta | \lambda_i) - \log P(\Theta | \lambda_j)]. \quad (5)$$

The problem with this formulation is that we do not know which model has generated the observation sequence, Θ , of the current view. To solve this problem, we compute the distance between all possible permutations of models, as given by equation (6).

$$\hat{D}_m = \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} D(\lambda_{i,m}, \lambda_{j,m}) \quad (6)$$

where $1 < |S| \leq N$. Equation 6 can be computed in realtime because it involves only additions and

subtractions, as the $\log P(O|\lambda_i)$ parts of equation 5 were already computed and stored during the learning phase of the HMMs.

3.3 Camera Movement Strategy

The above formulation provides us with a set of potential matches, S , and a method for computing the distance between any two (or more) object encodings in the shape-space. Based on this distance measure, we want to move the camera to a new location in order to eliminate some (hopefully all but one) members of S . The best direction to move the camera is the one that maximizes the distance between all HMMs representing the members of S at the new viewpoint. To achieve this objective, we use equation (7) to select the new camera location.

$$l = \arg \max_{\forall m; m \neq k} \hat{D}_m \quad (7)$$

where k is the index of current viewpoint and l is the index of the new viewpoint.

This particular search method is not the most efficient in terms of asymptotic runtime. Our method attempts to break the set S into $|S|$ distinct singleton subsets. This will produce $O(|S|)$ comparisons. A search method that attempts to break $|S|$ into two subsets each of size $|S|/2$ produces only $O(\log_2 |S|)$ total comparisons. In the best case, this is more efficient with respect to the number of comparisons, but our search method produces a single camera move while the latter produces $\log_2 |S|$ camera moves. Since active vision systems take many orders of magnitude longer to move than to perform comparisons, our method is more realtime efficient.

Another consideration is the magnitude of the commanded motions. Our search method will only command a motion large enough to produce separation in the shape-space manifolds. Consider a similar method based upon aspect graphs. Such a method would be required to command motions sufficient to cause a transition from one state in the aspect graph to an adjacent state. These commanded motions would be significantly larger than our method since the granularity of the representation for aspect graphs is much lower than the granularity of the representation for our manifolds.

Finally, a termination condition needs to be set to avoid an endless loop situation that might occur if the size of the matchings set never becomes less than two. Simply, we limit the maximum number of moves to the number of discretizations of the shape-space, M . If M moves have been done without solving the ambiguity, the system stops for external user assistance.

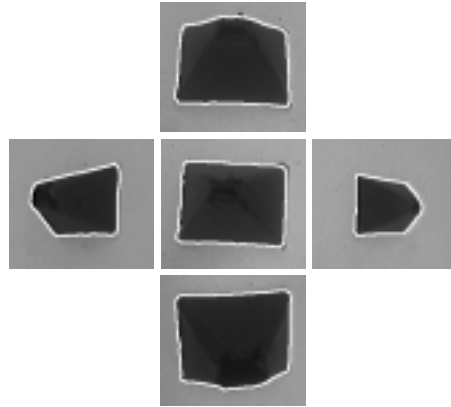


Figure 2: Contour Migration for a Pyramid.

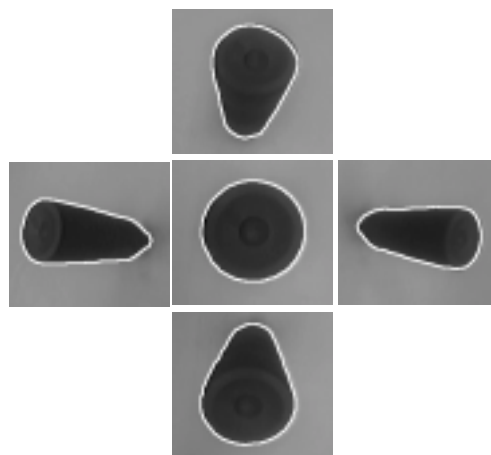


Figure 3: Contour Migration for a Water Bottle.

4 Experimental Results

To conduct our experiments, we used a Puma 560 manipulator with a Trident Robotics controller and a mini-camera with a 3mm lens. The output of the camera was fed to a Matrox Genesis vision processing board in a 1 GHz Pentium III computer running Windows 2000. The vision updates are transferred to the robot controller via a serial interface.

Our initial experiments used a small number of objects (i.e. $N = 4$) that typify the classic silhouette matching problem: a pyramid (figure 2), a water bottle (figure 3), a cylinder (figure 4) and a rectangular prism. It is not possible, in the general case, to uniquely disambiguate these objects given a single monocular viewpoint. The manifold of contours representing each object was discretized into five points (i.e. $M=5$). This gives a total of twenty ($M \times N$) HMMs.

The prism and the cylinder have ambiguous contours

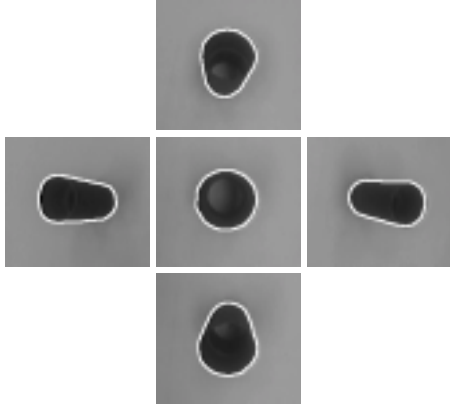


Figure 4: *Contour Migration for a Cylinder.*

when aligned such that the objects' principle axes are parallel to the image plane with the optical axis passing through the centroid of the object. On the other hand, the cylinder and the water bottle present ambiguous contours when their principle axes are aligned with the optical axis of the camera. Intuitively, we know that one must move the viewpoint such that the profile that discriminates the objects is visible and the system can make the correct classification/recognition.

Consider the cylinder and prism case (the more difficult case in our initial trials). Once the contour of the target is extracted and registered, the hypotheses set is derived. It will have both the cylinder and the prism as members since these two manifolds will both contain the rectangular contour presented by the target (see figure 5). Once this set has been derived, the method searches for a camera viewpoint that disambiguates the hypotheses objects' manifolds. Once found, the camera is moved to this point and a new contour is analyzed to determine which (if either) hypotheses match the observed contour.

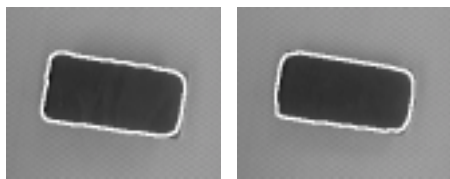


Figure 5: *An Ambiguous View of a Rectangular Prism and a Cylinder.*

For the rectangular prism and the cylinder, only a move along principle axis (translation) or about an axis orthogonal to the principle axis (rotation) will disambiguate these two hypotheses. In our experiments, the selected motion was a rotation (see figure 6) since the rotational motion produces larger differ-

ences with less motion and a rotational motion was kinematically safer due to the pose of the manipulator.

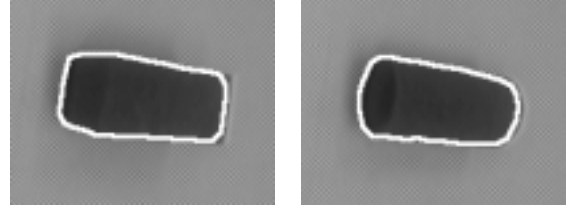


Figure 6: *Possible View After a Camera Rotation Selected by Contour Migration.*

In an experiment where the cylinder was presented in an ambiguous orientation with a prism, the contour migration was correctly selected that gives a camera view that splits the hypothesis set into two singleton sets. The experiment was repeated where the prism was placed. The contour migration selected allows the correct identification of the prism by giving a view point that disambiguates it from a cylinder.

A final trial is presented in figure 7 to show the migration required to disambiguate the cylinder from the water bottle. Since the kinematics of the robot being used did not allow the same range of rotational motion about the camera's x-axis as it did about the y-axis, the manifold was clipped. This biased the contour migration to select one of two particular ordinal directions. Also, any axis through the centroid of either the cylinder or the bottle can be the principle axis since the contours are circles. In practice, we discovered a bias since the active deformable model used would have a slight eccentricity (beyond what the human eye could discern) in its shape.

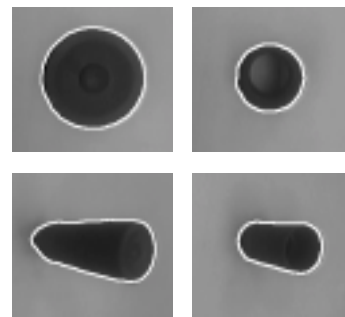


Figure 7: *Decided Motion to Disambiguate between a Cylinder or a Bottle.*

5 Conclusion

In this paper, we have presented a new technique, called contour migration, for disambiguating object

silhouettes by using a shape-space technique to plan camera motions. These motions result in new viewpoints that prune various incorrect object hypotheses resulting from the initial silhouette matching. This process of planning and purposive movement is repeated until the set of hypotheses has either one member (correct recognition) or zero members (an unknown object). Since the process of motion planning uses the hypotheses set and shape database together, we produce new viewpoints that reduce the size of the hypotheses set efficiently. We have demonstrated the efficiency of the approach through experiments using objects that are particularly challenging for a silhouette matching technique.

6 Future Work

This paper presents preliminary research results to prove the concept of contour migrations and, as such, has many directions for future work. The main areas for further research are: 1) refinement of the representation of shapes and contour migrations in the object database, 2) the incorporation of learning so that objects whose hypotheses set is disambiguated to the empty set can then be learned and stored for future use, and 3) the use of a hierarchical framework that performs shape classification first and then performs shape recognition, allowing more efficient recognitions by disambiguating larger shape class manifolds first and then performing shape recognition on limited number of specific object manifolds that belong to the shape class.

The number of discretizations of the shape-space is also another area that needs more investigation. It is obvious that M should be large enough to facilitate dynamic addition of new objects. Changing M dynamically is not a good approach as the database will need to be rebuilt.

Acknowledgments

This work was supported in part by the U.S. Department of Energy, under Grant No. DE-FG04-95EW55151, issued to the Manufacturing Engineering Program at the University of New Mexico, Sandia National Laboratories University Research Program (SURP) and Real-Time Innovations' (RTI) educational donations program.

References

[1] D. Perrin, O. Masoud, Christopher E. Smith, and N.P. Papanikolopoulos, "Unknown object grasping using statistical pressure models," in *2000 ICRA. IEEE International Conference on Robotics and Automation, 24-28 April 2000, San Francisco, CA, USA*.

[2] K. Lee and W. Street, "Model-based detection, segmentation, and classification for image analysis using online shape learning," *To appear, Machine Vision and Applications*, 2001.

[3] K.A. Tarabanis, R.Y. Tsai, and P.K. Allen, "The mvp sensor planning system for robotic vision tasks," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 1, pp. 72 – 85, February 1995.

[4] K. A. Tarabanis, P. K. Allen, and R. Y. Tsai, "A survey on sensor planning in computer vision," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 1, 2 1995.

[5] F.G. Callari and F.P. Ferrie, "Active object recognition: Looking for differences.," *International Journal of Computer Vision*, vol. 43, no. 3, pp. 189 – 204, 2001.

[6] D. Wilkes and J.K. Tsotsos, "Active object recognition.," in *Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 15-18 June 1992, Champaign, IL, USA*.

[7] S. Hutchinson and A. Kak, "Planning sensing strategies in a robot work cell with multi-sensor capabilities," *IEEE Transactions on Robotics and Automation*, pp. 765–783, 12 1989.

[8] Sven J. Dickinson, Henrik I. Christensen, John Tsotsos, and Goran Olofsson, "Active object recognition integrating attention and view point control," *Computer Vision and Image Understanding*, vol. 67, no. 3, 9 1997.

[9] Roboer M. Haralick, *Computer and Robot Vision*, vol. II, Addison-Wesely, 1993.

[10] J. Maver and R. Bajcsy, "How to decide from the first view where to look next," in *Proceeding of DARPA Image Understanding Workshop*, 1990, pp. 482–496.

[11] K. Kutulakos and C. Dyer, "Recovering shape by purposive viewpoint adjustment," *International Journal of Computer Vision*, , no. 12, pp. 113–136, 12 1994.

[12] N. Duta, A. Jain, and M. Dubuisson-Jolly, "Automatic construction of 2d shape models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 5, pp. 443–446, 5 2001.

[13] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proceedings of the First International Conference on Computer Vision*, 1987.

[14] W. Abd-Almageed and C. Smith, "Mixture models for dynamic statistical pressure snakes," in *the IEEE International Conference on Pattern Recognition, Quebec City, Canada, 2002*.

[15] W. Abd-Almageed and Christopher E. Smith, "Hidden markov models for silhouette classification," in *Proceedings of the World Automation Congress, Orlando, Florida, 2002*.

[16] B. H. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden markov models," *AT&T Technical Journal*, vol. 64, no. 2, 2 1985.