Proceedings of SPIE—The International Society for Optical Engineering

Volume 521

Intelligent Robots and Computer Vision

David P. Casasent/Ernest L. Hall
Chairmen/Editors

Cooperating Organizations
Carnegie-Mellon University, Robotics Institute • Catholic University of Leuven (Belgium)
Massachusetts Institute of Technology, Artificial Intelligence Laboratory
University of Cincinnati, Center for Robotics Research • University of Rhode Island, Robotics Research Center
UCLA Computer Science Department

November 5-8, 1984
Cambridge, Massachusetts

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Telex 46-7053

SPIE (The Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to advancing engineering and scientific applications of optical, electro-optical, and optoelectronic instrumentation, systems, and technology.

Recognizing partially hidden objects

J. L. Turney, T. N. Mudge and R. A. Volz

Department of Electrical Engineering and Computer Science,
University of Michigan, Ann Arbor, Michigan 48109

Abstract

In this paper, an approach is described for recognizing and locating partially hidden objects in an image. In the approach, templates are formed from the edge contours of the objects sought. Segments of each template are matched to segments of the edge contours of the image. A Bayesian approach is used to decide the probability an object has been located given that matches occur.

Introduction

A problem of great practical interest in machine vision is the recognition of objects that are partially hidden. As an application, imagine a kit of parts dumped on a table to be used in assembly, or a bin of parts from which a part must be extracted. In both of these cases it is likely that some of the parts will be partially hidden from the view of the camera.

Algorithms that are appropriate for solving this problem are numerous. Among these are the following: Duda and Hart [DuH72] used a version of the Hough transform to locate edge contours of objects in an image. The Hough technique is essentially histograming in multidimensional parameter space. It is useful, but many times finds false locations for objects. Perkins [Per78] looked for straight line and circular arc segments that he referred to as "concurves" in the slope angle-arclength representation of the edge contours of an image. With this approach he was able to locate objects that were overlapping as long as the overlap was not excessive. Ballard [Bal81] used a restricted form of the generalized Hough approach. He formed a template from the edge contours of the part he wish to locate and then compared edge points of the template to edge points of the image. He assumed that they matched if they had the same slope. Drawbacks to this algorithm are that templates have to be matched at all angles requiring considerable computation, and many false matches occur between the template points and the image points with the result that false locations can often be determined for objects. Bolles and Cain [BoC82] used an approach referred to as "local feature focus" in which features such as corners and holes were located in the image. Graphs connecting these features were formed and graphs formed from features of known objects were compared to those of the image. The algorithm relied on special features. Segen [Seg83] matched extrema in curvature in the edge contours of the image to extrema in the contours of the templates. His algorithm was computationally simple but relied on the location and determination of curvature for a few extrema points. Turney et al [TMV83],[TMV84] matched fixed length template contour segments to image edge contour segments of the same length in a space where the slope angle of a contour is parameterized by its arclength. Templates segments were weighted according to their "saliency". The algorithm was able to recognize objects even when they were heavily occluded, but require a large amount of computation. Bhanu and Faugeras [BhF84] use a relaxation approach to recognize objects. They approximated the template and the image boundary by boundaries composed of straight line segments. The algorithm was computationally intensive.

In [TMV83],[TMV84] emphasis was placed on the context in which an object is found. In industrial applications one generally knows the number and exact shape of the objects that are to appear in a scene, and it is advantageous to use this information to distinguish objects. This paper presents preliminary results from an algorithm that also uses this contextual information.

Bayesian approach

Before describing the algorithm itself, consider the following situation. Assume that an edge contour of an object appears in an image with its location and orientation about the origin as in Fig. 1a. In the terms of signal detection theory one might say that this contour represents a signal, S^0, which has been sent to the camera. Any other orientation or shift of this contour or any contour of another object that might appear in the image would represent a different signal, S^i, sent to the camera. Assume that R in Fig. 1b is the only portion of S^0 that is seen or received by the camera due to occlusion, noise or bad lighting. The question of interest is, "What is the probability that S^0 was sent given that R is received?" Denote this conditional probability by $P(S^0 \mid R)$. From Bayes rule this can be rewritten as:

$$P(S^0 \mid R) = \frac{P(R \mid S^0)P(S^0)}{\sum_i P(R \mid S^i)P(S^i)} \tag{1}$$

Note that the sum in the denominator is over all possible signals that might have been sent. Assume that R is composed of two fixed length nonoverlapping segments, r_1 and r_2, as shown in Fig. 1b. The reason for this restricted assumption will become obvious later. It will be assumed that under the condition that a signal S^i was sent the reception of r_1 is independent of the reception of r_2. It will also be assumed that the reception of these segments depends only on the segments of the signals S^i that correspond to the r_j. Call these segments s_j^i, using the subscript j to indicate the correspondence. Then

(a) (b)

Figure 1. Signal sent and signal received.

$$P(R \mid S^i) = P(r_1 \mid s_1^i)P(r_2 \mid s_2^i) \tag{2}$$

And Eqn. 1 becomes:

$$P(S^0 \mid R) = \frac{P(r_1 \mid s_1^0)P(r_2 \mid s_2^0)P(S^0)}{\sum_i P(r_1 \mid s_1^i)P(r_2 \mid s_2^i)P(S^i)} \tag{3}$$

Segments s_1^i and s_2^i could have been sent from different objects and could have accidentally aligned to produce the received signal, r_1 and r_2, but since there is generally a low probability of accidental alignment it will be assumed that s_1^i and s_2^i were sent from one object.

The probability $P(r_j \mid s_j^i)$ can be estimated as follows: $P(r_j \mid s_j^i)$ is the probability that r_j is received given that s_j^i was sent, which in other words is the probability that noise distorted s_j^i into r_j. One can assume a noise model for the image and then based on the type of edge detector employed in locating the edge contour, one can determined a function to estimate $P(r_j \mid s_j^i)$.

Even though the probabilities can be estimated it would still take a prohibitive amount of time to calculate the terms in the denominator in the above expression; therefore, it is necessary to approximate. The numerator of Eqn. 3 will generally be small unless r_1 and r_2 are simultaneously close to s_1^0 and s_2^0. Assume that this is the case, i.e. that $r_1 \approx s_1^0$ when $r_2 \approx s_2^0$, and denote the resulting denominator as D_{12}^0,

$$D_{12}^0 = \sum_i P(r_1 \approx s_1^0 \mid s_1^i)P(r_2 \approx s_2^0 \mid s_2^i)P(S^i) \tag{4}$$

Since D_{12}^0 is no longer a function of r_1 and r_2, one could calculate D_{12}^0 ahead of time, but as a simplification assume a further approximation. Assume that for any term in D_{12}^0 where s_1^i and s_2^i are simultaneously close to s_1^0 and s_2^0 respectively the term $P(r_1 \approx s_1^0 \mid s_1^i)P(r_2 \approx s_2^0 \mid s_2^i)$ is equal to one and to zero otherwise. This approximation can be used to overestimate the size of the D_{12}^0 and hence provide a lower bound on the probability. With these approximations D_{12}^0 can be calculated off-line and from D_{12}^0 the probability $P(S^0 \mid R)$ can be estimated during run-time.

In practice because of occlusion, one does not know the segments s_1^0 and s_2^0 that will be received as r_1 and r_2 and it is necessary to calculate a denominator D_{uv}^0 for each possible pair of segments s_u^0 and s_v^0 of the signal S^0. The method of determining D_{uv}^0 for all u and v is discussed in the section on training later in this paper.

Storing the denominators requires considerable memory space. Since a segment can be centered around each point on the contour, a large number of denominators must be calculated and stored ahead of time. For N possible segments taken two at a time $\frac{N(N-1)}{2}$ denominators must be stored. If three segments, s_1^0, s_2^0, s_3^0 had been considered a more accurate estimate of $P(S^0 \mid R)$ could be obtained but at the expense of storing $\frac{N(N-1)(N-2)}{6}$ denominators. This is why only two segments have been considered in the above development.

This Bayesian approach of estimating whether or not a signal has been sent given the reception of a pair of segments is used together with the matching approach to locate partially hidden objects. The matching approach is discussed in the next section.

Matching segments

To locate an object, segments of the template edge contours are matched to segments of the image edge contours. The approach used has been discussed in detail in [TMV83] and [TMV84], but is briefly repeated here for completeness.

In this approach, the template and image contours are represented in two spaces, in normal cartesian space and in slope angle - arclength space, or $\vartheta - s$ space. (Slope angle - arclength space is discussed in [Bal82].) The template and image contours in both the $\vartheta - s$ space representation and cartesian space are partitioned into segments of fixed arclength.

Matching is performed in $\vartheta - s$ space since it is more efficient than matching in cartesian space. A $\vartheta - s$ representation of the template segment (shown highlighted in Fig. 2) is moved along the s axis so that its center is aligned with the center of the image segment to which it is to be compared. The template segment is then shifted in the ϑ direction so that the mean ϑ value of the template segment has the same mean ϑ value as the image segment. This ϑ shift (see Fig. 2) measures the average slope angle difference between the template and image segments and will be referred to as the "angle of match". The difference in ϑ is found between corresponding points of the template and image segment. The sum of the squares of these differences is used to measure the similarity of the two segments. If they are similar, they are assumed to match.

If the template and image segments match in $\vartheta - s$ space, the match is recorded as follows: In cartesian space a vector from the center of the template segment to the template centroid is determined. This vector is rotated by the "angle of match" and translated so that its tail is centered at the same location as the center of the image segment (See Fig. 3a.) The location of the head of this vector represents a potential location of the

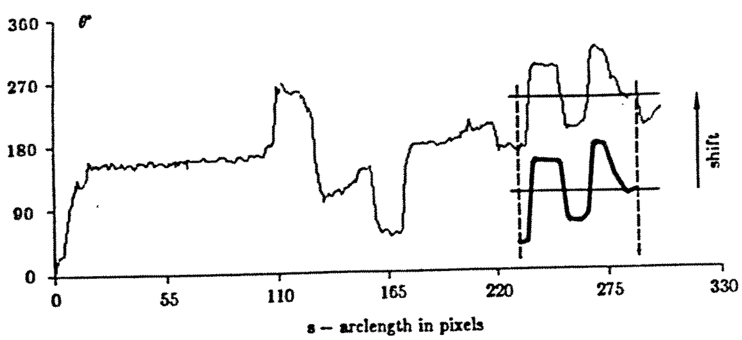
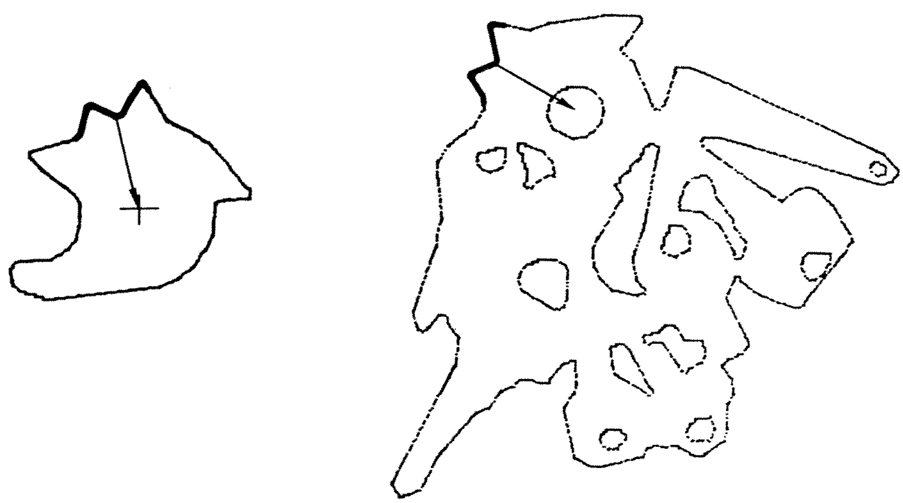


Figure 2. Matching in $\vartheta - s$ space.



(a)

(b)

Figure 3. Storing a record of the match.

centroid of the template in the image. Each pixel location in the image has an associated list. If the head of the vector falls on a particular pixel, a record containing the identity of the template segment and the "angle of match" is stored in the list at that pixel location (See Fig. 3b.)

Training

In the algorithm presented in this paper the denominators for the conditional probabilities of an object are trained off-line. The template of the object to be trained is matched to templates of all of the objects (including itself) that might appear in the image at all orientations and locations. From this matching information one can determine the denominators of the conditional probabilities.

The object whose conditional probabilities are to be determined is termed the training object. As before let D_{uv}^0 denote the denominator term for the conditional probability when $r_u \approx s_u^0$ and $r_v \approx s_v^0$, where s_u^0 and s_v^0 are segments of the training object. The calculation proceeds as follows: The segments of the template of the training object are matched to the segments of the template of one of the objects. After matching, the list of records at each pixel location is examined. If the list at a pixel location or the list of any nearby pixel contains a record of a match by segment s_u^0 and a record of a match by s_v^0 at roughly the same angle of match, then D_{uv}^0 is incremented by 1. After all lists generated for this match have been examined they are disposed of and the template of the training object is matched to the template of the next object generating new lists of match records. These lists are again analyzed, and so forth until all other objects have matched and all possible contributions to D_{uv}^0 have been counted.

Locating partially hidden objects

In locating an object, the segments of the template of the object are matched to the segments of the contours in the image, using the approach discussed previously. When a template segment of the object matches a image segment, a record of the match is stored in a list associated with a pixel at the location of a possible centroid of the template. Then the list at that pixel and of all neighboring pixels are examined to see if there is any previous record of a match with another template segment at the same match angle.

If such a record exits the denominator D_{uv}^0 is looked up in the training table using the identity of the two template segments (the one just stored, s_u^0 and the one previously stored, s_v^0) as indices. The probability, $P(S^0 \mid r_u \approx s_u^0, r_v \approx s_v^0)$, can be calculated from D_{uv}^0. If this probability is large enough an attempt is made to fit the template to the image. If the template fits sufficiently well, the object is assumed to have been found. If not, matching continues until the object is found.

Results

Fig. 4 shows the contours of the set of objects that were used during training.

Fig. 5 shows an example of the joint conditional probability, $P(S^0 \mid r_u \approx s_u^0, r_v \approx s_v^0)$. The bullet indicates the center of segment s_u^0. A vertical line is drawn from each possible center of s_v^0. The length of the line is proportional to the joint conditional probability $P(S^0 \mid r_u \approx s_u^0, r_v \approx s_v^0)$ that one would obtain if s_v^0 were centered about each of these possible locations. Note that since s_u^0 is a straight line segment any choice of s_v^0 along the

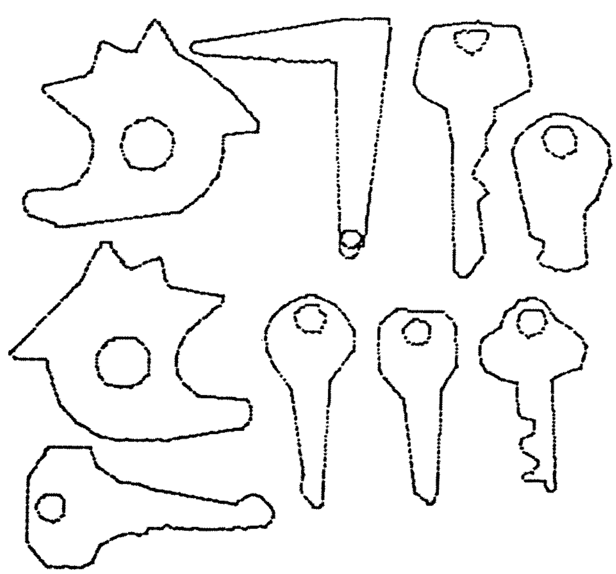


Figure 4. Training set.

same straight section would yield a low conditional probability (as indicated in Fig. 5 by the vertical lines of zero length originating from these points), since the reception of two such segments would not provide enough information to determine which object in the training set had been sent.

In the algorithm, several objects can be searched for simultaneously. Fig. 6 shows some preliminary results when two objects were sought. Total recognition time for both objects was 60 seconds. The objects were located repeatedly by different pairs of segments, r_u and r_v. When an object was located the template was drawn at the predicted position and orientation. No attempt was made to determine a final "best" fit for either of the objects. Note that no false locations were determined.

Summary

In this paper it has been shown that a Bayesian approach, together with template segment matching in a special space can be used as an effective approach to locate partially hidden objects. Efforts will be made to improve the speed of this algorithm, and to extend the approach to the recognition of scaled and tilted partially occluded parts. Work is also underway to extend this approach to three dimensions to aid in the location of parts from depth map information.

Acknowledgments

The research work leading to this paper was performed under Air Force Contract No. F49620-82-C-0089.

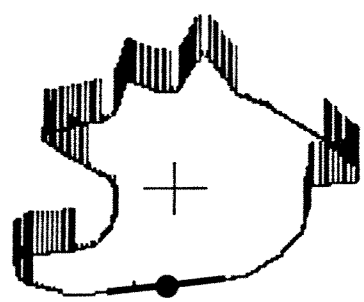


Figure 5. Joint conditional probability.

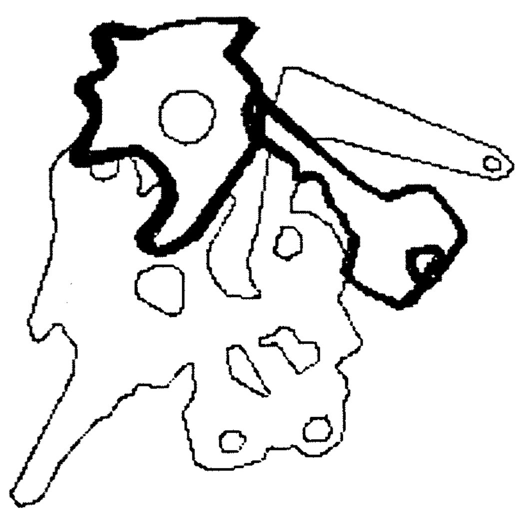


Figure 6. Objects located.

References

[Bal81] D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, Vol. 13, No. 2, 1981, pp. 111-122.

[Bal82] D. H. Ballard, C. M. Brown, "Computer Vision", Prentice Hall, Englewood Cliffs, N.J. 07632, 1982, p. 237.

[BhF84] B. Bhanu and O. D. Faugeras, "Shape Matching of Two-Dimensional Objects," *IEEE Pattern Analysis and Image Processing*, Vol. 6, No. 2, Mar. 1983, pp. 137,155.

[BoC82] R. C. Bolles, R. A. Cain "Recognizing and Locating Partially Visible Workpieces", Proceedings IEEE Conference on Pattern Recognition and Image Processing., pp. 498-503, June 1982.

[DuH72] R. O. Duda and P. E. Hart, "Use of the Hough Transform to Detect Lines and Curves in Pictures," CACM, Vol. 15, No. 1, pp 11-15, 1972.

[Per78] W. A. Perkins, "A Model-based Vision System for Industrial Parts," *IEEE Transactions on Computers*, Vol. C-27, No. 2, Feb. 1978, pp. 126-143.

[TMV83] J. L. Turney, T. N. Mudge and R. A. Volz, "Experiments in Occluded Parts Recognition," *Proceedings of the Society of Photo-optical Instrumentation Engineers Cambridge Symposium on Optical and Electro-optical Engineering*, Cambridge, MA, Nov. 1983,

[TMV84] J. L. Turney, T. N. Mudge and R. A. Volz, "Recognizing Partially Occluded Parts," *IEEE Pattern Analysis and Image Processing* (to appear).

[Seg83] Jakub Segen, "Locating Randomly Oriented Objects from Partial View," *Proceedings of the Society of Photo-optical Instrumentation Engineers Cambridge Symposium on Optical and Electro-optical Engineering*, Cambridge, MA, Nov. 1983,