A System for Model-based Recognition of Articulated Objects

Bir Bhanu and Joon Ahn
Center for Research in Intelligent Systems
University of California, Riverside, California 92521 USA
{bhanu, ahn}@vislab.ucr.edu

Abstract

This paper presents a model-based matching technique for recognition of articulated objects (with two parts) and the poses of these parts in SAR (Synthetic Aperture Radar) images. Using articulation invariants as features, the recognition system first hypothesizes the pose of the larger part and then the pose of the smaller part. Geometric reasoning is carried out to correct identification errors. The thresholds for the quality of match are determined dynamically by minimizing the probability of a random match. Results are presented using SAR images of three articulated objects. The system performance is evaluated with respect to identification performance, accuracy of estimates for the poses of the object parts and noise.

1. Introduction

Recognition of objects in SAR imagery is an active area of research in pattern recognition [4]. In this paper we focus on the problem of recognizing articulated objects (with two parts) and the poses of the articulated parts. Previous work in this area, [1] and [3], has used simple models (like scissors and lamps) in visual imagery and has used constraints around a joint to recognize these objects. Because of the unique characteristics of SAR image formation (specular reflection, multiple bounces, low resolution and non-literal nature of the sensor), it is difficult to extract linear features (commonly used in visual images), especially in SAR images of targets at six inch to a foot resolution. Previous recognition methods for SAR imagery using templates [4] or boundary contours are not suitable for the recognition of articulated objects, because articulation or occlusion changes the object outline and each different articulation configuration requires a different template leading to a combinatorial explosion.

The key contribution of the paper is the recognition of articulated objects and the articulation of the objects in SAR imagery using models based on articulation invariants. An *end-to-end system* has been developed whose input is a target chip and the final result is the identification of the target and the poses of its parts. The system has been extensively tested against occluded articulated objects and noise.

2. System Overview

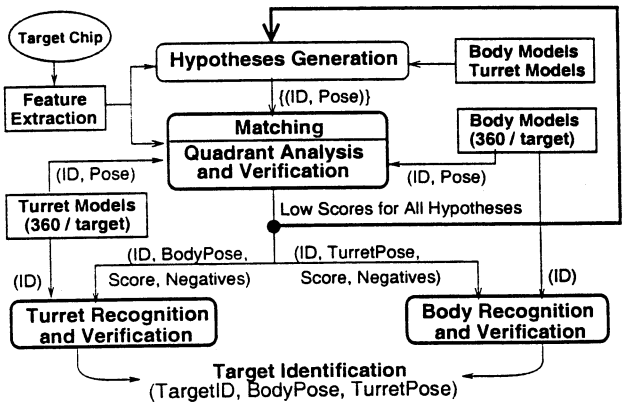


Figure 1. Our system for object recognition.

The basic assumption behind our approach is that more scattering centers (features) are from the larger part than from the smaller one so that the models for the larger part are used first to find the target identification (ID) and the pose of the larger part. For some targets like the M1a1 tank, occasionally more scattering centers come from the smaller part (turret) than from the larger one (body). If the system succeeds in recognizing the target and its body pose, those scattering centers used for the body part recognition (the positives) are eliminated from the test data. The remaining scattering centers (called the negatives) are supplied to the next stage to recognize the turret part and its pose. If the system fails to recognize the tar-

1051-4651/98 $10.00 © 1998 IEEE

get ID and its body pose, the system tries to recognize the target ID and its turret pose instead. After the recognition of the turret, it tries to recognize the body pose based on the negative scattering centers, which are left over from the turret part recognition. Figure 1 shows the matching components including the feedback (loop) used in geometric reasoning from the matching module to the indexing module.

• **Invariant feature extraction and model building:** We find that some of the scattering centers remain invariant to the turret articulations. The articulation invariances of M1a1, T72, and T80 tanks are 37%, 48%, and 53% respectively. For model building and experimentation, we use the invariants to build body models and a subset of variants to build turret models. We do not fix the number of scattering centers to be extracted from the target chips because we want to get as many features as possible from both body and turret.

• **Hypotheses generation and verification:** In the recognition phase, a set of hypotheses is generated using a geometric hashing technique. The program then finds the best data/model correspondence using a quadrant analysis technique which transforms the scattering centers from the model coordinate system to the image coordinate system. In this transformation, only translation is considered because rotation is handled by 360 models for every single degree of azimuth (note that there is no scaling involved in SAR image formation). This quadrant analysis technique allows positional error, ϵ_p, within one pixel. Finally, the system verifies the top ten hypotheses, from the quadrant analysis, which have the highest matching scores.

(a) Quadrant analysis: Each entry in the transformation space represents a transformation $\mathcal{T}$ and the value represents the number of correspondences between model and data scattering centers for exact matching positions. Since the feature extraction module uses eight-neighbor comparison, the closest two scattering centers are two pixels apart from each other. In order to allow correspondence between pixels that are one pixel apart, the quadrant analysis routine generates a new transformation space by adding all values at the four corners of each quadrant. A new transformation space, $\mathcal{Q}$, is constructed as follows: $\mathcal{Q}(x,y) = \mathcal{T}(x,y) + \mathcal{T}(x+1,y) + \mathcal{T}(x,y+1) + \mathcal{T}(x+1,y+1)$. Figure 2 shows an example of quadrant analysis.

(b) Dynamic selection of matching threshold: We use a dynamic matching threshold based on a statistical occupancy model developed by Grimson and Huttenlocher [2]. The main assumption underlying this model is that the extraneous features in an image will be uniformly randomly distributed with respect to a given object model. We use this approach with restrictions and modifications that are required for SAR images. In particular, we are concerned with scattering centers as point features in a SAR image and only the translation transformation between the model and data features.

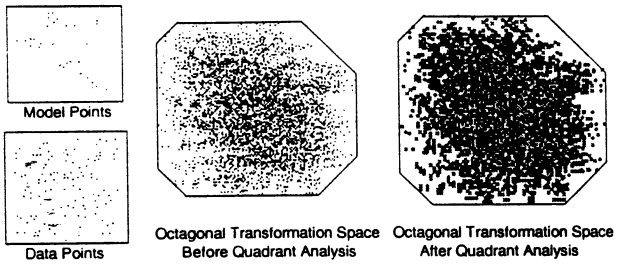


Figure 2. An example of new transformation space with convex octagonal bounding box.

Given a correspondence between data and model points for a fraction f ($f \epsilon [0,1]$) of the m model points, what is the relation between f and the probability δ that correspondences can occur at random? Once f and δ are known, the matching threshold can be dynamically selected for each input. In order to characterize the probability of a false match of a model to an image, statistical occupancy models are used. Let l be the number of random feature correspondences. Assuming the acceptable positional error to be within $\pm\epsilon_p$ in both x and y directions, which is a square of dimensions $2\epsilon_p \times 2\epsilon_p$, the volume of the range of feasible transformations V_{jJ} for single data-model pairing (j,J) is $c_{jJ} = 4{\epsilon_p}^2$. Let the sum of the sizes of all the transformation space volumes over the total size of the transformation space be λ. Then,

$$\lambda = \frac{sm4{\epsilon_p}^2}{A} = sm\bar{c}$$

where s and m are the number of scattering centers for data and model, respectively, A is the convex octagonal area (the bounding octagon) of the transformation space, and $\bar{c}$ is the average normalized volume size which is $4{\epsilon_p}^2/A$.

Given n cells and r events, what is the probability, p_k, that a given cell contains exactly k events? Two widely used models for the probability are *Maxwell-Boltzmann* and *Bose-Einstein* models. *Maxwell-Boltzmann* model assumes that the events are uniformly randomly distributed, such that all n^r possible placements of the r events in the n cells are equally probable. *Bose-Einstein* model, an alternative model, assumes that each *distinguishable distribution* of events across cells has an equal probability of occurrence. To

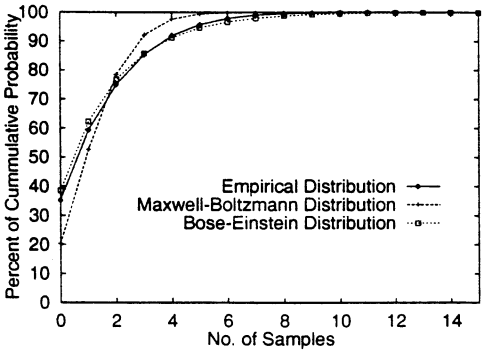


Figure 3. Comparison for the fitness of two occupancy models.

Table 1. D_n for Kolmogorov-Smirnov Test.

(Total 360 Models/Target)

	Body		Turret	
	Bose-Einstein	Maxwell-Boltzmann	Bose-Einstein	Maxwell-Boltzmann
M1a1	4.13 (%)	13.20 (%)	3.31 (%)	10.99 (%)
T72	12.52 (%)	32.41 (%)	3.64 (%)	9.38 (%)
T80	4.31 (%)	15.97 (%)	4.97 (%)	5.22 (%)
AVE	6.90 (%)	20.53 (%)	3.97 (%)	8.53 (%)

select a particular occupancy model for articulated object recognition, we generated test data with random scattering centers inside the bounding box whose size is the average size of all test SAR chips. We compute the cumulative empirical distribution as follows.

(1) For each pair of model and data points, compute a point in the transformation space. Enter an event into the cell containing this point. (2) Do the quadrant analysis to get a new transformation space. (3) Over all cells inside the convex octagonal hull of the events, count the number of events in each cell (the occupancy numbers), and tally the number of cells with each occupancy number. (4) For each entry in the tally, normalize the entry by the total number of cells, thus producing the empirical distribution of the number of events per cell. (5) Sum the normalized values to obtain the cumulative empirical distribution.

Figure 3 shows the fitness comparison of the two occupancy models for the empirical distribution generated using the model and data shown in Figure 2. Then we compute the total area of the convex octagonal hull A, and the average distribution of the events λ. This cumulative empirical distribution is supplied to the Kolmogorov-Smirnov test (K-S test). The K-S test measures the maximal difference between the empirical distribution and some hypothesized distribution (*e.g.*, Bose-Einstein model or Maxwell-Boltzmann model). All 360 models of both body and turret models for each target are tested against the data. Table 1 shows the results of the K-S test. For $\alpha = 0.05$ level, D_n's (the maximum difference between two cumulative distribution function) for Maxwell-Boltzmann model are large enough to reject H_0 while D_n's for Bose-Einstein model are not large enough to reject H_0. Accordingly, we choose Bose-Einstein model for the occupancy model.

From $Pr\{v \geq l\}$, probability that l or more of the volumes intersect at random, we can determine the fraction of model features f_0 such that the probability of mf_0 features being matched at random is less than some predefined level δ. Since δ is a function of the noise in the data measurements, and the uncertainty in position $4{\epsilon_p}^2$, we have

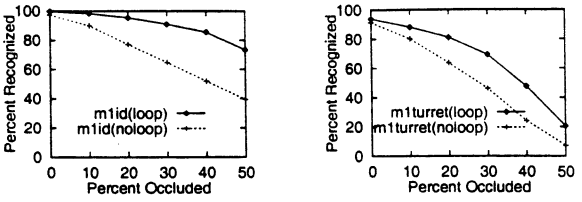
$$f_0 \geq \frac{\log(\frac{A}{4{\epsilon_p}^2})}{m\log(1+\frac{1}{ms\bar{c}})}.$$

In the verification stage, we consider the number of correspondences from the quadrant analysis, the number of data features, the number of model features, and the size of the transformation space. Based on this information, the verification stage calculates the matching score along with the formal threshold.

- **Geometric reasoning:** In our approach, we assume that a target has two parts one of which is larger than the other. For example, the body part is larger than the turret part and the body part has more articulation invariant points than the turret part. But there are some exceptions like the M1a1 tank which has a very large turret compared to the size of the body. Even though the body part is still larger than the turret part, there are more articulation invariant points on the turret than on the body for some configurations. In this case, the recognition of body part will fail because of the lack of articulation invariant points from the body. For this case we try to recognize the turret part first. Figure 4 shows the improvement obtained through geometric reasoning (the 'loop' cases). The identification does not consider the pose of each part. For the turret pose recognition, the correctness is within $\pm$ 5 degree accuracy. Improvement of recognition is significant for the identification and turret pose recognition cases.

3. Results

- **Models and Data:** In building models, we have



(a) Identifications (b) Turret pose

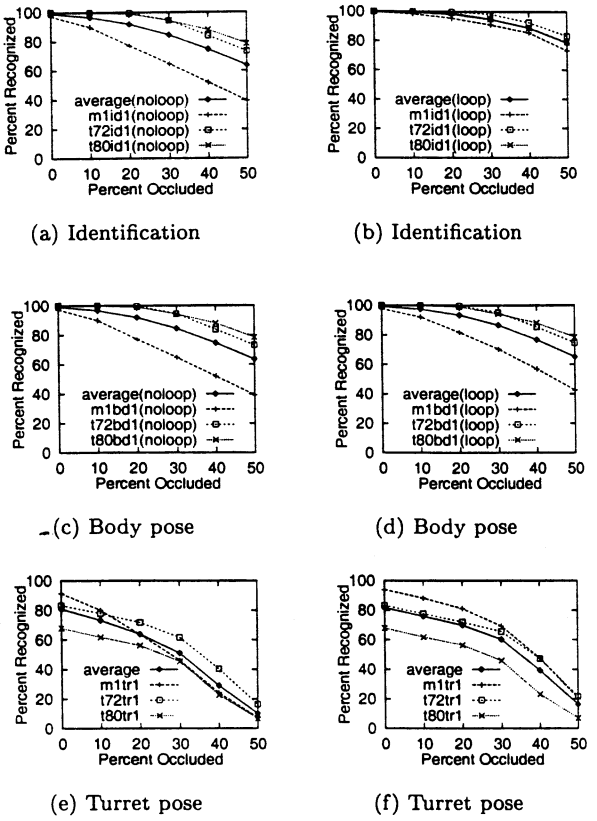
Figure 4. Effect of geometric reasoning.

used three targets: T-72 tank, M1a1 tank and T-80 tank. Each target has four different articulation configurations which are achieved by rotating the turret 0°, 30°, 60°, and 90° relative to the tank body. For each articulation configuration, we generated 360 images (one for each degree in azimuth) for a given depression angle of 15°. So, the total number of images generated is 4320 (3 targets × 4 articulations × 360). From each image, we extracted scattering centers from their signal returns as point features of the model. For each of the six experiments, we used two articulation configurations of each target (3 targets × 2 articulations × 360) to build 360 body models and 360 turret models. The other two remaining articulation configurations are used as test data. These results averaged to demonstrate the final performance.

• **Matching results:** In Figures 5 (a) through (f), *average, M1a1, T72, T80* curves represent the experimental results of the average and each individual target, respectively. The x-axis shows the occlusion rate which ranges from 0% to 50% in 10% steps. The y-axis shows the correct recognition percentages. The identification performance degrades gracefully as the occlusion rate increases. The recognition of T-72 and T-80 are similar while M1a1's recognition rate decreases faster as the occlusion rate increases. In the body and turret pose recognition results, correct pose recognition is within ± 5°. The low pose recognition rate of the M1a1 tank body is due to the fact that the relative size of the turret to the body is large. This characteristic is reflected in the pose recognition of the turret part. M1a1 turret pose recognition is much better than the other two targets. As the occlusion rate increases, the turret pose recognition drops rapidly beyond 30% occlusion. This is expected because the turret part is in the middle of the image, in general, and will have fewer valid point features as the occlusion rate increases.

4. Conclusions

We have developed an end-to-end system for recognition of articulated objects. We have demonstrated



(a) Identification (b) Identification

(c) Body pose (d) Body pose

(e) Turret pose (f) Turret pose

Figure 5. Recognition results without (a,c,e) and with (b,d,f) geometric reasoning.

the performance of our approach using extensive experiments. Currently we are extending the system with additional features.

Acknowledgement: This work was supported in parts by grants F49620-97-0184 and DAAH049510049.

References

[1] A. Beinglass and H. J. Wolfson. Articulated object recognition, or : How to generalize the generalized Hough transform. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 461–466, June 1991.

[2] W. E. L. Grimson and D. P. Huttenlocher. On the verification of hypothesized matches in model-based recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 61(3):1201–1213, Dec. 1991.

[3] Y. Hel-Or and M. Werman. Recognition and localization of articulated objects. In *Proc. IEEE Motion of Non-Rigid and Articulated Objects*, 1994.

[4] L. Novak, G. Owirka, and C. Netishen. Radar target identification using spatial matched filters. *Pattern Recognition*, 27(4):607–617, 1994.

Hosted by

APRS

IAPR

ICPR '98

Proceedings

14th International Conference on Pattern Recognition • Volume II

August 16–20, 1998 • Brisbane, Australia

Sponsored by

CSIRO AUSTRALIA

IIB

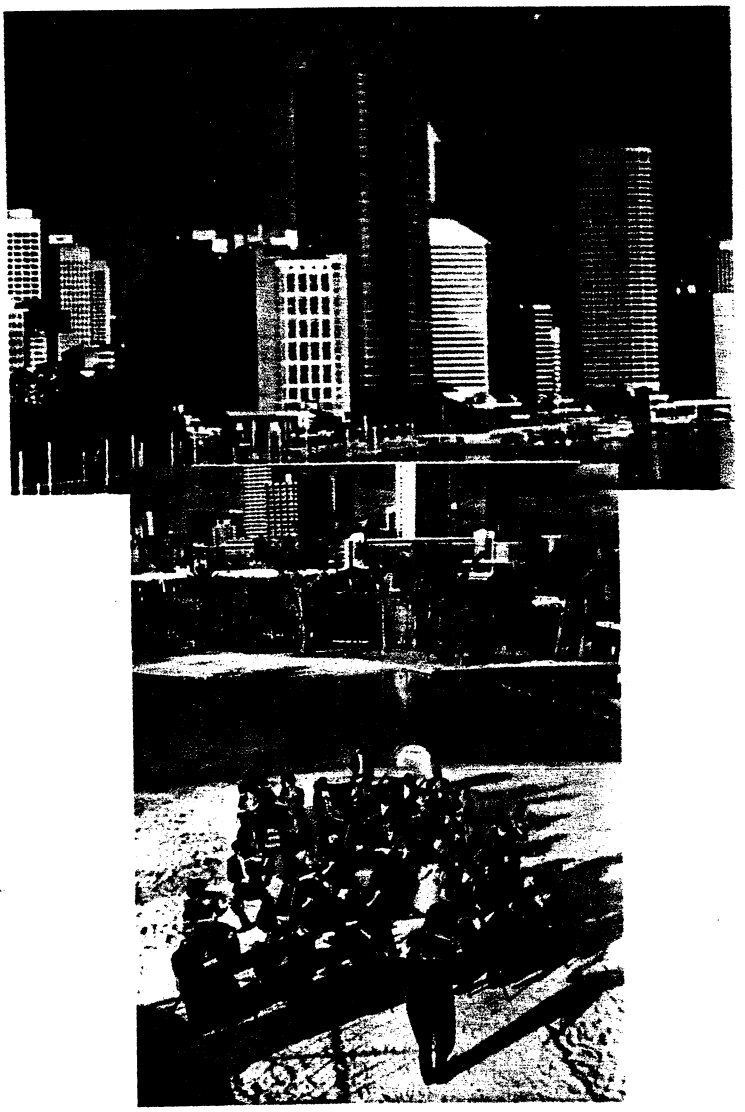
Department of Tourism, Small Business and Industry

University of Ballarat

curtin

THE UNIVERSITY OF QUEENSLAND

MONASH
AUSTRALIA'S INTERNATIONAL UNIVERSITY



Editors
Anil K. Jain, Svetha Venkatesh, and Brian C. Lovell

IEEE COMPUTER SOCIETY

IEEE

Proceedings

Fourteenth International Conference on Pattern Recognition

August 16 – 20, 1998

Convention and Exhibition Centre

Brisbane, Australia

Editors

Anil K. Jain

Svetha Venkatesh

Brian C. Lovell

Hosted by

International Association for Pattern Recognition

Australian Pattern Recognition Society

IEEE COMPUTER SOCIETY

Los Alamitos, California

Washington • Brussels • Tokyo