PERGAMON

Pattern Recognition 34 (2001) 469–485

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Recognizing articulated objects in SAR images

Grinnell Jones III, Bir Bhanu*

Center for Research in Intelligent Systems, University of California, Riverside, CA 92521, USA

Received 25 June 1999; received in revised form 27 October 1999; accepted 27 October 1999

Abstract

This paper presents the first sucessful approach for recognizing articulated vehicles in real synthetic aperture radar (SAR) images. This approach is based on invariant properties of the objects. Using SAR scattering center locations and magnitudes as features, the invariance of these features with articulation (e.g. turret rotation of a tank) is shown for XPATCH-generated synthetic SAR signatures and actual signatures from the MSTAR (public) data. Although related to geometric hashing, our recognition approach is specifically designed for SAR, taking into account the great azimuthal variation and moderate articulation invariance of SAR signatures. We present a basic recognition system for the XPATCH data, using scatterer relative locations, and an improved recognition system, using scatterer locations and magnitudes, that achieves excellent results with the more limited articulation invariance encountered with the real SAR targets in the MSTAR data. The articulation invariant properties of the objects are used to characterize recognition system performance in terms of probability of correct identification as a function of percent invariance with articulation.

Keywords: Automatic target recognition; Geometric invariance; Magnitude invariance; Azimuth variance; Model-based recognition; Characteristics of scattering centers

1. Introduction

This paper establishes the existence of articulation invariant object features in SAR images and presents the first successful approach to recognizing articulated objects in real SAR images. We are concerned with recognizing articulated vehicles, starting with SAR image chips of various target vehicles and ending with identification of the specific vehicle type (e.g., a T72 tank). The major challenge is that the vehicles can be in articulated configurations such as a tank with its turret rotated or the SCUD missile erect on its mobile launcher. We approach the problem of recognizing articulated objects from the fundamentals of SAR images. We characterize the SAR image azimuthal variance to determine the number of models required. We demonstrate (and measure) the invariance with target articulation. Based on these invariants, we develop SAR specific recognition methods that use non-articulated models to successfully recognize articulated versions of the same objects. This is done for both synthetic target images, generated by the XPATCH radar signature prediction code [1] and real SAR images of actual vehicles from the MSTAR (Public) targets data set [2]. The XPATCH data allow a large number of precisely controlled experiments, while the MSTAR data provide more challenging real-world cases. The paper is organized so that the three sections following Section 2 on Background and Related Work: SAR Object Characteristics, SAR Recognition System, and Recognition Results, each address the simulated XPATCH data and then the real MSTAR data.

2. Background and related work

2.1. Background

Automatic target recognition (ATR) is the use of computer processing to detect and recognize object (target)

* Corresponding author. Tel.: +1-909-787-3954; fax: +1-909-787-3188.

E-mail address: bhanu@cris.ucr.edu (B. Bhanu).

0031-3203/00/$20.00
PII: S0031-3203(99)00218-6

signatures in sensor data. General reviews of ATR concepts and technologies can be found in Refs. [3–6]. ATR systems generally have separate detection and recognition stages. The goal of the detection stage is to eliminate most of the sensor data from further consideration and find small regions potentially containing the objects of interest. The goal of the recognition stage is to classify or identify the objects. The term *classification* is generally used for a coarse categorization (e.g., the object is a tank, not a truck) and *identification* is used for a fine categorization (e.g., a specific type of tank). Detecting objects in SAR imagery typically involves a prescreen stage (e.g., a constant false alarm rate (CFAR) thresholding technique [7]) and a discriminator stage to separate the objects from the background clutter (a detailed description of features can be found in Ref. [8]). Other methods for object detection include: filters [9], using the variations in return from man-made objects and natural clutter with changes in image resolution [10], likelihood images with image relational graphs [11] and neural networks [12]. There are several different categories of algorithms used for recognition in ATR systems: detection theory, pattern recognition, artificial neural networks and model-based recognition [6]. Several of these can be used for detection as well as recognition and several ATR systems use combinations of approaches.

2.1.1. Detection theory

The detection theory approach uses filters to separate the distributions of object and clutter signatures so they can be distinguished by a simple statistical test. This approach has been applied to classification of different objects in SAR images using minimum noise and correlation energy (MINACE) filters [9] and optimal trade-off distance classifier correlation filters (OTDCCFs) [13] (which use a distance measure for classification instead of just the output correlation peak). Because of the azimuth variation of SAR signatures, multiple filters are used for each object, each covering a range of target azimuths.

2.1.2. Pattern recognition

The pattern recognition approach uses SAR image templates [14] or feature vectors (such as topographical primal sketch (TPS) features, based on zero crossings of directional derivatives [15]) and determines the best match between the object image and an exemplar database. In contrast to detection theory algorithms that are derived by using statistical models of raw data, pattern recognition uses more ad hoc approaches to the definition and extraction of the features used to characterize objects. In contrast to the later model-based approach, the pattern recognition approaches usually use global measures to represent the object. These global features are very susceptible to occlusion and articulation effects. Many of the pattern recognition approaches suffer from an exhaustive search of a database of exemplars, although a hierarchical index of distance transforms with composite models had been used [16] to convert the problem to a tree search.

2.1.3. Neural networks

The neural network approach uses learning by example to discover and use signature differences that distinguish different types of objects. Adaptive resonance theory networks [12] have been used to categorize SAR object aspects and learn SAR object templates. Feature space trajectory neural networks have been used for SAR detection and classification [17]. One problem for neural network approaches is to achieve good performance with a range of object signatures (with occlusion, articulation and non-standard configurations) and varying background clutter, given limited amounts of training data.

2.1.4. Model-based recognition

The model-based recognition approach typically uses multiple local features (involving object parts and the relationships between the parts) and matching of sensor data to predictions based on hypotheses about the object type and pose. A current state of the art example is the moving and stationary object acquisition and recognition (MSTAR) program that uses a search module for hypothesis generation and refinement [18], a feature prediction module that captures the object signature phenomenology [19] and a matching module [20]. While most model-based systems are optimized for unobscured objects, one model-based approach, partial evidence reconstruction from object restricted measures (PERFORM) [21], uses a linear signal decomposition, direction of arrival pose estimation technique and attempts to overcome the difficulties of recognizing an obscured object by breaking the problem into smaller parts (e.g. separately recognize the front, center and rear regions of the vehicle and fuse the results).

2.2. Related work

There is no published paper, by others, on the recognition of articulated objects in SAR images. The detection theory, pattern recognition and neural network approaches to SAR recognition, as discussed in the previous subsection, all tend to use global features that are optimized for standard, non-articulated, non-occluded configurations. Approaches that rely on global features are not appropriate for recognizing articulated objects because articulation changes global features like the object outline and major axis [22]. Template matching (whether the templates are built by hand, as in Ref. [14], or developed by neural networks, as in Ref. [12]) has been successfully used for object recognition with SAR images, but this approach is not suitable for recognizing articulated objects since there will be a combinatorial explosion of the number of templates with varying articu-

lations. Some of the ATR techniques, e.g., MINACE filters [23] and PERFORM [21], have reported limited test results for small amounts of occlusion, typically 25% or less, which would indicate some potential for articulated object recognition. The standard approaches used for articulated object recognition in optical images (such as recognition by components [24], constrained models of parts [25] and joint articulation [26]) are not appropriate for the relatively low-resolution, non-literal nature and complex part interactions of SAR images; which are successfully handled by using articulation invariants in our approach presented in this paper. Our previous work in this area [27] was limited to XPATCH data and only used the location invariance of the SAR scattering centers with articulation as the basis for recognition. This paper establishes that where the scattering center locations are invariant with articulation, the corresponding magnitudes are also invariant (within a small tolerance); it significantly expands the prior work to real SAR data; and it demonstrates that an improved recognition system, using the scatterer magnitudes as well as locations, can successfully overcome the greater variability of the real SAR data.

3. SAR object characteristics

The typical detailed edge and straight line features of man-made objects in the visual world, do not have good counterparts in SAR images for sub-components of vehicle-sized objects at 6 in to 1 ft resolution. However, there are a wealth of peaks corresponding to scattering centers. The relative locations of SAR scattering centers, determined from local peaks in the radar return, are related to the aspect and physical geometry of the object, independent of translation and serve as distinguishing features. In addition to the scatterer locations, the magnitudes of the peaks are also features that we use in this paper.

3.1. Objects, SAR images and scattering centers

3.1.1. XPATCH objects, SAR images and scattering centers

The XPATCH radar signature prediction code is used to generate image chips of objects at 360 azimuth angles (at a 15° depression angle) from CAD models of articulated and non-articulated versions of three tanks (T72, T80 and M1a1) and a SCUD missile launcher. Fig. 1 shows the CAD models in some articulated positions; these are modified commercial CAD models with numbers of surface facets ranging from 5345 to 32,954. The articulated versions of the tanks have 60 and 90° counter-clockwise turret rotations, the SCUD launcher has the missile erect. The scattering center locations are determined by finding local eight-neighbor maxima in the radar image. Examples, at various azimuths, of the object geometry, SAR image and scattering center locations (as black squares) are shown for both articulated and non-articulated cases of the T72 tank (Fig. 2) and the T80 tank (Fig. 3). (Figs. 2 and 3 are not to scale and the image is displayed at eight intensity levels, the scattering center map at 256 levels.) Unless otherwise noted, the XPATCH data and results are for 6 in resolution SAR.

3.1.2. MSTAR objects, SAR images and scattering centers

Photo images of the MSTAR articulated objects used in this paper, T72 tank serial number (#) a64 and ZSU 23/4 antiaircraft gun #d08, are shown in Figs. 4 and 5. The articulated test data have the turret rotated 315° counter-clockwise, compared to the non-articulated models with the turret straight forward. Regions of interest (ROI) are found in the MSTAR SAR object chips by reducing speckle noise using the Crimmins algorithm in Khoros [28], thresholding at the mean plus two standard deviations, dilating to fill small gaps among regions, eroding to have one large ROI and little regions, discarding the small regions with a size filter

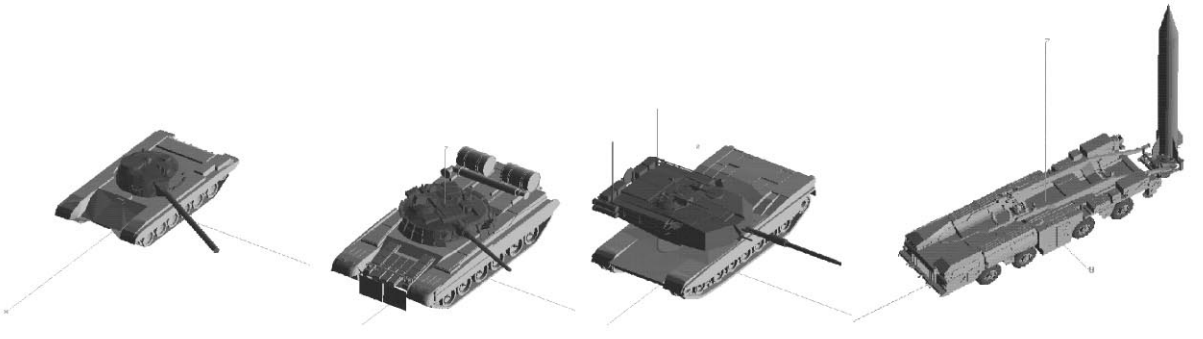


Fig. 1. Models of articulated objects (not to scale) for XPATCH.

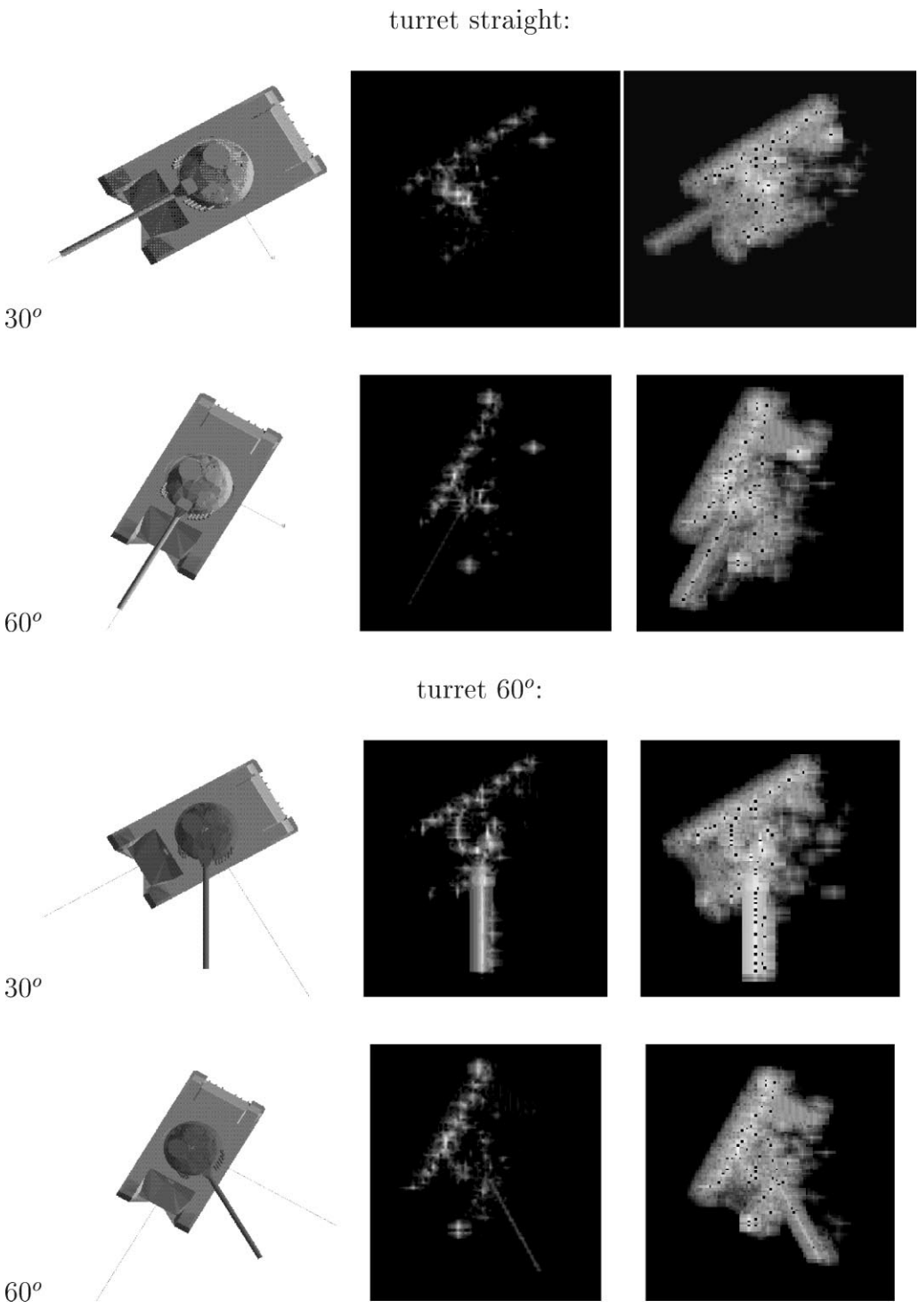


Fig. 2. T72 tank geometry, XPATCH SAR image and scattering centers for 30 and 60° hull azimuths.

and dilating to expand the extracted ROI. The parameters used in extracting ROIs are held constant for all the results reported. The scattering centers are extracted from the SAR magnitude data (within the boundary contour of the ROI) by finding local eight-neighbor maxima. Example SAR images and the

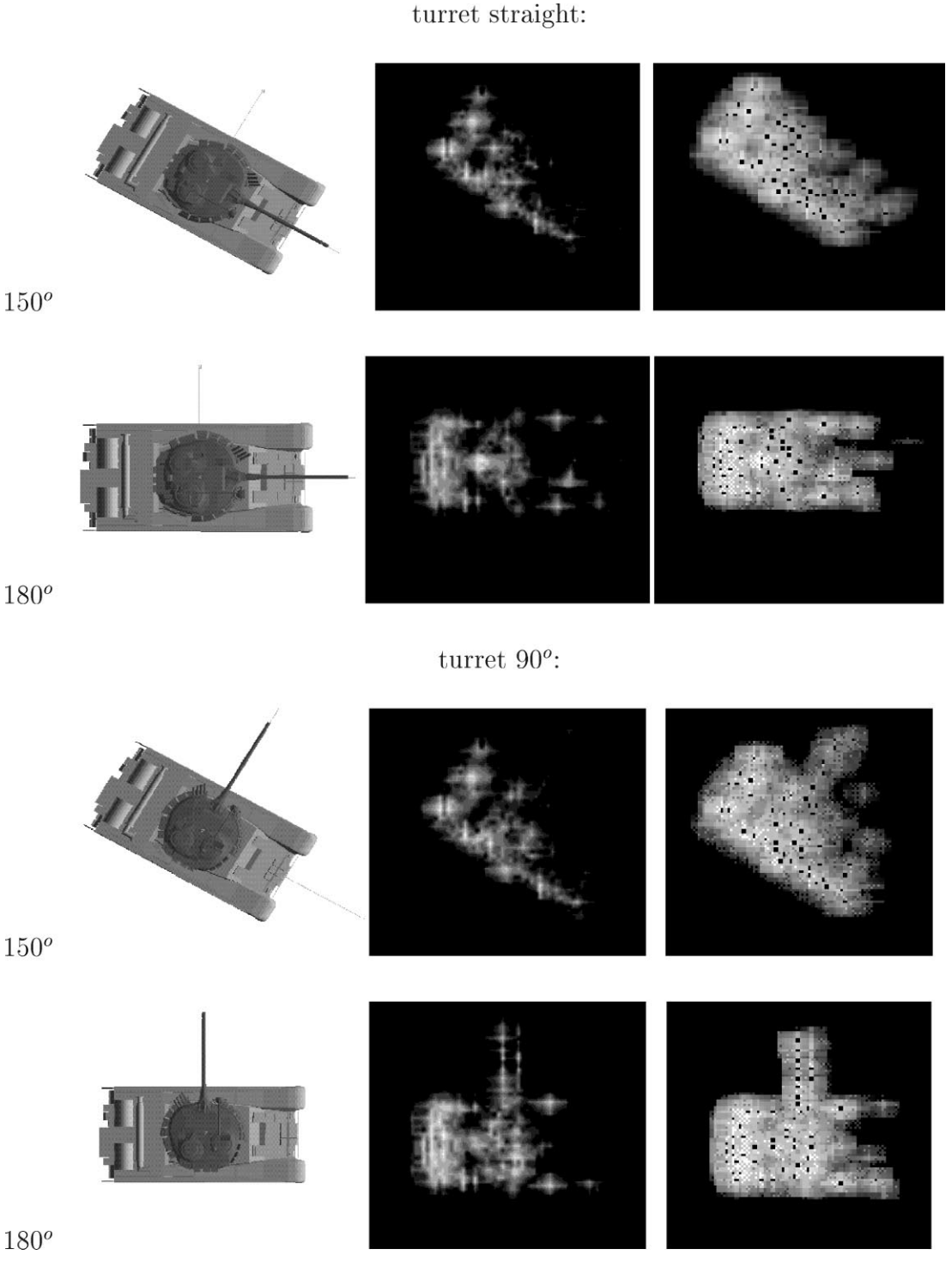


Fig. 3. T80 tank geometry, XPATCH SAR image and scattering centers for 150 and 180° hull azimuths.

regions of interest (ROI), with the locations of the scattering centers superimposed (as black squares), are shown in Fig. 6 for baseline and articulated versions of the T72 and ZSU. The MSTAR data are all at 1 ft resolution and the articulated object data are at 30° depression angle.

(a) turret straight. (b) turret articulated.

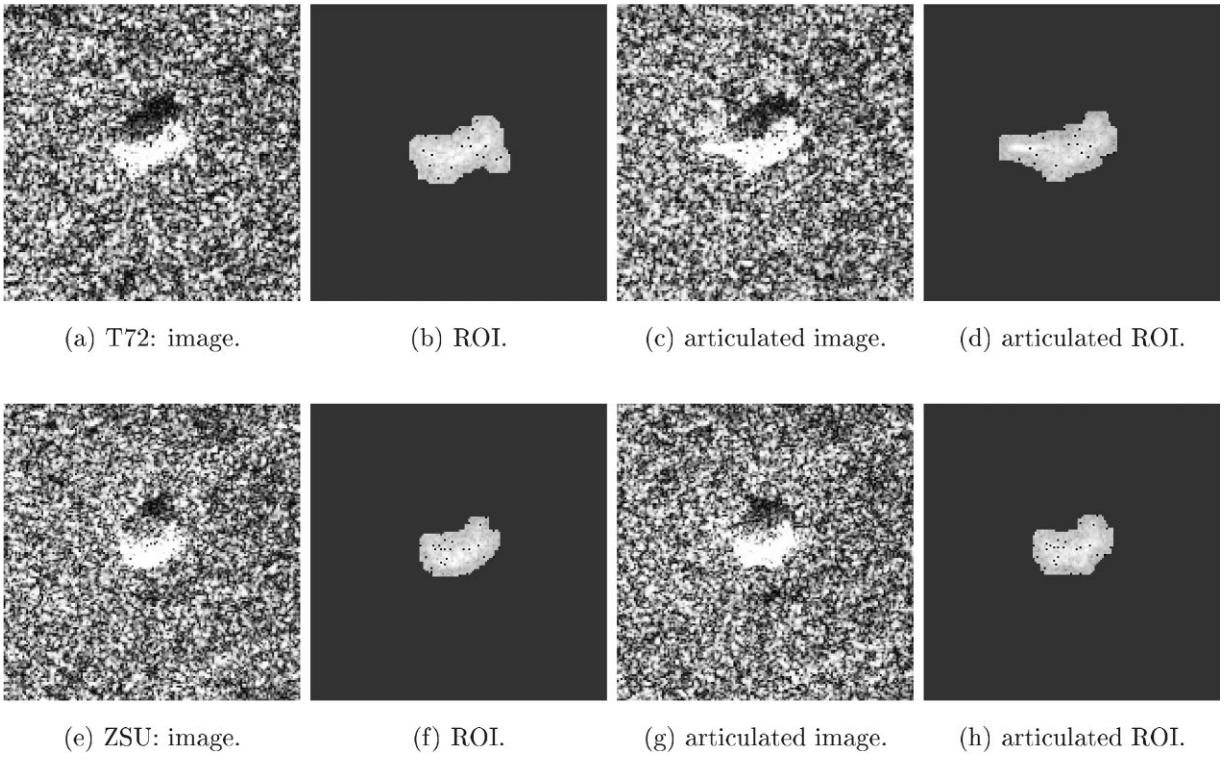
Fig. 4. T72 tank #a64.

(a) turret straight. (b) turret articulated.

Fig. 5. ZSU 23/4 antiaircraft gun #d08.

3.2. Azimuthal variance

The typical rigid-body rotational transformations for viewing objects in the visual world do not apply much for the specular radar reflections of SAR images. This is because a *significant* number of features *do not* typically persist in radar images over a few degrees of object rotation. Since the radar depression angle is generally known, the significant unknown object rotation is (360°) in azimuth. *Azimuth invariance* or persistence can be expressed in terms of the percentage of scattering center locations that are unchanged over a certain span of azimuth angles. It can be measured (for some base azimuth θ_0) by rotating the pixel locations of the scattering centers from an image at azimuth θ_0 by an angle $\Delta\theta$ and comparing the resulting range and cross-range locations with the scatterer locations from an image of the same object at azimuth $\theta_0 + \Delta\theta$. We consider two cases for the precision of the scatterer location match. In the 'exact match' cases the center of the rotated scatterer pixel from the image at θ_0 azimuth is within the pixel boundaries of a corresponding scatterer in the image at $\theta_0 + \Delta\theta$. In the 'within 1 pixel' cases, the scatterer location is allowed to move into one of the 8 adjacent pixel locations. To determine scattering center locations that persist over a span of angles, there is an additional constraint that for a matching scattering center to 'persist' at the kth span $\Delta\theta_k$, it must have been a persistent scattering center at all smaller spans $\Delta\theta_j$, where $0 \leqslant j < k$. Averaging the results of these unchanged scattering center locations over 360 base azimuths gives a 'mean location invariance' for an object.



(a) T72: image. (b) ROI. (c) articulated image. (d) articulated ROI.

(e) ZSU: image. (f) ROI. (g) articulated image. (h) articulated ROI.

Fig. 6. MSTAR SAR images and ROIs (with peaks) for T72 tank #a64 and ZSU 23/4 #d08 at 66° azimuth.

3.2.1. XPATCH Azimuthal Variance

The XPATCH images are generated with a consistent origin so no translation is required for registration and the corresponding scatterers are evident. (With the XPATCH data we accomplish the rotation in the slant plane, instead of the projection to the ground plane — rotation — projection back to the slant plane, but at a 15° depression angle the difference is small.) Fig. 7 shows the location invariance of the strongest 50 T72 tank scattering centers as a function of azimuth span for both 6 in and 1 ft resolution XPATCH data. The invariance of the 1 ft resolution cases is similar to (but slightly less than) the 6 in data. For both the exact match and within one-pixel cases, significant numbers of scattering centers do not persist over a few degrees of rotation. Because of this azimuthal variation and the goal of recognizing articulated objects, in this research we use object models at 1° intervals.

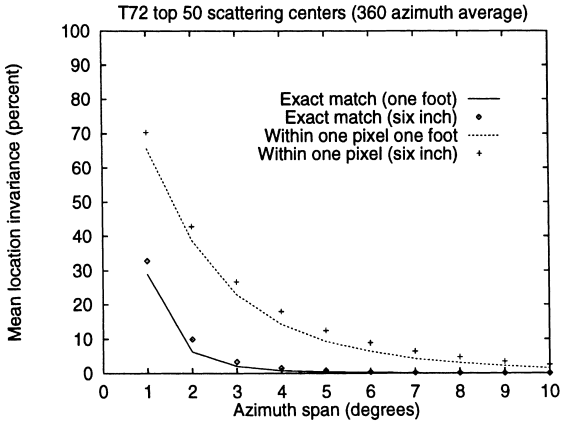


Fig. 7. XPATCH scatterer location persistence.

3.2.2. MSTAR Azimuthal variance

With the 30° depression angle MSTAR data we accomplish the azimuthal rotation by projection from the slant plane to the ground plane, rotation in the ground plane and projection back to the slant plane. Since the objects in the MSTAR chips are not registered, we calculate the azimuth invariance as the maximum number of corresponding scattering centers (whose locations match within a given tolerance) for the optimum integer pixel translation. This method of registration by finding the translation that yields the maximum number of correpondences has the limitation that for very small or no

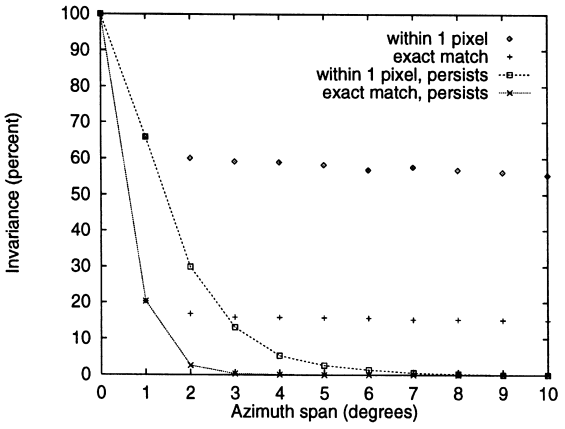


Fig. 8. Scatterer location persistence, MSTAR T72 #132.

actual invariance it may find some false correspondences and report a slightly higher invariance than in fact exists. Fig. 8 shows an example of the mean scatterer location invariance (for the 40 strongest scatterers) as a function of azimuth angle span using T72 tank #132, with various definitions of persistence. The cases labeled 'persists' in Fig. 8 enforce the constraint that the scatterer exist for the entire span of angles and are similar to the results previously shown in Fig. 7 with XPATCH data. For both the XPATCH simulated data and the MSTAR real data, very few scatterers continuously persist for even 5°. In the two cases not labeled 'persists' in Fig. 8, scintillation is allowed (i.e. a scatterer can disappear and then reappear as a result of constructive/destructive interference) and the location invariance declines slowly with azimuth span. The 'within one-pixel' results (that allow scintillation) are consistent with the one foot ISAR results of Dudgeon [29], whose definition of persistence allowed scintillation.

3.3. Geometric invariance with articulation

The existance of scattering center locations that are invariant with object articulation is crucial to successfully recognizing articulated objects with non-articulated models, thus avoiding the combinatorial problem of modeling 360 articulations × 360 azimuths.

3.3.1. XPATCH location invariance

Fig. 9 demonstrates the existence of articulation invariants with XPATCH data, where the locations of scattering centers are indicated by the black squares. In the SCUD launcher example, with the radar directed (from the left in Fig. 9) at the cab end of the launcher, many of the scattering center locations on the launcher itself are independent of whether the missile is erect or down. In the similar view of the T72 tank, many of the details from the hull are independent of the turret angle. An example of XPATCH articulation invariance is shown in Fig. 10(a), which plots the percentage of the strongest 50 scattering centers for the T72 tank that are in exactly the same location (the 'exact match' case) with the turret

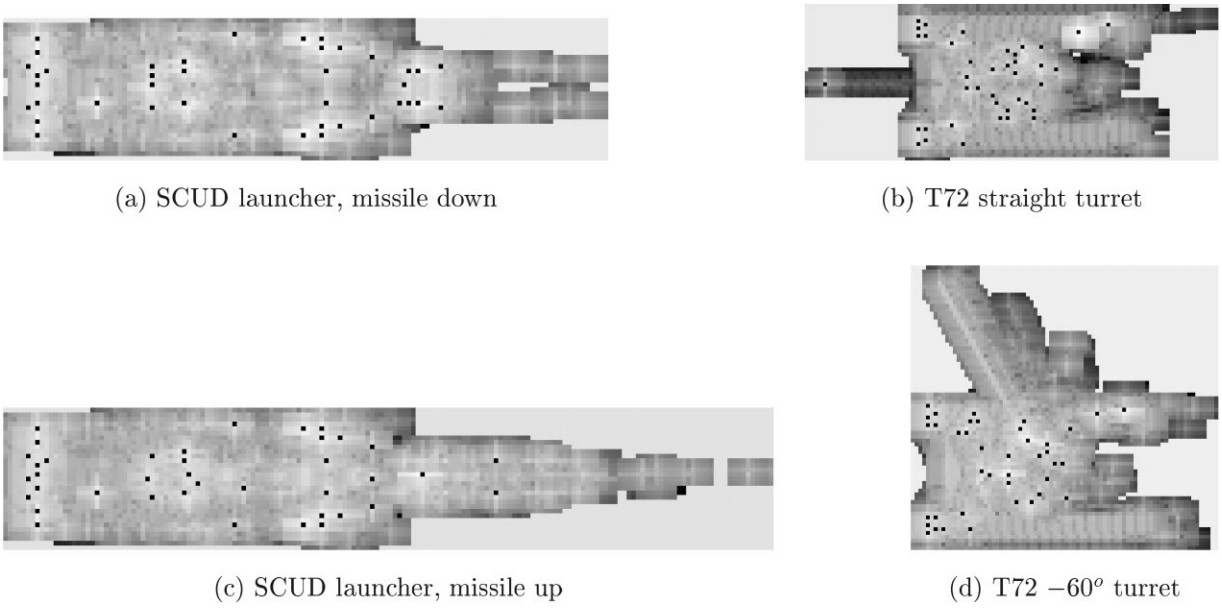
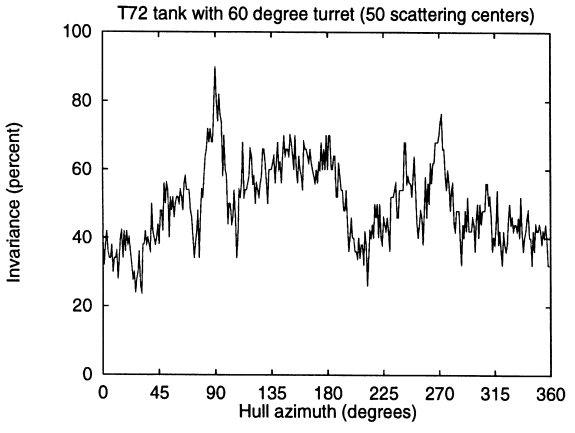
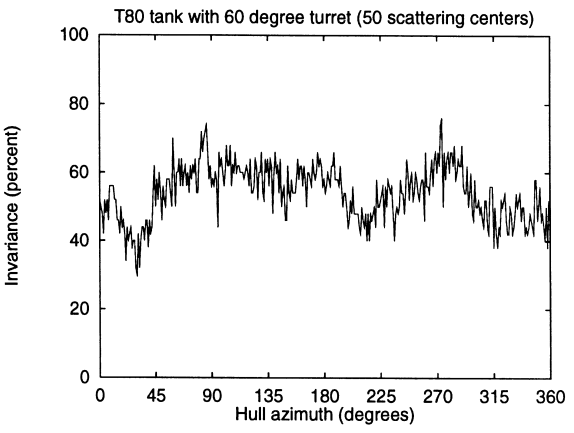


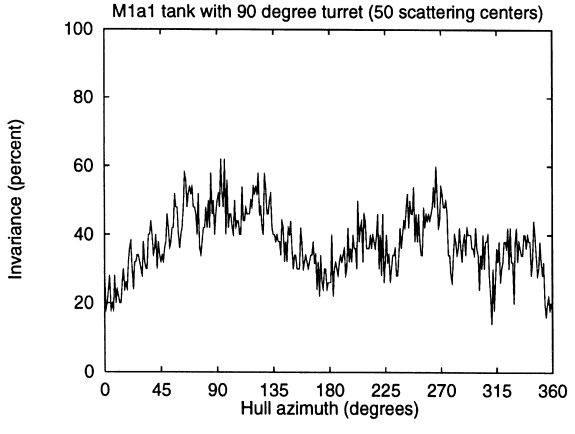
Fig. 9. XPATCH SCUD launcher and T72 tank articulation examples.



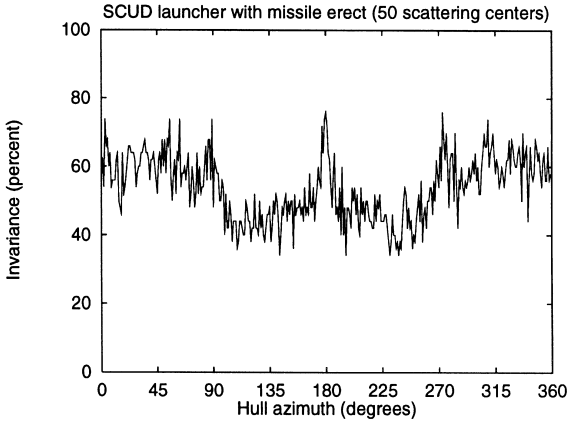
(a) T72 (60^o turret).



(b) T80 (60^o turret).



(c) M1a1 (90^o turret).



(d) SCUD launcher (missile up).

Fig. 10. XPATCH articulation invariants.

rotated 60° as they are with the turret straight forward, for each of 360 azimuths. Similar examples of articulation invariance are shown in Fig. 10 for the T80 tank with 60° turret, M1a1 tank with 90° turret, and the SCUD launcher with the missile erect. Fig. 11 shows the articulation invariance at 6 in and 1 ft resolutions for the T72 tank with a 90° turret; on average 48% of the locations are invariant for the six inch data and the one foot XPATCH results are similar.

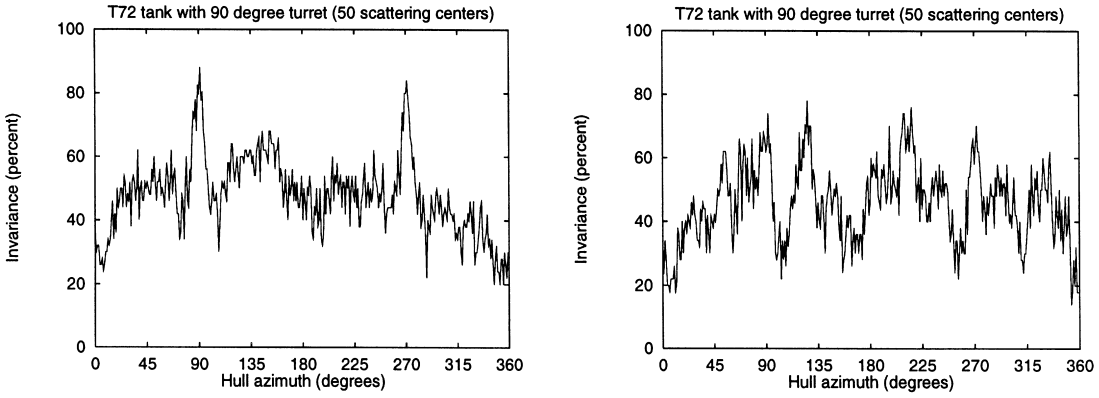
Table 1 shows the mean and standard deviation values of the average (over 360 azimuths) percent articulation invariance for the various XPATCH (6 in resolution) articulated objects and the overall average. The mean values are similar for the 20 and 50 scattering center cases. The smaller average articulation invariance for the M1a1 tank is expected, because the M1a1 tank has a comparatively much larger turret than the other tanks (see Fig. 1(c)).

3.3.2. MSTAR location invariance

Because the object and ROI are not registered in the MSTAR data, we define the scattering center location invariance with respect to *articulation* as the maximum number of corresponding scattering centers (whose locations match within a stated tolerance) for the optimum integer pixel translation. Given an original version of a SAR object image with n scattering centers, represented by points at pixel locations $P_i = (x_i, y_i)$ for $1 \leqslant i \leqslant n$ and a translated, distorted version $P'_j = (x'_j, y'_j)$ $(1 \leqslant j \leqslant n)$ at a translation $t = (t_x, t_y)$, we define a *match* between points P'_j and P_i as

$$M_{ij}(t) = \begin{cases} 1 & \text{if } |x'_j - t_x - x_i| \leqslant l \text{ and } |y'_j - t_y - y_i| \leqslant l, \\ 0 & \text{otherwise,} \end{cases}$$

where $l = 0$ for an 'exact' match and $l = 1$ for a match 'within one pixel'. The scatterer location invariance, L_n,



(a) Six inch resolution. (b) One foot resolution.

Fig. 11. Effect of resolution on XPATCH articulation invariance.

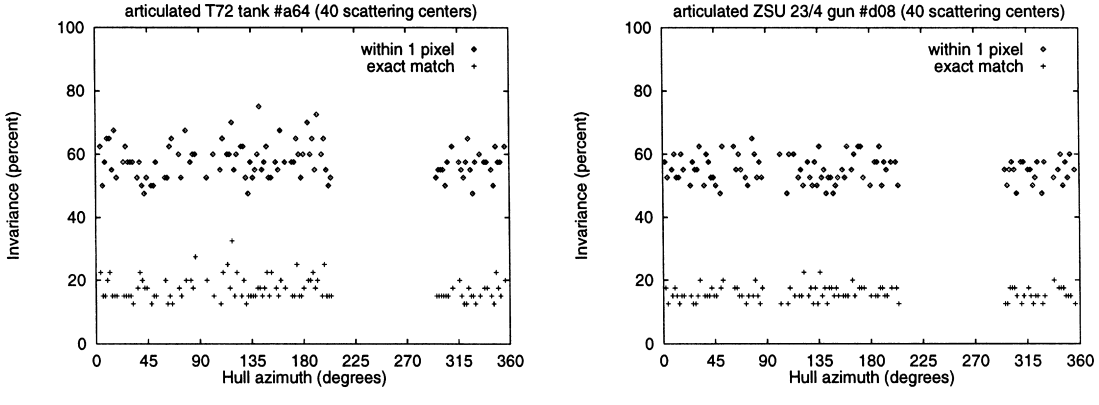
Table 1
Scatterer location invariance for XPATCH objects with articulation (in %)

	20 scatterers		50 scatterers	
	Mean	Std. dev.	Mean	Std. dev.
SCUD missile up	52.76	12.98	53.67	9.79
T72 60°turret	47.92	15.96	49.94	11.71
90°turret	45.22	15.61	48.00	12.03
M1a1 60°turret	35.96	16.79	37.96	9.32
90°turret	37.18	15.35	37.66	9.06
T80 60°turret	58.49	16.52	54.17	8.02
90°turret	58.31	14.61	53.84	7.94
Average	47.98	15.41	47.89	9.70

of n scatterers, expressed as a percentage of matching points, is given by

$$L_n = \max_t \left\{ \frac{100}{n} \sum_{j=1}^{n} \min\left(\left(\sum_{i=1}^{n} M_{ij}(t) \right), 1 \right) \right\},$$

where each point P'_j is restricted to at most one match. (The 'min' distinguishes between the no match case, 0, and one or more matches, 1). Fig. 12 shows the location invariance, L_{40}, of the strongest 40 scattering centers with articulation for MSTAR T72 tank #a64 and ZSU 23/4 antiaircraft gun #d08 (at a 30° depression angle) as a function of the hull azimuth. The mean and standard deviation for percent location invariance with articulation of the MSTAR T72 and ZSU23/4 are shown in Table 2 for the exact match and within one pixel match cases. Note that the 16.5% exact match articulation invariance in Table 2 for the real MSTAR data is



(a) T72 tank. (b) ZSU 23/4.

Fig. 12. MSTAR scatterer location invariance with articulation.

Table 2
Scatterer location invariance for MSTAR objects with articulation (in %)

Object	Turret angle	Exact match invariance		Within 1 pixel invariance	
		Mean	Std. dev.	Mean	Std. dev.
T72 #a64	315	17.17	1.47	57.83	2.23
ZSU #d08	315	15.69	0.91	55.05	1.72
Average		16.45		56.47	

considerably less than the comparable 48% invariance for the XPATCH simulated data in Table 1.

3.4. Magnitude invariance with articulation

This section establishes another useful invariant feature that the scatterer magnitudes associated with the invariant scattering center locations are also invariant (within limits).

3.4.1. XPATCH magnitude invariance

Fig. 13 shows a comparison of the magnitude of the peak return (from the strongest scattering center) for the XPATCH SCUD launcher with the missile erect versus the launcher with the missile down as a function of azimuth. For the 222 (of 360) azimuth values where the range and cross-range pixel locations of these peak returns are identical (i.e. location invariant with articulation) the mean magnitude difference is 0.06% and the standard deviation is 1.25%.

3.4.2. MSTAR magnitude invariance

Using a scaled scatterer amplitude (S), expressed as a radar cross section in square meters, given by $S = 100 + 10\log_{10}(i^2 + q^2)$, where i and q are the components of the complex radar return, we define a percent amplitude change (A_{jk}) as: $A_{jk} = 100(S_j - S_k)/S_j$. (This form allows a larger variation for the stronger signal returns.) A location and magnitude match $Q_{jk}(t)$ is given by

$$Q_{jk}(t) = \begin{cases} 1 & \text{if } M_{jk}(t) = 1 \text{ and } |A_{jk}| \leqslant l_A, \\ 0 & \text{otherwise,} \end{cases}$$

where l_A is the percent amplitude change tolerance. The scatterer magnitude and location invariance (I_n), expressed as a percentage of n scatterers, is given by:

$$I_n = \max_t \left\{ \frac{100}{n} \sum_{k=1}^{n} \min\left(\left(\sum_{j=1}^{n} Q_{jk}(t) \right), 1 \right) \right\}.$$

Fig. 14 shows the probability mass functions (PMFs) for percent amplitude change for the strongest 40 articulated vs. non-articulated scattering centers of MSTAR T72 tank #a64 and ZSU #d08. Curves are shown both for the cases where the scattering center locations correspond within a 1-pixel tolerance and for all the combinations of scatterers whose locations do not match. The mean and standard deviation for these matching and non-matching scatterers and the crossover points for the PMFs are given in Table 3. Table 4 shows the mean and standard deviation for the percent location and magnitude invariance (within a one-pixel location tolerance and an amplitude change tolerance of $\pm 9\%$) of the strongest 40 scatterers for the MSTAR articulated data.

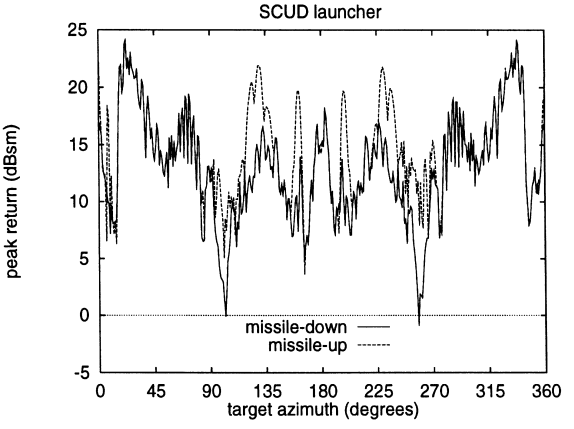
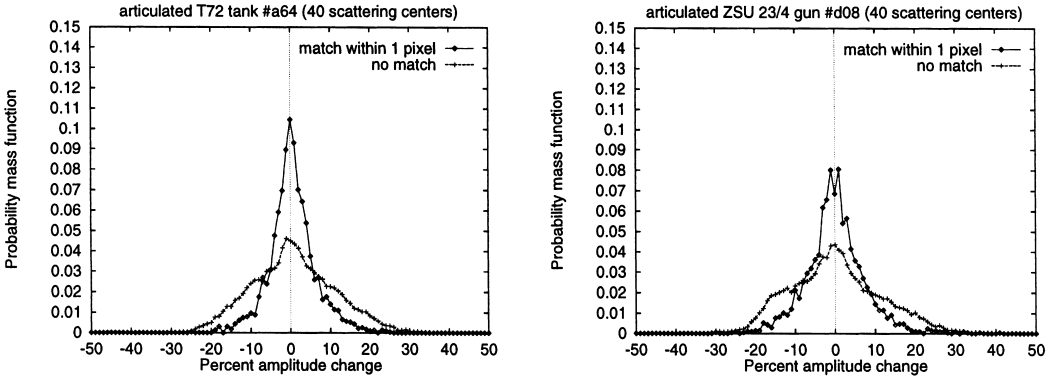


Fig. 13. XPATCH SCUD launcher peak return invariance.

4. SAR recognition system

Establishing an appropriate local coordinate reference frame is critical to reliably identifying objects (based on locations of features) in SAR images of articulated objects. The object articulation problem requires the use of a local coordinate system; global coordinates and global constraints do not work, as illustrated in Figs. 2, 3, 6 and 9, where the center of mass and the principal axes of the object change with articulation. In the geometry of a SAR sensor the 'squint angle', the angle between the flight path (cross-range direction) and the radar beam (range direction), can be known and fixed at 90°. Given the SAR squint angle, the image range and cross-range directions are known and any local reference point chosen, such as a scattering center location, establishes a reference coordinate system. (The scattering centers are local maxima in the radar return signal.) The relative distance and direction of the other scattering centers can be expressed in radar range and cross-range coordinates, and naturally tessellated into integer buckets that correspond to the radar range/cross-range bins. The recognition system takes advantage of this natural system for



(a) T72 tank. (b) ZSU 23/4.

Fig. 14. MSTAR scatterer magnitude invariance with articulation.

Table 3
MSTAR scatterer percent amplitude change

Articulation	Within 1 pixel		No match		
	Mean	Std. dev.	Mean	Std. dev.	x-over
T72 #a64	0.51	5.91	0.75	10.44	$-5/+6$
ZSU #d08	0.06	7.44	0.08	11.37	± 9

Table 4
MSTAR scatterer location and magnitude invariance (in %)

Articulation	Mean	Std. dev.
T72 #a64	53.47	2.63
ZSU #d08	47.98	2.22
Average	50.78	

SAR, where selecting a single basis point performs the translational transformation and fixes the coordinate system to a 'local' origin.

Our model-based recognition system uses standard non-articulated models of the objects (at 1° azimuth increments) to recognize the same objects in non-standard articulated configurations. Using a technique like geometric hashing [30], the relative positions of the scattering centers in the range (R) and cross-range (C) directions are indices to a look-up table of labels that give the associated object type and pose. This is an efficient search for *positive* evidence that generates votes for the appropriate object (and azimuth). The models and recognition engine have evolved from the 2D version [27] for the XPATCH data, which uses only the relative distances and the 'exact' scatterer locations, to a 6D version for the more challenging MSTAR data, which uses more local features and accommodates a 'within 1 pixel' scatterer location uncertainty. In the 6D version the model look-up table labels contain four additional features: range and cross-range position of the 'origin' and the magnitudes (S) of the two scatterers. The model construction algorithm for the 6D recognition system is outlined in Fig. 15.

For ideal data one could use the strongest scatterer as the origin, however any given scatterer could actually be spurious or missing due to the effects of noise, articulation, occlusion, or non-standard configurations. Thus, for both the 2D and 6D versions, we model and use (for recognition) all the scattering center locations in turn as the origin, so the size of the look-up table models and the number of nominal relative distances considered in the recognition of a test image is $n(n-1)/2$, where n is the number of the strongest scattering centers used.

In contrast to many model-based approaches to recognition [31], we are not 'searching' all the models; instead we are doing table look-ups based on relative distances between the strongest scatterers in the test image. Each query of the look-up table may generate votes for one or more potential candidate solutions. In the 2D version, the look-up table results directly generate votes for object–azimuth pairs and when all the combinations of scatterer pairs are considered, the object–azimuth pair with the most accumulated votes is chosen.

In the 6D version, further comparison of each test data pair of scatterers with the model look-up table result(s) provides information on the range and cross-range translation and the percent magnitude changes for the two scatterers. Limits on allowable values for translations and magnitude changes are used as constraints to reduce the number of false matches. (The number of scattering centers used and the various constraint limits are design parameters that are optimized, based on experiments, to produce the best recognition results.) Here, votes are accumulated in a 4D space: object, azimuth, range and cross-range translation. Also a (city-block) weighted

1. For each model Object do 2
2. For each model Azimuth do 3, 4, 5
3. Obtain the location (R, C) and magnitude (S) of the strongest n scatterers.
4. Order (R, C, S) triples by descending S.
5. For each origin O from 1 to n do 6
6. For each point P from O+1 to n do 7, 8
7. $dR = R_P - R_O$; $dC = C_P - C_O$.
8. At look-up table location dR, dC append to list entry with: Object, Azimuth, R_O, C_O, S_O, S_P.

Fig. 15. 6D model construction algorithm.

1. Obtain from test image the location (R, C) and magnitude (S) of n strongest scatterers.
2. Order (R, C, S) triples by descending S.
3. For each origin O from 1 to n do 4
4. For each point P from O+1 to n do 5, 6
5. $dR = R_P - R_O$; $dC = C_P - C_O$.
6. For DR from dR-1 to dR+1 do 7
7. For DC from dC-1 to dC+1 do 8, 9, 10
8. weighted_vote = |DR| + |DC|.
9. Look up list of model entries at DR, DC.
10. For each model entry E in the list do 11
11. IF $|\text{tr} = R_O - R_E| <$ translation_limit and $|\text{tc} = C_O - C_E| <$ translation_limit
and $|1 - S_{EO}/S_O| <$ magnitude_limit and $|1 - S_{EP}/S_P| <$ magnitude_limit
THEN increment accumulator array [Object, Azimuth, tr, tc] by weighted_vote.
12. Query accumulator array for each Object, Azimuth, tr and tc, summing the votes in a 3x3 neighborhood in translation subspace about tr, tc; record the maximum vote_sum and the corresponding Object.
13. IF maximum vote_sum > threshold
THEN result is Object ELSE result is "unknown".

Fig. 16. 6D recognition algorithm.

voting method is used to reduce the impact of the more common small relative distances. To accommodate some uncertainty in the scattering center locations, the eight neighbors of the nominal range and cross-range relative location are also probed in the look-up table and the final translation results are summed over a 3×3 neighborhood in the translation subspace. This voting in translation space, in effect, converts the consideration of scatterer pairs back into a group of scatterers at a consistent translation. The process is repeated with different scattering centers as reference points, providing multiple 'looks' at the model database to handle spurious scatterers that arise due to articulation, noise or other factors. To handle identification with 'unknown' objects, we introduce a criteria for the quality of the recognition result (e.g., the votes for the potential winning object exceed some threshold $v_{\min}$). The recognition algorithm for the 6D system is given in Fig. 16.

5. Recognition results

5.1. Results using XPATCH SAR data

Results with the XPATCH data are based on using a 2D recognition algorithm that is an earlier, simpler version of the 6D algorithm described in Fig. 16. The 2D algorithm uses only the relative range and cross-range distances; it does not compute the appropriate translation; it only considers the 'exact match' scatterer location; and it does not use the scatterer magnitude information. The experimental results of 2520 trials with XPATCH generated articulated test objects (SCUD launcher with the missile up, T72, M1a1 and T80 tanks with 60° and 90° turrets) for the 2D recognition system using 50 scattering centers and non-articulated models (missile down, turrets straight forward) are shown as a confusion matrix in Table 5. The overall forced recognition performance is a 93.14% probability of correct identification (PCI). The 'number of scattering centers used' is a design parameter that can be tuned to optimize the performance of the 2D recognition system as shown in Fig. 17, where each point on the plot of PCI vs. 'number of scattering centers used' is the result of 2520 trials. While the maximum recognition rate is achieved at 50 scattering centers (93.14%), a more optimal system with 35 scattering centers achieves similar performance (92%) with slightly less than half the storage and twice the speed of 50 scattering centers.

The detailed recognition results can be related to the articulation invariance of the objects. The 33 recognition failures for the T72 tank with the turret at 90° in Table 5 are plotted (as 'diamonds') on the curve of percent invariance vs. azimuth in Fig. 18. These results show that recognition failures generally occur for azimuths where the percent invariance is low. Fig. 19 shows how the PCI varies with the percent articulation invariance for the forced recognition case with the 2D recognition system,

Table 5
Confusion matrix for XPATCH articulated forced identification results (50 scatterers)

Articulated test objects	Non-articulated models			
	SCUD Down	T72	M1a1	T80
		0°	turret	
SCUD missile up	360			
T72 60° turret		335	7	18
90° turret		327	8	25
M1a1 60° turret		1	300	59
90° turret		2	305	53
T80 60° turret				360
90° turret				360

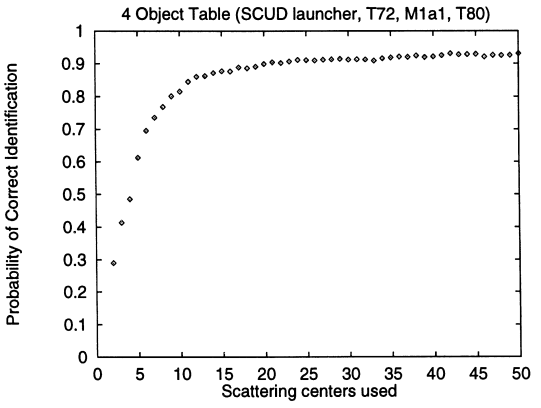


Fig. 17. Effect of the number of scattering centers on forced recognition rate (XPATCH).

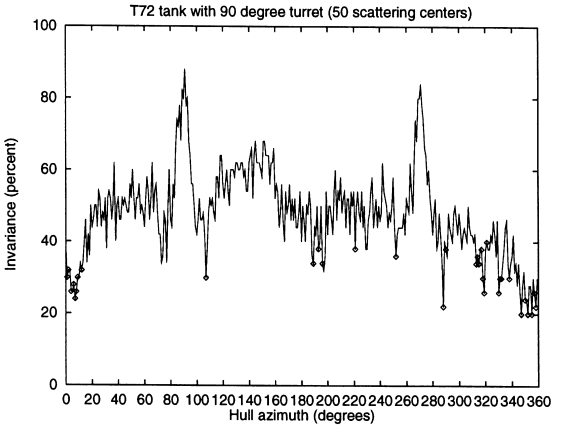


Fig. 18. XPATCH T72 tank (turret 90°) recognition failure plot (◇) on articulation invariance curve.

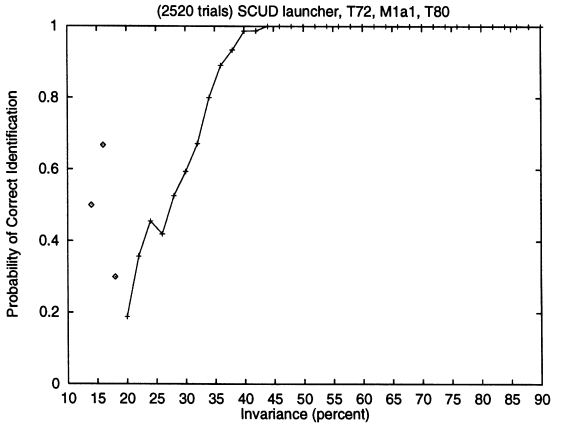
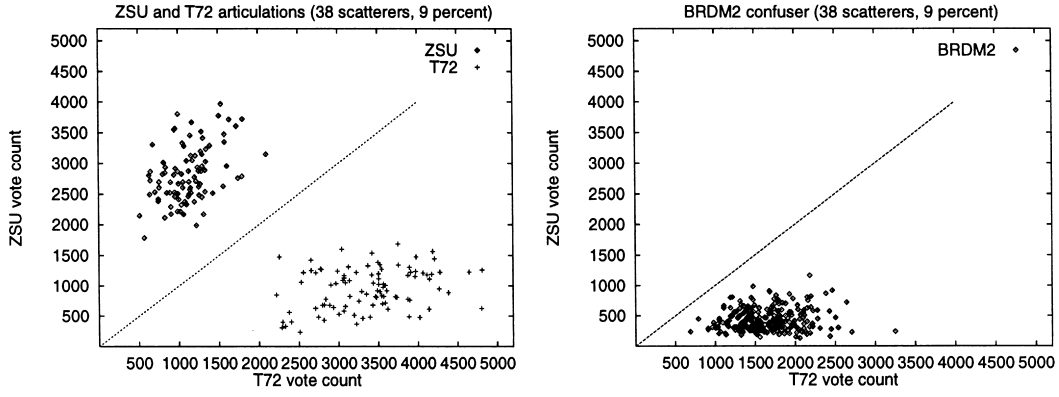


Fig. 19. Forced recognition rate and articulation invariance (50 scatterers, average of four XPATCH objects).



(a) ZSU vs. T72. (b) BRDM2 confuser.

Fig. 20. Scatter plots for 6D engine MSTAR results with articulation.

where the invariance is measured for an exact match of pixel locations. The points at low invariance values (plotted as 'diamonds in Fig. 19) are misleading. They are due to a few correct identifications for the M1a1 tank, where the invariance (measured with respect to the hull) is low, yet a correct identification is made from a number of features on the large turret, which are not accounted for in the hull invariance measure.

5.2. Results using MSTAR SAR data

In the MSTAR experiments the models are non-articulated versions of T72 #a64 and ZSU23/4 #d08 and the test data are the articulated versions of these same serial number objects and BRDM2 #e71 as an "unknown" confuser vehicle (all at 30° depression angle). (The confuser is not modeled and represents some unknown object that the recognition system should reject.) Results are optimum with the 6D recognition system using 38 scattering centers, a translation limit of ± 5 pixels and a percent magnitude change of less than $\pm 9\%$. The overall forced recognition rate (without a confuser, using only a decision rule that the winning object has the most votes) is 100% over a range from 14 to 40 scattering centers. Figs. 20(a) and (b) show 6D recognition scatter plot results in ZSU-T72 vote space for articulation of the ZSU 23/4 gun and the T72 tank and for the BRDM2 confuser. The results for the ZSU 23/4 and T72 are widely separated and away from the equal votes line, giving a 100% forced recognition rate. Fig. 20(b) shows that while the BRDM2 is always classified as a T72, a unique threshold in the range of 2000–2500 T72 votes will eliminate most of the false alarms at the cost of only a few T72s moved to the "unknown" classification. A common threshold applied to votes for either the T72 or the ZSU has a higher cost because some ZSUs would be moved to "unknown". This can be seen in Table 6 for a common threshold of 2100 votes, where the T72s are all correctly classified and only 2 ZSUs are classified as unknown, for an overall PCI of 0.990 and a probability of false alarm (PFA) of 0.126.

Table 6
MSTAR articulated object confusion matrix

MSTAR (Public) articulated test objects		Identification results		
		T72	ZSU	Unknown
T72	315° turret	98	0	0
ZSU	315° turret	0	92	2
BRDM2		32	0	222

A form of receiver operating characteristic (ROC) curve, with PCI vs. PFA, can be generated from the scatter plot data in Fig. 20 by varying the vote threshold (e.g. from 1500 to 4000 in 50 vote increments). The ROC curve in Fig. 21 shows the excellent recognition results that are obtained with the 6D recognition system for the MSTAR articulated objects. Fig. 22 shows how the PCI varies with the percent articulation invariance (for a within one-pixel location match) for the 6D recognition engine. The sets of curves are shown with different vote thresholds from 1700 to 2700 to generate failures that illustrate the effect of location invariance on recognition rate. The behavior shown in Fig. 22 for the 6D system with MSTAR data and 'within one-pixel' location match is consistent with the 2D system with XPATCH data and 'exact' location match previously shown in Fig. 19.

6. Conclusions

The large azimuthal variations in the SAR signatures of objects can be successfully captured by using models at

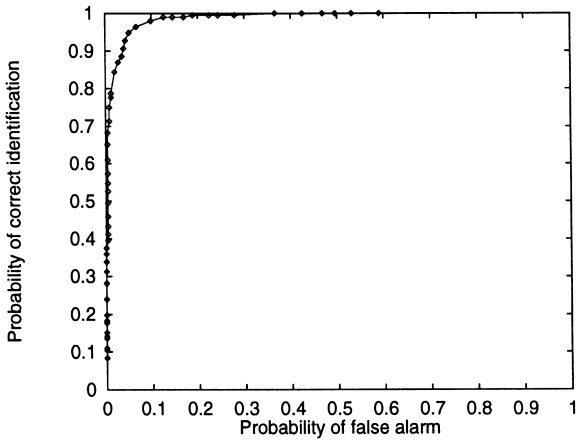


Fig. 21. Receiver Operating Characteristics for recognizing MSTAR articulation.

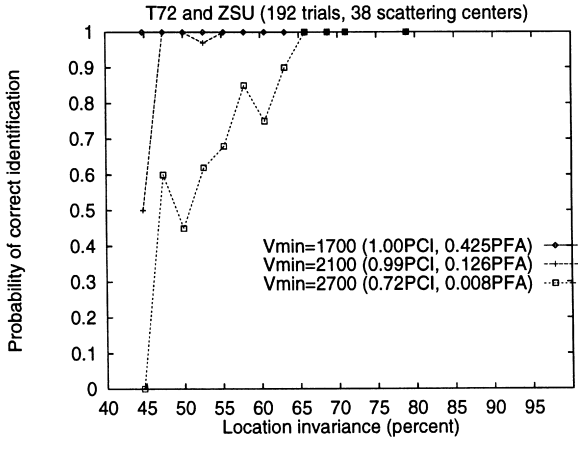


Fig. 22. MSTAR recognition rate and articulation invariance.

1° degree increments for a fixed depression angle. Useful articulation invariant features exist in SAR images of vehicles. For the XPATCH simulated SAR data, 48% of the scattering center locations exactly match after object articulation. For the real MSTAR SAR data, only 16.5% exactly match, however 56.5% of the locations match within a 3×3 pixel tolerance. In addition, where the scattering center locations are invariant with articulation, the corresponding magnitudes are also invariant within a small tolerance (typically less than a 10% change for the MSTAR data). The feasibility of a new concept for a system to recognize articulated objects in SAR images based on non-articulated models is demonstrated. A basic version of the system, using only relative locations of the scattering centers is sufficient for recognizing XPATCH generated objects that have a relatively high articulation invariance. The much greater variability of the real MSTAR SAR data can be successfully overcome by using the scatterer magnitude as well as location, by accommodating one-pixel uncertainty in the scatterer location and by considering an object as a group of scatterers at a consistent translation. These systems achieve excellent articulated object recognition results for simulated and real SAR data.

References

[1] D. Andersch, S. Lee, H. Ling, C. Yu, XPATCH: a high frequency electromagnetic scattering prediction code using shooting and bouncing ray, Proceedings of Ground Target Modeling and Validation Conference, August 1994, pp. 498–507.

[2] T. Ross, S. Worrell, V. Velten, J. Mossing, M. Bryant, Standard SAR ATR evaluation experiments using the MSTAR public release data set, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery V, Vol. 3370, April 1998, pp. 566–573.

[3] B. Bhanu, Automatic target recognition: state-of-the-art survey, IEEE Trans. Aerosp. Electron. Systems 22 (1986) 364–379.

[4] B. Bhanu, D. Dudgeon, E. Zelnio, A. Rosenfeld, D. Casasent, I. Reed, Introduction to the special issue on automatic targe detection and recognition, IEEE Trans. Image Process. 6 (1) (1997) 1–6.

[5] B. Bhanu, T. Jones, Image understanding research for automatic target recognition, Proceedings ARPA Image Understanding Workshop, January 1992, pp. 249–254.

[6] D. Dudgeon, R. Lacoss, An overview of automatic target recognition, Lincoln Lab. J. 6 (1) (1993) 3–9.

[7] L. Novak, G. Owirka, C. Netishen, Performance of a high-resolution polarimetric SAR automatic target recognition system, Lincoln Lab. J. 6 (1) (1993) 11–24.

[8] D. Kreithen, S. Halverson, G. Owirka, Discriminating targets from clutter, Lincoln Lab. J. 6 (1) (1993) 25–51.

[9] D. Casasent, R. Shenoy, Synthetic aperture radar detection and clutter rejection MINACE filters, Pattern Recognition 30 (1) (1997) 151–162.

[10] W. Irving, A. Willsky, L. Novak, A multiresolution approach to discriminating targets from clutter in SAR imagery, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery II, Vol. 2487, April 1995, pp. 272–299.

[11] J. Starch, R. Sharma, S. Shaw, A unified approach to feature extraction for model-based ATR, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery III, Vol. 2757, April 1996, pp. 294–305.

[12] A. Waxman, M. Seibert, A. Bernardon, D. Fay, Neural systems for automatic target learning and recognition, Lincoln Lab. J. 6 (1) (1993) 77–116.

[13] D. Carlson, B. Kumar, R Mitchell, M. Hoffelder, Optimal trade-off distance classifier correlation filters (OTDCCFs) for synthetic aperture radar automatic target recognition, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery IV, Vol. 3070, April 1997, pp. 110–120.

[14] J. Verly, R. Delanoy, C. Lazott, Principles and evaluation of an automatic target recognition system for synthetic aperture radar imagery based on the use of functional templates, SPIE Proceedings: Automatic Object Recognition III, Vol. 1960, April 1993, pp. 57–71.
[15] R. Meth, R. Chellappa, Automatic classification of targets in synthetic aperture radar imagery using topographic features, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery III, Vol. 2757, April 1996, pp. 186–193.
[16] T. Ryan, B. Egaas, SAR target indexing with hierarchical distance transforms, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery III, Vol. 2757, April 1996, pp. 243–252.
[17] D. Casasent, R. Shenoy, Feature space trajectory for distorted-object classification and pose estimation in SAR, Opt. Eng. 36 (1997) 2719–2728.
[18] J. Wissinger, R. Washburn, D. Morgan, C. Chong, N. Friedland, A. Nowicki, R. Fung, Search algorithms for model-based SAR ATR, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery III, Vol. 2757, April 1996, pp. 279–293.
[19] E. Keydel, S. Lee, Signature prediction for model-based automatic target recognition, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery III, Vol. 2757, April 1996, pp. 306–317.
[20] G. Ettinger, G. Klanderman, W. Wells, W. Grimson, A probabilistic optimization approach to SAR feature matching, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery III, Vol. 2757, April 1996, pp. 318–329.
[21] W. Jachimczyk, D. Cyganski, Enhancements of pose-tagged partial evidence fusion SAR ATR, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery IV, Vol. 3070, April 1997, pp. 334–345.
[22] J. H. Yi, B. Bhanu, M. Li, Target Indexing in SAR images using scattering centers and the Hausdorff distance, Pattern Recognition Lett. 17 (1996) 1191–1198.
[23] D. Casasent, S. Ashizawa, Synthetic aperture radar detection, recognition and clutter rejection with new minimum noise and correlation energy filters, Opt. Eng. 36 (1997) 2729–2736.
[24] I. Biederman, Recognition-by-components: a theory of human image understanding, Psychol. Rev. 94 (1987) 115–147.
[25] Y. Hel-Or, M. Werman, Recognition and localization of articulated objects, Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, November 1994, pp. 116–123.
[26] A. Beinglass, H. Wolfson, Articulated object cognition, or: How to generalize the generalized Hough transform, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 1991, pp. 461–466.
[27] G. Jones III, B. Bhanu, Recognition of articulated and occluded objects, IEEE Trans. Pattern Anal. Mach. Intell. 21 (1999) 7.
[28] Khoros Pro v2.2 User's Guide, Addison-Wesley, Longman Inc., New York, 1998.
[29] D. Dudgeon, R. Lacoss, C. Lazott, J. Verly, Use of persistent scatterers for model-based recognition, SPIE Proceedings: Algorithms for Synthetic Aperture Radar Imagery, Vol. 2230, April 1994, pp. 356–368.
[30] Y. Lamden, H. Wolfson, Geometric hashing: A general and efficient model-based recognition scheme, Proceedings of International Conference on Computer Vision, December 1988, pp. 238–249.
[31] W.E.L. Grimson, Object Recognition by Computer: The Role of Geometric Constraints, The MIT Press, Cambridge, MA, 1990.

About the Author—GRINNELL JONES III was a National Merit Scholar who received his BS degree in mechanical engineering from the Massachusetts Institute of Technology in 1966, MS in aerospace engineering (with Distinction) from the Air Force Institute of Technology in 1971 and MS in computer science from the University of California, Riverside in 1997. After a 25 year career in development engineering, missile operations and acquisition management with the U.S. Air Force, Lt. Col. (USAF Retired) Jones has been conducting research in automatic target recognition using Synthetic Aperture Radar imagery for the past five years with the University of California, Riverside. His research interests include object recognition, computer vision, machine learning, image and video databases, and systems applications.

About the Author—BIR BHANU received the S.M. and E.E. degrees in electrical engineering and Computer Science from the Massachusetts Institute of Technology, Cambridge, the Ph.D. degree in electrical engineering from the Image Processing Institute, University of Southern California, Los Angeles and the M.B.A. degree from the University of California at Irvine. Currently he is a Professor of electrical engineering and computer science and Director of Center for Research in Intelligent Systems at the University of California, Riverside. Previously, he was a Senior Honeywell Fellow at Honeywell Systems and Research Center, Minneapolis, MN. He has been the principal investigator of various programs for DARPA, NASA, NSF, AFOSR, ARO and other agencies and industries in the areas of learning and vision, image understanding, pattern recognition, target recognition, navigation, image databases, and machine vision applications. He is the co-author of books on "Computational Learning for Adaptive Computer Vision", (Plenum, Forthcoming), "Genetic Learning for Adaptive Image Segmentation" (Kluwer, 1994), and "Qualitative Motion Understanding" (Kluwer, 1992). He holds 10 U.S. and international patents and over 200 reviewed technical publications in the areas of his interest. Dr. Bhanu is a Fellow of the IEEE and the AAAS. He is a member of ACM, AAAI, Sigma Xi, Pattern Recognition Society, and SPIE.