Image Understanding Research at UC Riverside: Integrated Recognition, Learning and Image Databases

Bir Bhanu
College of Engineering
University of California, Riverside, California 92521
Email: bhanu@engr.ucr.edu
URL: http://constitution.ucr.edu

Abstract

This report summarizes the image understanding (IU) research being conducted at the University of California at Riverside (UCR) under the DARPA sponsored programs in learning, target recognition and image databases. The goal of our research is to develop robust, reliable and efficient algorithms and systems that can work effectively in real-world applications. The principal areas of investigation include physically-based approaches utilizing multiple representations for target detection and recognition using multisensor data, multistrategy learning-based approaches for IU, and image databases. Automatic target recognition, image exploitation, surveillance, dynamic multisensor databases are the principal applications areas of our research.

1 Introduction

The University of California at Riverside is conducting research in several different aspects of image understanding. We summarize the technical objectives and scientific issues of the new University Research Initiative effort just starting, and the important progress made in the areas of learning for image understanding and automatic target recognition using multisensor (SAR and FLIR) imagery during the period from November 1995 to March 1997.

2 Learning Integrated Visual Database for Image Exploitation

The DOD has critical needs for robust high performance automated systems that can recognize objects in reconnaissance imagery acquired under dynamically changing conditions and for systems that can efficiently extract information from enormous image databases. Our new research addresses two interrelated problems with the effectiveness and efficiency of automated/semi-automated techniques for image understanding. First, the lack of robustness in algorithms and systems for object recognition with changing environments. Second, the lack of scalable intelligent strategies for quickly extracting meaningful information from enormous, dynamically changing image databases. This project is distinguished from other image databases in the following areas: (a) The content-based image retrieval image database technology is used for designing reliable IU algorithms. (b) The system has learning capability, improving its performance with use, both in terms of processing speed and matching with the user's perception; (c) users can query the images as well as the processing algorithms; (d) an extensive amount of image-related information is stored for characterization of various features and algorithms.

We focus on the task of image exploitation. The operational goal is to monitor military forces (vehicles and equipment) in a small geographic area (10 Sq. miles) that move, sit and then move. This requires robust high performance IU systems for recognizing objects/events in multisensor imagery acquired under dynamically changing conditions, and efficiently extracting "information" from enormous dynamic databases and exploiting it to develop reliable IU systems that will adapt to changing environments. The results of this program will provide a significant tool for military and intelligence information systems that will directly contribute to meeting the DoD goal of dominant battlefield awareness.

Objectives: The overall scientific goal of this project is to demonstrate that the conjunction of learning, recognition, and content and context-based retrieval (CCBR) are necessary and sufficient for reliable IU. We believe that for the development of robust and reliable IU systems we need a new generation of IU research that integrates target recognition, learning and CCBR technologies. Each alone or any combination of two is not sufficient to develop reliable IU systems operating in dynamic real-world environments. We must combine them in an integrated system to develop the science for image recognition. The specific subgoals are:
(a) Techniques for adapting recognition algorithms and models to different theater of operations and target types.

(b) Algorithms for handling target configuration differences, articulations and occlusions without combinatorial explosion.

(c) Methods for database queries by example with multiple objects and relationships (semantic queries) for recognition of events or scenarios.

2.1 Adaptive Recognition Models for Different Environments

State-of-the-art image understanding (IU) algorithms and systems for image exploitation from SAR images generally use static algorithms. They possess no learning ability and cannot improve their performance with experience achieved over time. Since they possess no adaptive capability to adjust to varying sensor operational conditions (such as sensor differences, depression angles, and multiple polarizations) and deployment environments (such as desert, forest, agricultural, urban areas and their seasonal variations) they cannot migrate from one theater of operations to another. The objective of our research is an approach that applies adaptive learning algorithms to exploit context information and feedback on performance results to improve the performance of IU based force monitoring systems. We allow the image exploitation system to adapt itself to a variety of SAR clutter types and perform optimally under different operating conditions. The learning takes place to (a) adapt clutter models with changes in sensor operating conditions, (b) adapt classifiers for different clutter types, and (c) adapt parameters employed within feature groups based on target recognition results. The changes for different deployment environments such as forest, desert, jungle, arctic, etc.) are primarily reflected in the characteristics of the image background clutter. Thus, adapting IU system to varying clutters is of fundamental importance. The research contains the following innovative ideas:

Variety of Feature Groups to Build SAR Clutter Models: No single feature may capture all possible statistical/structural variations for different clutters involved in a SAR deployment environment. We use several groups of features based on (a) multiscale Gabor wavelet (b) self-similarity in natural scenes, (c) statistics of geometrical/structural elements, and (d) statistical features.

Learning Background Clutter Models Through a Supervised Self-Organizing Process: Instead of artificially assigning a distribution to clutter models, we build clutter models from examples through a supervised learning process. These clutter models are represented by compact self-organizing maps (SOMs) which capture the distribution of the training data without the need to store a large number of examples. The SOM technique is extended in our approach to an incremental supervised learning process for clutter characterization. We also use the self-organizing map to classify a given region of an image into a clutter or a target area. The classification algorithm is adapted for different clutter types.

Stochastic Reinforcement Learning Technique to Adapt Clutter Models to SAR Sensor Operating Conditions: Different Sensor operating conditions correspond to varying weights of different feature groups which together constitute a model for a particular clutter type. The relationship between operating conditions and the weight of feature groups is optimized through a stochastic reinforcement learning process. This learning paradigm is used here since the human supervisor (man-in-the-loop) will only be told the system that it is doing a "good job" rather than helping the computer in finding the association between operating conditions and the weights of different features.

Delayed Reinforcement Learning for Learning Clutter Model Parameters Based on Target Recognition Results: The image exploitation process requires a sequence of algorithms for CFAR (Constant False Alarm Rate) detection, feature extraction, clutter characterization and target recognition. It is inherently a multi-stage process that has delay from stage-to-stage. Since we cannot determine the goodness of different stages until we have seen the final recognition result, it is natural to evaluate the quality of earlier stages based on the final recognition results and a delayed reinforcement learning technique fits this situation exactly.

2.2 Algorithms for Handling Articulations and Occlusions

Current methods for target recognition in SAR imagery cannot handle target articulation, configuration differences or moderate occlusion. The objective of the research is to focus on the challenging problems caused by target variations due to articulation or configuration differences. Our approach to the problem of automatic model construction and recognition of articulated, non-standard targets in SAR imagery is based on local features and local reference coordinate systems. We have a systematic method for constructing recognition models of objects that are not articulated and then we employ local image features to match these models and recognize the same objects in articulated positions or non-standard configurations. The key features of the approach are:

Sensor Specific Design Approach for SAR Target Recognition: The unique characteristics and physics of SAR sensors are recognized and accommodated
by our design approach. The natural range/cross-range coordinates and tessellation are directly incorporated. The translation invariance is captured by using relative positions of SAR specific features and the large rotational variances are accommodated by modeling an appropriate number of azimuths.

**Models Based on Articulation Invariants:** Our approach for SAR target recognition makes use of the existence of articulation invariants. The models are stored for standard non-articulated objects. Thus, it avoids the combinatorial explosion of model configurations and is inherently directly applicable to matching the un-occluded regions (of occluded objects).

**Physically Based Local SAR Image Features Accommodate Articulation:** The relative distances between scattering centers (and other features such as topographic primal sketch features, reflector geometry, feature sequences based on location and relative amplitude, and polarization based features) are related to the shape and physical dimensions of the detailed target geometry. The local coordinate approach to local features (vs. global approaches or even a local neighborhood approach) accommodates articulation/occlusion without precluding use of widely separated features (which are good discriminators).

**Efficient Search for Positive Evidence is Designed to Accommodate Spurious Data:** A powerful combination of a true look-up table and a voting technique that searches for positive evidence reduces the work on all non-matching cases to the random coincidences and makes the method scale gracefully.

**Super-Resolution Target Chips:** Super-resolution (e.g. six inch) provides rich feature sets that allow matching the non-articulated or un-occluded regions of the target. Since it is not clear that the problem is solvable at one foot resolution, we have taken the unique approach of demonstrating feasibility at six inch resolution, and then investigating the performance degradation at one foot resolution real data.

**Hierarchical Approach to Indexing and Matching for Handling the Exceptional Cases:** The basic approach for indexing and matching based on the relative locations of HH-polarization signal strength maxima will be extended to other features (such as other polarizations and using the complex components) to handle the exceptional cases and additional matching modules, based on other features, will be applied to discriminate among ambiguous results. In addition, we explore a promising stochastic hidden Markov modeling (HMM) based approach for indexing/matching.

### 2.3 Database Semantic Queries for Recognition of Objects and Events

There are basically two approaches for searching image databases to identify objects. The first approach uses the traditional object recognition that requires the understanding of images. The second approach uses features for content-based retrieval to select images based on the chosen measure of similarity. It does not require the full understanding of images. We want to combine these two approaches for image exploitation application and investigate ways of using contextual data and domain knowledge for image interpretation. The key features of the approach are:

**Flexible Similarity Measures and Indexing Functions:** Current techniques for feature-based retrieval use a fixed set of features, similarity measures and indexing strategies that are determined in advance. We develop learning algorithms for feature selection, similarity metrics and associated indexing structures. We allow generation of run-time features and handle data and index management on the fly. We investigate techniques that permit efficient search of high dimensional space. This will allow improved performance in terms of retrieval speed and a quality of results approaching human perception of similarity.

In practice the relevance of each feature in classifying a new object may be different. In addition, the relevance of a feature may depend on the user and the object being classified. Inclusion of features with low relevance leads to high dimensionality of the feature vector and can degrade performance. What is required is to find the local relevance of each feature and use that information to define a flexible similarity measure that closely resembles human perception. Our approach for content-based retrieval is to learn the most salient features and develop flexible similarity measures that best resemble human perception of similarity for image exploitation.

Another important problem in large visual databases is the indexing structure to reduce the search space, allowing quick browsing. Since multiple features are normally necessary to represent an image, a multidimensional indexing structure is required. The performance of existing techniques for query by example critically depends on the selection of the features, the similarity measures, the user and the application context. In our approach the database is indexed by the order of the most significant factor/eigenfeature, the second most significant factor/eigenfeature, and so on. The query search in our approach consists of two stages: the pre-query stage and on-line stage. The learning is at two-levels: first to determine the local relevance of each feature, the ranking and selection of the features, and the indexing structure for the current user, query and application, and sec-
ond to select from the knowledge base the ranking of features and the indexing structure using the contextual information related with the application.

Data Models and Queries: We define a complete set of data models that the image database system is designed to handle. This includes data models for contextual information and the design of data structures for indexing and retrieval. There exist fairly complete data models for image formats and intermediate data types, which can be used as a basis of our development. The key issue is to devise an integrated model of images, image-related information, and processing algorithms.

We develop database access methods based upon content and context with common-sense and temporal reasoning capability, develop suitable query constructs and the semantics of linguistic constraints that allow one to express image-oriented queries, and designing an image data model that is sufficiently powerful, flexible, and extensible. Query methods can learn the selection of features and similarity measures that match with the user's perception, and the associated indexing structure. The learning approach will lead to improvement of performance with the use, both in terms of retrieval speed and user's perception of similarity.

Query language will perform associative search on images, features and algorithms. It will be sufficiently expressive and be capable of handling imprecise and incomplete data. What image resolution to use for query processing is an important optimization problem. We will investigate whether low-resolution intermediate results can be used to reduce the processing cost of image queries. Incomplete information often results in imperfect database schema, which need to evolve though learning, monitoring, or user overwriting. The schema may change at the data representation level (e.g. the attributes of an object, the class an object belongs to, etc.). It may also change at the conceptual level (e.g., the change of the class hierarchy). Both types of evolution will be studied.

Bayesian-Based Factor Analysis: Principal component analysis is a commonly used technique in image processing and has been recently used in visual databases. However, there are several limitations of this technique and the factor analysis model [28] has several important advantages: first, the factor analysis model permits a noise term, second, the factor analysis model postulates a linear model for the basic data vectors, and, finally, the factor analysis model is much more general, and is driven by a need to find and retain a meaningful correlation structure for the data that can be explained by a few linear combinations of some latent factors.

The method we develop and apply in this context involves scoring the image according to the Bayesian factor analysis model, which is ideally suited for image databases. It provides us a compact representation, contextual information for image exploitation can be explicitly accounted for in the model, and it is suitable for indexing, image recognition and classification.

Image Characterization: A variety of features are used in content-based retrieval for visible images. Many of these features are not useful for SAR, FLIR and multispectral images which are important for image exploitation. We develop image content based on wavelet (e.g. Gabor wavelet based representation has energy patterns that are localized both in the spatial domain and in the frequency domain) and information complexity measures (such as minimum description length) to characterize multisensor images.

2.4 Prototype System

Our new research will be build upon the Visual Intelligence Datablade system being developed by Virage Inc. This system [23] is based on a basic model called Visual Information Management System (VIMSYS) developed by Virage Inc. This model has four layers of information abstraction: the raw image, the processed image, the user features of interest and the user events of interest. The top three layers form the content of the image. There are mechanisms for defining and installing new similarity measures, called primitives. In addition, Virage has tools for graphical user interface, query canvas (query-by-sketch), light table (for displaying query result), and command line interface.

As part of the project we will develop algorithms and tools for image exploitation in the context of large databases. In addition, we will develop a research testbed to integrate image and context databases with both human customers and the target detection and recognition system algorithm customers.

2.5 Evaluation Plan

Our evaluation plan provides a significant emphasis on algorithm evaluation and will allow the subsystem technology developed to be evaluated in the context of overall system effectiveness.

The overall system performance metrics are a probability of detection (Pd) and a false alarm rate (FAR). Demonstration results for the clutter modeling will be expressed in terms of Pd and FAR, later results for recognition will be in terms of Prc (probability of correct classification) and Pci (probability of correct identification). In addition, the performance of the learning system will be reported as a learning rate expressed in terms of performance versus the number of exemplars experienced. We plan to
use available SAR, visible and multispectral imagery and the imagery that may become available during the program and simulated SAR scenes produced by XPATCH at various depression angles and for different environments (e.g. forest, agriculture, desert shrub and desert) to populate the database.

The critical experiments are (a) demonstration of the capabilities of various feature groups and self organizing clutter models to distinguish man-made objects from natural clutter in actual SAR images and to show test results for scenes in simulated imagery. (b) the use of reinforcement learning to adapt the natural clutter models to sensor operating conditions. (c) learning rate (performance vs. experience) for retraining a clutter model with data from a different depression angle. (d) demonstrate the performance of clutter models that are adapted to new deployment environments (for example agriculture, and desert shrub) and report the learning rate results. (e) demonstrate the system level performance of the recognition elements integrated with the clutter models. (f) demonstrate the system level performance with the clutter models adapted to the matching results and also report both the learning rate and the point where learning transitions from supervised to unsupervised.

3 Multistrategy Learning for Image Understanding

The multistrategy learning-based IU approach selectively applies machine learning techniques at multiple levels of the IU process to achieve robust recognition performance. At each level, appropriate evaluation criteria are employed to monitor the performance and self-improvement of the system [5, 18].

With the goal of achieving robustness, our research at UCR is directed towards learning parameters, feedback, contexts, features, concepts, and strategies of IU algorithms for model-based object recognition. The progress made during the last year includes the following: (a) development of approaches based on reinforcement learning for controlling feedback between segmentation and recognition components in an object recognition system, and using it to learn segmentation and feature extraction parameters. (b) development of an approach based on reinforcement learning for integrating context with clutter models to reduce false alarms and improve target detection performance in FLIR images (c) development of a methodology to improve performance of an IU algorithm by adapting the input data into the desired form for a given algorithm. (d) development of a case-based reasoning approach for learning recognition strategies for image exploitation by categorization of images.

Earlier we have demonstrated the scalability of the genetic learning-based approach for adaptive image segmentation [12, 17]. We also developed basic ideas applicable to integrating information from multisensors or integrating recognition and motion analysis, using multiobjective optimization [2, 9].

3.1 Learning Recognition Strategies

We have developed several techniques for learning recognition strategies. These techniques are based on reinforcement learning and case-based reasoning.

3.1.1 Reinforcement Learning for Adaptive Algorithms, Parameters and Feedback in an IU System

**Problem:** To automate acquisition of recognition strategies in dynamic environments to develop theoretically sound approaches to control feedback which are based on the results of recognition and to learn segmentation and feature extraction parameters for robust model-based recognition.

**Approach:** We have developed two approaches based on reinforcement learning for closed-loop object recognition in a multi-level vision system. These approaches use the team of learning automata algorithm [26] and the delayed reinforcement learning algorithm [27].

The closed-loop object recognition system evaluates the performance of segmentation and feature extraction by using the recognition algorithm as part of the evaluation function. Recognition confidence is used as a reinforcement signal to the image segmentation or feature extraction processes. By using the recognition algorithm as part of the evaluation function, the system is able to develop recognition strategies automatically, and to recognize objects accurately in newly acquired images. As compared to the genetic algorithm [9, 10] which simply searches a set of parameters that optimize a prespecified evaluation function, here we have a recognition algorithm as part of the evaluation function [26].

In order to speed up the above algorithms we have developed a general approach [3] to image segmentation and object recognition that can adapt the image segmentation algorithm parameters to the changing environmental conditions. The edge-border coincidence is used for both local and global segmentation evaluation. However, since this measure is not reliable (see Figures 1 and 2) for object recognition, it is used in conjunction with model matching in a closed-loop object recognition system. Segmentation parameters are learned using a reinforcement learning algorithm that is based on a team of learning automata and uses edge-border coincidence or the results of model matching as reinforcement signals. The edge-border coincidence is used initially to
select image segmentation parameters using the reinforcement learning algorithm. Subsequently, feature extraction and model matching are carried out for each connected component which passes through the size filter based on the expected size of objects of interest in the image. The control switches between learning integrated global and local segmentation based on the quality of segmentation and model matching.

Accomplishments: Using the Phoenix algorithm for the segmentation of color images, a clustering-based algorithm for the recognition of occluded 2-D objects [11] and a team of learning automata [26] algorithm, or a delayed reinforcement learning algorithm [27], we show that in real images with varying environmental conditions and camera motion, effective low-level image analysis and feature extraction can be performed. We have shown performance improvement of an IU system combined with learning over an IU system with no learning [26, 27]. Figure 3 gives an example for performance improvement for both image segmentation and object recognition with experience. In this figure the traffic sign shown in the first column of images (taken at different times) is to be recognized. The second column shows the segmented results when the learning process is stopped and the traffic sign has been recognized. Figure 4 demonstrates the learning behavior - a reduction in CPU time to recognize the traffic sign in one run of 12 images. Figure 5 shows the improvement in speed between the two schemes - scheme1 [26] and scheme2 [3]. Scheme 2 makes use of edge-border coincidence and global/local image segmentation to speed up the recognition process. Both schemes use the same learning algorithm.

Future Work: (a) Develop a complete reinforcement learning-based system for 3-D model-based ob-
ject recognition with feedback among various levels. (b) Evaluate the performance of the technique for ATR application, (c) Learn algorithm parameters, develop algorithms and evaluation criteria for multisensor image segmentation and recognition, (d) Learn the optimal sensor combinations and cross-sensor validation of segmentation results.

3.1.2 Case-based reasoning for adaptive IU System

Problem: To automate acquisition of IU strategies, to integrate context with image properties, recognition algorithms and their parameters.

Approach: Most current model-based approaches to object recognition utilize geometric descriptions of object models, i.e., they emphasize the recognition problem as a characteristic of individual object models only. Various other factors, however, may influence the outcome of recognition in a real application such as photointerpretation. These factors include contextual information, sensor type, target type, scene models, and other non-image information. Using Case-Based Reasoning (CBR), successful recognition strategies (contextual information, algorithms, features, parameters, etc.) are stored in memory as cases and are used to solve new problems.

Since there are no algorithms that show acceptable performance over all different image sets that can be input to a system, we categorize images into classes and find the best algorithm for each class. When a new image is provided to recognize an object such as a particular aircraft type, the new image is first
Figure 5: Comparison of accumulated CPU time for 5 different runs on 12 images.

categorized into the most similar class and then processed using the best algorithm known beforehand.

Categorization of images is, however, a very difficult problem. Instead of categorizing an image, a region of interest (ROI) is classified. For training images, ROIs are acquired and divided into classes by a human operator. The best algorithm is also selected by a human operator during training. Once images are categorized, characteristics of image sets are compiled statistically. These compiled probability distributions of values for each characteristic feature are utilized to find the most similar class. Characteristic features fall into two categories: contextual information and pure image metric information. Weather, time of image acquisition, and viewing angles are used as contextual information. Homogeneity factor, convexity factor, and agglomeration factor are suggested as pure image metrics information.

Accomplishments: We have developed the basic elements of the CBR paradigm. We have experimented extensively with a C-based algorithms for aircraft recognition in aerial photographs [19, 20, 21]. We have written code for characterizing image data sets.

Future Work: (a) Develop a prototype system which will have all the basic elements of CBR. (b) Select the best image metrics based on the discriminating power for categorizing images. (c) Develop reasoning, adaptation and indexing approaches that will make CBR an effective approach for IU applications [25].

3.2 Learning to Integrate Context with Clutter Models

Problem: To integrate contextual information with clutter models for target detection and recognition. Current image metrics commonly used to characterize images do not correlate well with the performance of target recognition systems.

Approach: The contextual parameters, which describe the environmental conditions for each training example, are used in a reinforcement learning paradigm to improve the clutter models and enhance target detection performance under multi-scenario situations [29]. New Gabor transform-based features and other statistical image features are used to capture the statistical properties of natural backgrounds in visible and FLIR images. The non-incremental self-organizing map approach commonly used in an unsupervised mode is extended, by the addition of a near-miss injection algorithm, and used as an incremental supervised learning process for clutter characterization [30].

Accomplishments: A fast algorithm to compute the Gabor transform of a given image has been implemented. We have implemented two new Gabor transform-based feature groups and tested their classification performance on natural backgrounds. Experimental results show that the two feature groups could capture certain characteristics of the backgrounds, which are consistent with our theoretical expectations based on the physical meaning of each attribute within the feature group.

Using 40 second generation FLIR images and four contextual parameters (time of the day, depression angle, range to the target and air temperature) and 5 feature groups, we find 100% detection rate, 10% false alarm rate and significant improvement in the confidence for classifying a feature cell (rectangular regions in an image) as a clutter or a target. The results have been compared with and without contextual information [30].

Future Work: (a) Prove the convergence of the stochastic reinforcement learning algorithm for multi-feature cases. (b) Test the approach on a larger data set with a variety of contextual parameters. (c) Find the most influential environmental parameters for a given sensor, find how a feature group is affected by a given environmental parameter and find if we can make a feature invariant with respect to a given environmental parameter through normalization of the sensor data.

3.3 Learning for Input Adaptation and Feature Extraction

Problem: To improve the performance of an IU algorithm by adapting its input data to the desired form so that it is optimal for the given algorithm.

Approach: Two general methodologies for the performance improvement of an IU system are based on optimization of algorithm parameters and adaptation of the input. Unlike the genetic learning case for adaptive image segmentation, here we focus on the second methodology and use modified Hebbian learning rules to build adaptive feature extractors which transform the input data into the desired form for a given algorithm [35, 34]. Learning rules are based on different loss functions and are suitable for extracting expressive or discriminating features from the input.
Accomplishments: The feasibility of the approach is shown by designing an input adaptor for a thresholding algorithm for target detection in SAR, FLIR and color images. The results are excellent with input adaptor compared to the case with no input adaptor.

Future Work: (a) Develop transformations from input data to salient features needed for various classes of algorithms. (b) Compare performance with/without input adaptor for algorithms used in applications such as automatic target recognition.

4 Automatic Target Detection and Recognition

The goals of our ATR research are to use sensor and geometric models and multiple representations (called physically-based modeling) for developing techniques for the recognition targets in multisensor imagery [6] and generic object recognition in complex aerial images. Our initial approach for indexing/matching in SAR images was based on using scattering centers and the Hausdorff distance measure [7, 32]. Since then we have focused on recognition of articulated and occluded objects and this approach is not suitable for it. We have made progress in the areas of recognition of articulated and occluded targets in SAR images using invariants and stochastic models [1, 16, 24]. We have also developed a Bayesian approach for the segmentation of SAR images and an approach for automatic model construction from inverse synthetic aperture radar images. Earlier we have developed and tested approaches based on Gabor wavelet representation [8] for (a) distortion-tolerant flexible matching for the recognition of occluded and nonoccluded targets in FLIR images, (b) computing salient structures in cluttered images, and (c) approaches for target detection in complex multimodal FLIR images.

4.1 Recognition of Targets in SAR Images

Problem: Develop techniques for indexing and matching to recognize articulated/occluded targets in SAR images.

Approach: Our invariants based approach is based on relative distances among the scattering centers to access a look-up table that generates the votes for the appropriate target and the azimuth. Using these results we can identify features which are on the turret and which are on the hull and can identify target, its body pose and the turret pose [1]. The power of the techniques is derived form the fact that it makes use of “azimuthal variance”, both local and global constraints, high resolution data, “articulation invariants” and a voting mechanism as positive evidence for an efficient search [24].

Accomplishments: Figure 6 shows four sample targets which are used in our experiments. Using XPATCH generated data at 6in. resolution (10.0 GHz center frequency, 1.0 GHz bandwidth, 5.6° angular span), we have found that significant number of features do not typically persist over a few degrees of rotation. Averaging the results for 360 azimuths of the T72 tank, only about one-third of the 50 strongest scattering center locations remain unchanged for 1° azimuth (see Figure 7) and less than 5% persist for 10°. Figure 8 shows the articulation invariants. It shows the percentage of the strongest 50 scattering centers for the T72 tank that are in exactly the same location with the turret rotated 60° as they are with the turret straight forward. Figure 9 shows how the probability of correct identification varies with the percent invariance. Note that the recognition performance is excellent for invariance values greater than 40% (i.e., down to 60% spurious data). Recognition rate for varying amounts of occlusion (288,00 test cases) is shown in Figure 10. Note that it is consistent with the previous figure. Figure 11 compares (51,840 tests) the performance results of the articulated and occluded articulated targets for cases with the same number of valid scatterers. It shows the importance of relatively long distances and shows that object recognition approaches that combine both local and global constraints will be better than those which rely on local constraints only.

Future Work: (a) Test the approaches using real SAR data and quantify the performance, (b) Develop techniques for feature selection, (c) Develop matching techniques that account for complex feature types and 3D geometry [7], (d) optimize recognition performance with respect to feature extraction and feature types, (e) Develop a model for per-
Figure 7: T72 azimuthal invariance.

Figure 8: An example of Articulation invariants.

Figure 9: Recognition rate and articulation invariance (50 scatterers, average of 4 objects).

Figure 10: Recognition rate and occlusion percent.

Figure 11: Articulated object and occluded articulated object performance results.
Figure 12: Example of an observation sequence superimposed on an image of T72 tank.

Performance prediction.

4.2 Hidden Markov Modeling (HMM)-Based Approach for Indexing/Matching

Problem: Develop stochastic model-based techniques for indexing and matching to recognize articulated targets in SAR images.

Approach: The targets with pattern distortion caused by articulation and occlusion cannot be recognized by template matching. An alternative is to use a statistical method that can handle the possible configuration variations of the same object. Because of its stochastic nature HMM is suitable for describing patterns of variation. The key elements of HMM include: finding the probability of the observations given the model, finding the most likely state trajectory given the model and observation, and adjusting the parameters of HMM to model the observation sequence better. The basic idea is imagining the features as emitting patterns of some hidden statistical model. We can sort available features to get appropriate sequences as the observation sequences of HMM. These sequences will represent the particular pattern of point features. It is reasonable to use the relative locations and relative magnitudes of these point features to obtain observation sequences (see Figure 12). Sequence based on relative amplitudes of SAR image is O1 = Mag 1, Mag 2, Mag 3, ..., Mag n. Selected sequences based on geometrical relationship are: O2 = d(1,2), d(2,3), d(3,4), ..., d(n-1,n), d(n,1), O3 = d(1,2), d(1,3), ..., d(1,n), O4 = d(2,1), d(2,3), ..., d(2,n), O5 = d(3,1), d(3,2), ..., d(3,n).

Accomplishments: We have used HMM for recognition of occluded objects in XPATCH generated data as described above. Examples of occlusion in training and test cases are shown in Figure 13. During training we find the optimal number of symbols (4) and states (5) for HMM. Using 325,000 training samples (5-10% occlusion) and 81,000 testing samples (5-50% occlusion) for 5 classes we find the results as shown in Figure 14. The results obtained from individual models are combined by an algorithm to achieve the results shown in this figure. The dotted lines show the worst and the best performance that was achieved with 5 HMM models (O1 to O5).

Future Work: (a) Test the approach for articulation, and occlusion with articulation, (b) Test the approach on real SAR data, (c) Develop techniques to find the optimal number of HMM models for various targets, (d) Develop methods to integrate results from different HMM models.

4.3 Other SAR related work

We have developed a Bayesian approach [22] using dynamically selected neighborhoods for the segmentation of SAR images. The approach allows a variety of prior information to be explicitly included in the image segmentation task. Preliminary results have been shown on simulated data.

We have also constructed the models automatically for object recognition using ISAR images. Given a set of ISAR data of an object of interest, structural
features are extracted from the images. Statistical analysis and geometrical reasoning are then used to analyze the features to find spatial and statistical invariance so that a structural model of the object suitable for object recognition can be constructed. A novel feature of the approach is that it uses the persistency of scattering centers computed during training phase to extract good scattering centers during testing phase. Four objects (Camaro, Dodge Van, Dodge Pickup and Bulldozer) are used to demonstrate the results. There are 351 images in both the training and the testing data for each object. The testing data are offset by 0.2 degree in azimuth from the corresponding training data. The results are significantly better compared to the case when we do not use the persistency of scattering centers during on-line phase. The results show that this approach is promising for automatic model construction [33].

4.4 Gabor Wavelets for Target Recognition and Target Detection

Using Gabor wavelets representation we have developed model-based, distortion-tolerant flexible matching techniques for recognition of occluded and nonoccluded targets under varying environmental conditions [31]. The key idea is to use magnitude, phase and frequency measures of Gabor wavelet representation in an innovative flexible matching approach that can provide robust recognition. The Gabor grid, a topology-preserving map, efficiently encodes both signal energy and structural information of an object in a sparse multi-resolution representation. Flexible matching between the model and the image minimizes a cost function based on local similarity and geometric distortion of the Gabor grid. Grid erosion and repairing is performed whenever a collapsed grid, due to object occlusion, is detected. We have performed a variety of experiments with second generation FLIR data and synthetic targets exhibiting varying signatures with changing environmental conditions. The results are reported in [31].

We have developed a new feature ("composite phase") based on Gabor wavelets. Also we have developed techniques for the computation of salient structures and target detection using wavelets [2].

5 Other Research

Other areas of ongoing work include navigation and obstacle detection [4, 13, 15]. We are developing a mobile testbed, called UC Rover for experiments in perception and learning. We have developed model-based generic object recognition approaches for qualitative recognition of aircraft in perspective aerial imagery and tested them on complex aerial images [19, 20, 21]. We have also done research on terrain interpretation using multispectral images [14].

6 Conclusions

We have developed promising approaches and obtained good results to solve some of the fundamental problems in IU that will have strong impact in solving real-world applications. In the coming years our focus will be the development of new algorithms and the end-to-end complete system that integrates recognition, learning and image databases for image exploitation using SAR, visible and multispectral imagery. We shall emphasize the performance evaluation of our algorithms and systems to measure improvements over current approaches.

References


