


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

Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, California 92521
Email: bhanu@cris.ucr.edu
URL: http://www.cris.ucr.edu

Abstract

This report summarizes the image understanding (IU) research being conducted at the University of California at Riverside (UCR) under the DoD sponsored programs in target recognition, learning 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 target detection and recognition, multi-strategy learning-based approaches for IU, and image databases. Automatic target recognition, image exploitation, surveillance, dynamic multisensor databases and interactive systems are the principal applications areas of our research.

1 Introduction

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 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 and extended operating conditions.

Second, the lack of scalable intelligent strategies for quickly extracting meaningful information from enormous, dynamically changing image databases.

Our research is aimed at developing IU algorithms and systems that have performance prediction and learning capabilities and that can improve their performance with experience, in terms of quality of results, processing speed and matching with the user's perception.

The overall scientific goal of our research is to demonstrate that the conjunction of recognition, learning and context and content-based retrieval (CCBR) is 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 explored during the reporting period are:

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

(b) Performance prediction of recognition algorithms and fundamental lower and upper performance bounds.

(c) Techniques for adapting recognition algorithms and models to different operating conditions, and target types.

(d) Methods for image database queries by example where the computer learns from user interaction for online indexing.

(e) Content-based retrieval oriented approach for object recognition.

In the following we describe the major accomplishments achieved since the last IU Workshop held in May 1997. Specific aspects of the research are given in greater detail in separate papers in these proceed-
2 Invariants for the Recognition of Articulated and Non-standard Targets [17, 3, 4, 5, 2, 11]

Using SAR scattering center locations and magnitudes as quasi-invariant features we have developed new techniques for recognizing articulated, nonstandard and occluded targets in real SAR images from the MSTAR public data. These recognition results are a substantial improvement over the performance of the earlier recognition approach with real SAR data.

We find that a significant percentage (56.5 - 61.4%) of the SAR scattering center locations are quasi-invariant (within a 3x3 pixel tolerance) for object articulation (e.g., turret rotation for the T72 tank and ZSU 23/4 gun), configuration differences (e.g., fuel barrels, searchlights, wire cables, etc.) and small depression angle changes. The magnitudes of these quasi-invariant scatterers (expressed as a radar cross section) typically change by less than ±10%. As an example, SAR images and the regions of interest (ROI), with the locations of the scattering centers superimposed, are shown in Figure 1 for baseline and articulated versions of the T72. Figure 2 shows the location invariance of the strongest 40 scattering centers with articulation for MSTAR T72 #a64 (at a 30° depression angle) as a function of the hull azimuth. The average invariance is 17.2% for an exact match of scattering centers and 57.8% for a location match within a one pixel (3x3 neighborhood) tolerance. Figure 3 shows the probability mass function (PMF) for percent amplitude change for the strongest 40 scattering centers of T72 #e162 vs. #e12 (at a 15° depression angle).

The positions and magnitudes of pairs of these quasi-invariant scatterers are used in a 6D recognition engine to achieve good recognition results with real SAR data for object articulation, configuration differences and small depression angle changes. As an example Figure 4 shows the ROC curve for depression angle changes. The model includes T72 (models 132 and BMP2 (e231) at a 17° depression angle and the test data consists of the same serial numbers at a 15° depression angle with BTRC70 (e271) as an "unknown" confuser. The recognition rate is related to the percent of location and magnitude invariant scattering centers [17].

While these three problems (articulation, configuration differences and change in depression angle) are similar, the differences among configurations of an object type are a more significant challenge for recognition than articulation and depression angle changes, where the model and test data are the same physical object under different conditions.

In other related work [2], we have developed separate body and turret models for tank targets that are independent of the relative positions between the body and the turret. These models are used in a subsequent matching technique to refine the pose of the body and determine the pose of the turret. The thresholds for the quality of match are dynamically determined by minimizing the probability of a random match for the recognition system. The system performance is evaluated with respect to the number of hypothesis, classification performance, accuracy of estimates for body and turret poses and computation time.

The future work will incorporate additional features in the recognition engine which should lead to further performance improvements and accommodate combined cases such as configuration variants along with depression angle changes.

3 Recognition Performance Prediction and Fundamental Performance Bounds [12, 13, 14]

We have developed novel techniques for predicting the performance of model-based object recognition systems in the presence of data uncertainty, occlusion and clutter. These techniques determine fundamental performance bounds (lower and upper) and set the limits on what is possible for a feature-based object recognition system. The proposed techniques capture the structural similarity between model objects, which is a fundamental factor in determining the recognition performance.

We predict bounds on the probability of correct recognition (PCR) by considering:

- 1. Scene-Data Factors: data uncertainty (due to measurement error), occlusion (missing scene-object features), and clutter (spurious scene features), and
- 2. Model Factors: the structural similarity between model objects. Intuitively, the probability of failing to recognize an object, in a distorted scene, is directly proportional to the degree of similarity between this object and the rest of the model objects.

We assume that model objects and some data are represented by 2-D point features, where each feature is represented by its positional information. Further, we assume that the decision criterion is vote-based; i.e., the object/pole hypothesis with the maximum number of consistent features is selected as the valid one.

Given a bound on data uncertainty, we determine the structural similarity between every pair of model objects. This is done by computing the number of consistent features between the two objects as a function of the relative transformation between them. Similarity information is then used, along with statistical models for data distortion, to estimate the probability of correct recognition (PCR) as a function of occlusion and clutter rates.

We have validated theory by comparing predicted PCR plots with ones that are obtained experimentally using MSTAR data. Each model target is represented by a number of SAR views which sample the signature at a specific depression angle (ψ), and a variety of azimuth angles. The model database consists of views corresponding to three targets: T72 (231 views), BMP2 (233 views) and BTR (333 views). 

Figure 1: MSTAR SAR images and ROIs (with peaks shown as +) for T72 tank #a64 at 56° azimuth.

Figure 2: Articulated T72 scatterer location invariance.

Figure 3: Example T72 scatterer percent amplitude change with configuration.

Figure 4: Receiver Operating Characteristics for recognizing MSTAR depression angle changes.
views at $\theta_t = 17^\circ$. Each of the views is treated as an independent object for recognition purposes. In our case, the space of applicable transformations is 2-D translation in the image plane \[3\]. Scattering centers, peaks in the image, are used as point features for recognition. These peaks are extracted by comparing the value of each pixel with its eight neighbors. We have chosen the strongest 20 scattering centers to represent both model and some data.

Data distortion includes:

- (a) Occlusion: a number of features are selected randomly and deleted,
- (b) Clutter: the same number of features are randomly generated within a rectangular area whose center is the bounding box of the target and whose area is nine times the target size. The locations of clutter features are restricted such that none is consistent with the occluded features of the object under consideration.
- (c) Uncertainty: each unoccluded feature is randomly perturbed within its 4-neighbor region, subject to the restriction that the perturbed feature is not consistent with any occluded feature.

The above restrictions are imposed to control the number of votes for the object under consideration. Since we are considering a fixed number of scattering centers, the occlusion and clutter rates in an image are always the same.

Recognition is performed based on geometric hashing. The recognition algorithm examines almost all of the problem space (target, azimuth, translation $x$, translation $y$), and so its performance is almost optimal.

Figure 5 and Figure 6 show experimentally determined PCR plots along with PCR bounds which are predicted using our method, in the cases of no data uncertainty (Figure 5) and 4-neighbor uncertainty regions (Figure 6). Comparing actual PCR plots with predicted bounds, we observe that our method successfully determines tight bounds on PCR.

The future work will include predicting the performance using MSTAR data with target articulation and configuration differences, and developing appropriate occlusion and clutter models.

4 Adaptive Target Recognition Using Reinforcement Learning

[6, 19, 20, 7, 22]

This research focuses on the following problem: Given a particular point in the ROC performance curve at which one desires to operate the ATR system, how can we tune the parameters of the system so that the system operates as close as possible to the specified point on the ROC curve? We have applied a reinforcement learning-based approach to an end-to-end SAR target recognition system where the parameters to be learned are the number of features and the decision rule (ratio of votes for the potential winner to votes for the next best object).

Fundamentally, target recognition is a multi-level process requiring a sequence of algorithms at low, intermediate, and high levels. Generally, such systems are open loop with no feedback between levels and ensuring their robustness is a key challenge in computer vision and pattern recognition research.

In our approach the parameters of a multi-level system employed for model-based object recognition are learned. The method improves recognition results over time by using the output at the highest level as feedback for the learning system. In our current work the focus has been adapting the parameters at the high level. In our earlier work we have shown the experimental validation of the method for learning the parameters of image segmentation and feature extraction. Our approach systematically controls feedback in a multi-level vision system and shows promise in approaching a long-standing problem in the field of computer vision and pattern recognition.

As an example, the user desires to direct the SAR engine to perform at or around the performance point (PCI, PF). The learning algorithm can affect the performance of SAR engines by choosing the parameters of the SAR engine.

The parameters of SAR engine now used by the learning algorithm are number of scattering centers (#SC) and vote ratio (VR). Once the input image and parameters (#SC, VR) are given, after the SAR engine runs over input images with the given parameters, we get the ROC curve (PCI vs. PF) of the SAR engine. Figure 7 shows the effect of #SC and VR on PCI, and Figure 8 shows the effect of #SC and VR on PF. The desired solutions are high PCI and low PF value. But in order to get high PCI, we should use low VR, in order to get low PF, we should use high VR. So, these two factors are conflict. Figure 9 shows the Q-value for each pair of parameters. The Q-Value measures the quality of the pair of parameters. The larger the Q-Value, the better its corresponding parameters. The goal of the learning algorithm is to find the good parameters as fast as possible. For Figure 9, the user specified 0.80 PCI and 0.10 PF. After the learning algorithm succeeds 10 times, the Q-Value of each pair of parameters is shown in Figure 9, where the best pair of parameters found by the learning algorithm is #SC of 35 and VR of 1.1. This pair of parameters is very close to the optimum pair of parameters (36, 1.1). The PCI and PF resulted from (35, 1.1) are 0.80 and 0.12 respectively, very close to the PCI and PF specified by the user.

The future work will extend the learning approach to more sophisticated recognition engines using more features and parameters where the alternative of an exhaustive search to optimize the parameter set becomes prohibitively expensive.
5 Learning Integrated Online Indexing for Retrieval in Image Databases [23, 8, 9]

Most of the current image retrieval systems use "one-shot" queries to a database to retrieve similar images. Typically a K-NN (nearest neighbor) kind of algorithm is used where weights measuring feature importance along input dimensions remain fixed (or manually tweaked by the user) in the computation of a given similarity metric. However, the similarity does not vary with equal strength or in the same proportion in all directions in the feature space emanating from the query image. Moreover, a similarity metric is not optimal for all kinds of images in a database. The manual adjustment of these weights is time-consuming and exhausting. Moreover, it requires a very sophisticated user.

We have developed a novel image retrieval system that continuously learns the weights of features and selects an appropriate similarity metric based on the user's feedback given as positive or negative image examples. The approach is highly adaptive to query locations. Experimental results are presented that provide the objective evidence of learning behavior of the system for image retrieval. Currently we have used Gabor wavelet based features for texture characterization in images.

Figure 10 shows a particular retrieval result obtained by our technique on the MIT image database [23] with no learning, that is, each dimension is weighted equally in the distance computation, where a retrieval precision of 25% is achieved. In contrast, Figure 11 shows the retrieval results after learning has taken place, where the results in Figure 10 provide relevance feedback. In this case, a retrieval precision of 95% is achieved. This illustrates that capturing local feature relevance indeed helps improve retrieval procedures.

In the future we will apply our approach on multisensor data (color video, multispectral, SAR, IR).

6 Database Retrieval Oriented Approach for Target Recognition [24]

Recognition of objects when the number of model objects becomes large has been a challenging problem. This requires a systematic approach that can organize the model data using a suitable representation, thus motivating a database retrieval oriented stance for model-based object recognition. We are developing a new hierarchical approach for the organization of model databases for indexing and retrieval of a ranked list of probable matching models for an input query object. The approach basically consists of three steps: (1) classification of models according to the independence of feature data associated with the models; (2) clustering of the model data in the classification space; and (3) a ranking method based on mutual information for ordering the retrievals. In step (1) we use a variant of factor analysis, called correspondence analysis, wherein the model data and their extracted feature sets are analyzed in a common, reduced-dimension factor space. In step (2) first, we find clusters of models in the factor space. Then, we automatically select a few "key" models in each cluster and generate a triangle inequality based pruning method using the model-to-model distances in the factor space. This eliminates a large set of models from candidate list of matching models for an input (query) object, in an efficient manner. Finally, in step (3) we rank the model candidates based on their mutual information with the input (query) object and retrieve the top model candidates as most probable matches. Since, in our approach, correspondence analysis considers the set of models according to data independence, the eliminated model candidates have low mutual information with a query object and this avoids a large number of unnecessary but expensive computations of mutual information between a query object and database models.

The results on MSTAR data using our approach are summarized in Figure 12. The three databases, Database 1, Database 2 and Database 3, correspond to depression angle differences, configuration differences and articulation differences, respectively.

In the future we will investigate ways to make our approach more time efficient, especially the ranking of the hypothesis.

7 Other Research

Other areas of ongoing work include navigation and obstacle detection. We are developing a mobile testbed, called UCRover for experiments in perception and learning [21]. 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 [16, 15, 18]. We have also done research on terrain interpretation using multispectral images [10].

8 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 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


975
Figure 12: Performance curves for the SAR target database.


