IR and SAR Automatic Target Detection Benchmarks

U. Jaeger[a], H. Maier-Herburger[b], C. Stahl[c], N. Heinze[a], D. Willersinn[a]
[a]Fraunhofer Institute for Information and Data Processing (IITB), D-76131 Karlsruhe, Germany
[b]EADS-Dornier GmbH, D-88039 Friedrichshafen, Germany
[c]EADS Military Aircraft, D-81663 Ottobrunn, Germany

ABSTRACT

This contribution describes the results of a collaboration the objective of which was to technically validate an assessment approach for automatic target recognition (ATR) components[1]. The approach is intended to become a standard for component specification and acceptance test during development and procurement and includes the provision of appropriate tools and data.

The collaboration was coordinated by the German Federal Office for Defense Technology and Procurement (BWB). Partners besides the BWB and the group Assessment of Fraunhofer IITB were ATR development groups of EADS Military Aircraft, EADS Dornier and Fraunhofer IITB.

The ATR development group of IITB contributed ATR results and developer´s expertise to the collaboration while the industrial partners contributed ATR results and their expertise both from the developer's and the system integrator's point of view. The assessment group's responsibility was to provide task-relevant data and assessment tools, to carry out performance analyses and to document major milestones.

The result of the collaboration is twofold: the validation of the assessment approach by all partners, and two approved benchmarks for specific military target detection tasks in IR and SAR images. The tasks are defined by parameters including sensor, viewing geometries, targets, background etc. The benchmarks contain IR and SAR sensor data, respectively. Truth data and assessment tools are available for performance measurement and analysis. The datasets are split into training data for ATR optimization and test data exclusively used for performance analyses during acceptance tests. Training data and assessment tools are available for ATR developers upon request.

The work reported in this contribution was supported by the German Federal Office for Defense Technology and Procurement (BWB), EADS Dornier, and EADS Military Aircraft.

1 Introduction

Today high-resolution imagery from reconnaissance and surveillance systems provides a most valuable source of information to the military decision making process. Due to an increasing number of available sensors, computer power and by means of world-wide data networks, imagery of different spectral bands are readily available from nearly every place of the world, at nearly any time and even as continuous data streams. Experiences show that the exploitation of such huge amount of data in a very short time constitutes a today´s bottle-neck.

A sequence of preparing steps is necessary to overcome this problem by means of computer algorithms, e.g. acquisition of the underlying task and checking the existence of available solutions, thus leading to a goal-oriented development or adjustment of existing components. In any case the military user or those individuals responsible for procurement must either check the algorithm´s usefulness in accordance to the user´s needs as a kind of acceptance test or give advice for the specification of a new algorithm which may lead to a development process. Both cases are supported by the assessment approach described in this paper which is intended to become a standard for ATR component specification during development and acceptance tests for procurement purposes.

The assessment approach is a result of a collaboration coordinated by the German Federal Office for Defense Technology and Procurement (BWB). Partners besides the BWB and the group Assessment of Fraunhofer IITB were

Automatic Target Recognition XIV, edited by Firooz A. Sadjadi, Proceedings of SPIE Vol. 5426
(SPIE, Bellingham, WA, 2004) · 0277-786X/04/$15 · doi: 10.1117/12.542102

ATR development groups of EADS Military Aircraft, EADS Dornier and Fraunhofer IITB. The industrial partners not only contributed ATR results but also put in their expertise both from the developer's and the system integrator's point of view. The assessment group's responsibility was to provide task-relevant data and assessment tools, to carry out performance analyses and to document major milestones.

A software tool named ACoViS (Assessment of Computer Vision Systems) has been developed by the assessment group of IITB to support the entire assessment procedure. ACoViS has been used for validation of the approach by assessing a total number of 9 different ATR algorithms for infrared (IR) and SAR imagery provided by the different development groups. A more detailed description of ACoViS can be found in chapter 4.

Input data consist of imagery data and truth data which describes the desired output of the ATR algorithm. The datasets are split into training data for ATR optimization and test data for performance analyses during the succeeding acceptance test. During the training phase truth data is delivered to the developers, too. The underlying algorithm task, imagery data and the ACoViS tool are now made available to other individuals upon request and thus could establish rational performance goals that will lead to superior algorithm performance. This, in turn, may be called a benchmark (R.C. Camp[2]) which is well suited to the assessment of other ATR algorithms if they are dealing with a similar task, similar sensors and similar scenarios. Moreover, by using the benchmarks there is no need for expensive measurement campaigns and time consuming truth annotations which imply laborious on-site verification (e.g. documenting exact type, location and orientation) of all objects under consideration. Furthermore, reusing the benchmarks not only provides some kind of quality control of the development process but at the same time simplifies documentation of the development progress. Last but not least, using a benchmark which is well known, transparent to and widely accepted by the community is beneficial to the marketing aspect as well because ATR software which has been assessed according to the approach could be thought of having undergone a certain level of quality check.

The outline of the paper is as follows: section 2 gives an overview of the assessment approach. Section 3 describes the benchmark specification. Section 4 clarifies the application of the benchmarks by means of ACoViS. Conclusions and acknowledgement are found in section 5 and 6.

2 Assessment Approach

Since a couple of years Fraunhofer IITB has performed research work on the evaluation framework for the assessment of ATR algorithms in order to objectively determine their usefulness[1]. The underlying evaluation process first defines and determines appropriate performance measures to create an algorithm´s performance profile. Then, from the user´s point of view an expectation profile is defined describing what the user ideally expects from the ATR algorithm. Basically, assessment is done by comparing both the performance profile and the expectation profile. Additional weighting and combining of deviations between these two profiles by an appropriate assessment function lead to a single number (Figure Of Merit, FOM) which can be used for comparison and ranking of different ATR algorithms.

The following details are closely related to past work of Fraunhofer IITB presented in a preceding SPIE-paper[1]. To rely on fixed and standardized assessment procedures 3 main algorithm classes detection, classification and estimation were introduced and to proper organize the sequence of necessary steps a 3-phase approach, consisting of the definition phase, the tuning phase, and the evaluation phase was developed. The definition phase sets up the foundation of the entire evaluation process including definition of algorithm task, task requirements and a list of parameters which might affect performance. Also part of this phase is the acquisition of appropriate data, the definition of performance measures and the definition of the assessment function. Depending on the algorithm´s development state, analysis of preliminary results might be done as well. During the tuning phase refinement of algorithm specification and assessment procedure takes place. E.g. the algorithm is adopted to the given application. Therefore, a subset of the evaluation data including truth data (which reflects the desired results) and ACoViS is given to the developer. The tuning phase is concluded by ATR specification and checking for specification completeness. Finally, evaluation is accomplished during the subsequent evaluation phase carried out by the assessment group of Fraunhofer IITB. Evaluation uses explicit test data which must be distinct from the data used in tuning phase and should have never been used by the developers before. Evaluation ends up with a statistical analysis and, if there is more than one algorithm, ranking their usefulness in a specific order. For details about the ranking procedure please refer to the paper of S. Fries[3] given in session 5426-21 of this conference.

3 Benchmark Specification

Benchmark specification is based on a well defined military target detection task which in turn leads to an appropriate image exploitation task the ATR algorithm has to deal with. The image exploitation task must be supported by an appropriate set of imagery and detailed truth data.

The algorithms considered in this paper are designed to detect (non-camouflaged) military land vehicles in infrared (IR) and synthetic aperture radar (SAR) imagery. The algorithm´s output is a set of detection hypotheses, providing the center coordinates of possible targets as well as a number of certainty. Appropriate truth data which is provided by the assessment group of Fraunhofer IITB consist of a set of polygons surrounding objects under consideration. The polygons are drawn by hand using the underlying image and, if available, documents from an on-site measurement team as a reference.

By default 6 different truth object types are distinguished (see table below) but any appropriate object description could be added to the list.

Vehicles	the type of signatures the algorithm should detect
Clutter	signatures within the images which look like vehicle types but which from measured ground truth are not a vehicle for sure.
Unknown	we don´t know what it is, but it could be mixed up with a vehicle
OMVs	Other Military Vehicles which are not under consideration at the moment (e.g. camouflaged or concealed)
Urban Area	areas of non-military buildings e.g. homes and houses, villages, cities etc.
Landmarks	artificial hot spots made by special on-site hardware (e.g. heaters for IR or corner reflectors for SAR) for sensor calibration purposes

The different truth object types enable in-depth analyses of the algorithm´s behavior.

Truth annotation in real imagery was found to be a critical part within the assessment process because of its direct impact on ATR performance. The assessment group of Fraunhofer IITB had to guaranty a minimum truth quality (e.g. considering precision, completeness, consistency etc.) in huge imagery where usually only a small area had been fully documented by an on-site measurement team.

Both IR and SAR imagery comes from the IITB´s imagery data base. The IR sensor is a 8-13µ Thermal Infrared Camera Module (TICM) mounted on an airplane. Using different attitudes and a set of different aspect angles (see table below) the data set consists of 15 different classes of aspect views with a total of 521 images including about 1300 truth object types and covering a ground area of about 40 km².

	attitude	field of view	roll angle	pitch angle
1	3000 ft	W	0	0
2	3000 ft	W	0	45
3	3000 ft	W	45	0
4	3000 ft	W	45	45
5	3000 ft	T	0	0
6	3000 ft	T	0	45
7	3000 ft	T	45	0
8	3000 ft	T	45	45
9	1500 ft	W	0	45
10	1500 ft	W	45	0
11	1500 ft	W	45	45
12	1500 ft	T	0	0
13	1500 ft	T	0	45
14	1500 ft	T	45	0
15	1500 ft	T	45	45
W: (15 x 10)° T: (6 x 4)°				

table 1: viewing geometries of IR sensor

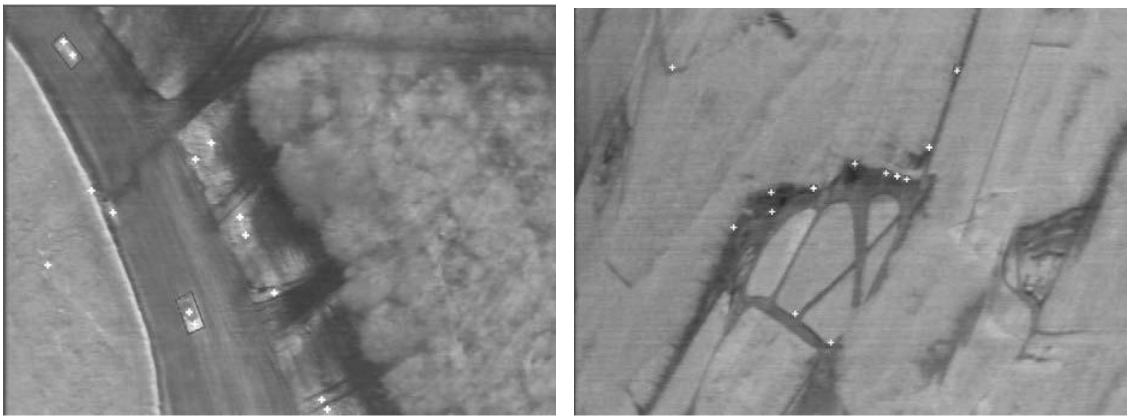


figure 1: 1500-w-0-0 (left side) 3000-w-45-45 (right side) sample images (rectangles = truth, crosses = algorithm results)

SAR imagery originates from EADS Dornier´s experimental X-band SAR (DOSAR) and consists of 26 images with nearly 1600 annotated truth objects. Ground covered area sums up to about 135 km². The sample SAR image below shows an excerpt of the different truth objects mentioned above.

figure 2: SAR image sample (section). (v)ehicle, (u)nknown, (c)lutter, (omo) other military objects and algorithm results (bright dots)

4 Benchmark application by means of ACoViS

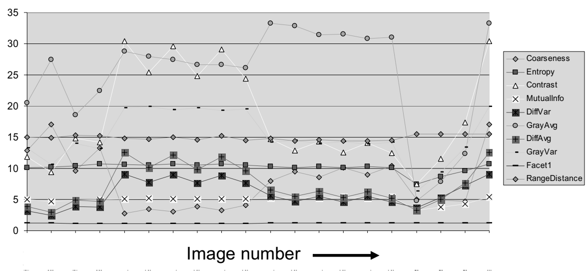
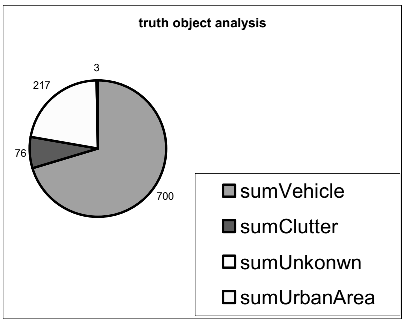
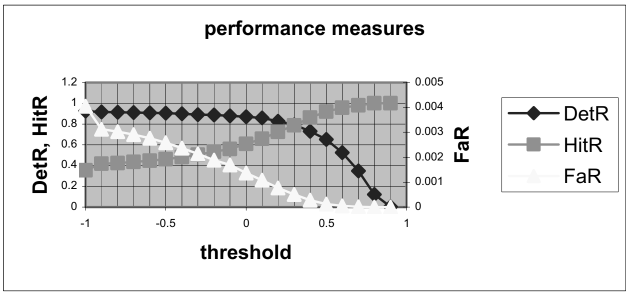
During its work the assessment group of Fraunhofer IITB developed several software tools which melt together into a unique integration platform called ACoViS (Assessment of Computer Vision Systems). Thus, ACoViS reflects the assessment approach outlined above and has been profoundly tested within the collaboration considering the benchmarks described in the previous section. Although ACoViS has migrated from pure experimental state towards a product-like one it is still a candidate for improvements due to the experiences from our daily work.

ACoViS components are implemented in JAVA and have been successfully ported to several hardware platforms including WINDOWS, LINUX and HP UNIX derivatives. Its main components support visualization, truth annotation and assessment of 3 main algorithm classes, namely detection, classification and estimation. Assessment of detection algorithm has undergone in-depth verification during the collaboration by applying the whole assessment process to 9 different algorithms provided by industry and developers of IITB. Figure 3 shows ACoViS with its viewing component ACoView.

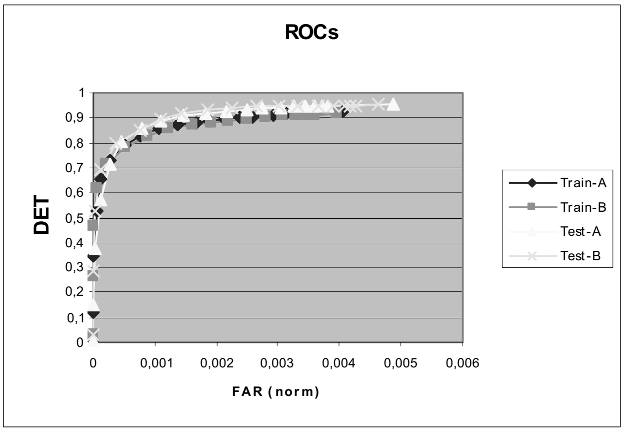
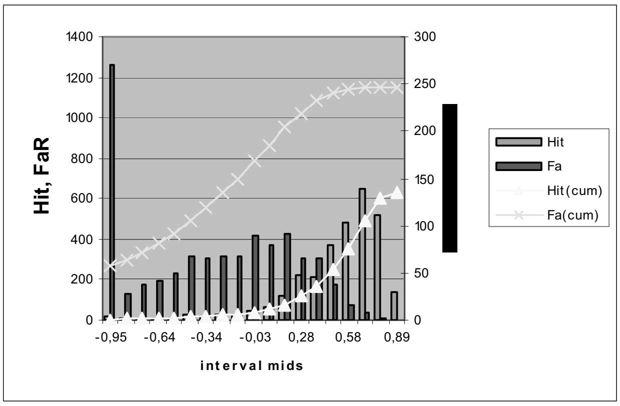
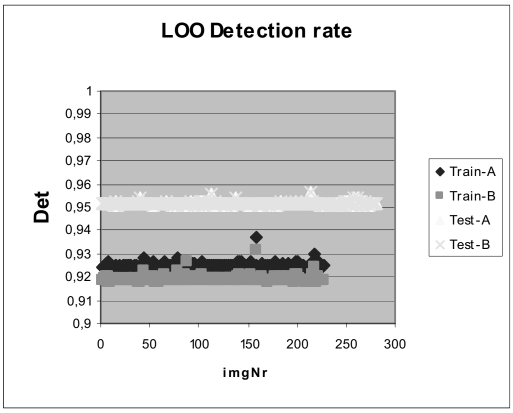
figure 3: ACoViS with its viewing component ACoView

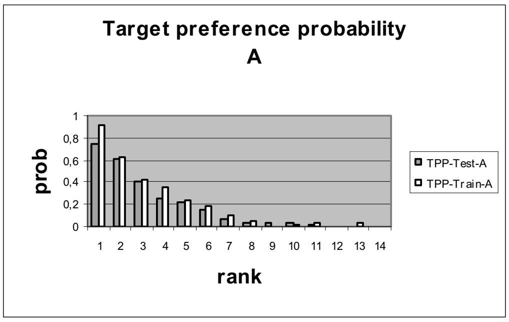
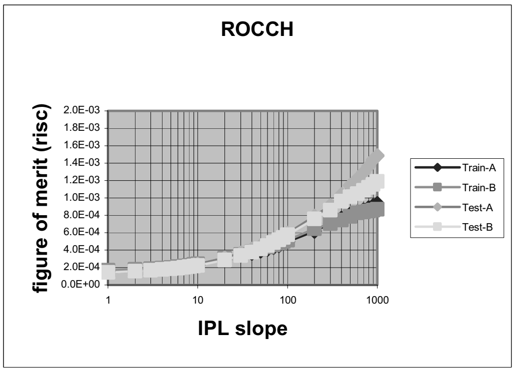
The following table highlights important steps of the entire assessment process supported by ACoViS. Example graphics (which do not correspond to the results of the collaboration) are added at the right side of the table for better understanding of the results.

The assessment process normally starts with a comprehensive data analysis, both on imagery and truth data. Image property measures for example give an idea of how similar training and test data are (1). A follow-up truth analysis (2) examines the frequency of truth objects both on each image and the whole image set and assures that the image exploitation task is covered by the imagery. In step (3) algorithm results are examined and compared to the truth data. Using a point-in-polygon-approach adequate performance measures are calculated and displayed. Because of its convenience and usefulness the receiver operation characteristic (ROC) which shows the detection rate as a function of the false alarm rate is displayed in step (4). Constructing the ROC relies to the fact that a "sensitivity parameter" is available for the algorithm which could be changed for optimal performance. Practically, when calculating detection rate and false alarm rate only those algorithm results are considered whose certainty measures lie beyond a variable threshold. A scoring analysis (5) helps investigating a certain threshold value which may optimally separate detections from false alarms. A leave-one-out-analysis (6) gives an idea of the influences of individual images to the performance measures and thus helps to identify outliers. In especially, assessing ATR algorithms for weapon systems requires some knowledge about the target preference probability (7), which answers the question of how often the strongest (second strongest … etc.) detection hypothesis corresponds to a real target. The assessment process finally ends up with the calculation of a single absolute number (figure of merit, FOM) reflecting the performance of the algorithm in accordance to the user´s requirements, thus enabling a ranking between competitive algorithms and determining the best suited one for the given application (8). This step requires the knowledge of cost factors for wrong decisions (e.g. false alarms and non-detections) which must be provided by the user. Calculating the FOM for different applications is based on the ROCCH (receiver operation characteristic convex hull) principle. Details are found in a companion session of this conference[3].

No	ACoViS feature	examples (for illustration purposes only, values do not relate to the work within the collaboration)
1	**image property measures** e.g. • coarseness • entropy • contrast • mutual info • diff var • gray var • facet1 • range distance allow comparison of properties between different image data sets	
2	**truth object analysis** e.g. examining frequency and mean area of the following truth-types • vehicle • clutter • unknown • urban area • landmark • other military objects	
3	**performance measures** e.g. • detection rate • false alarm rate • hit rate • multiple detection rate	

<table>
<tr>
<td>4</td>
<td>receiver operation characteristic (ROC)

displays detection rate as a function of false alarm rate while varying the threshold level for the certainty measure of hypotheses given by the algorithm</td>
<td>
</td>
</tr>
<tr>
<td>5</td>
<td>scoring analysis
using a variable threshold to answer the question of how many detections and false alarms lie above the threshold level, thus helping to determine the operation point for optimal separation between detections and false alarms.</td>
<td>
</td>
</tr>
<tr>
<td>6</td>
<td>leave-one-out analysis
gives an idea of the influence of every single image, detects outliers</td>
<td>
</td>
</tr>
</table>

7	**target and false alarm preference probability** answers the question of how often the strongest (second strongest…) hypothesis corresponds to a real target (detection) or a false alarm	
8	**comparison (ranking)** of competitive algorithms according to a risk number as a figure of merit based on the ROCCH principle which uses the slope of an IPL (iso-performance line) to consider different applications. For a detailed description please refer to S. Fries[3].	

5 Conclusion

This contribution describes an assessment approach for automatic target recognition (ATR) components which was developed and validated during a collaboration between Fraunhofer IITB (with both the ATR development group and the assessment group) and ATR development groups of EADS Dornier and EADS military aircraft using 9 different algorithms for IR and SAR imagery. The collaboration was coordinated by the German Federal Office for Defense Technology and Procurement (BWB). The algorithm developers of IITB and industry contributed ATR algorithm results and their development expertise to the collaboration. The industrial partners also added their expertise from the system integrator's point of view. The IITB assessment group's responsibility was to provide task-relevant data and assessment tools, to carry out performance analyses and to document major milestones. The approach is intended to become a standard for ATR component specification during development and acceptance test for procurement purposes and includes the provision of appropriate tools and data. From the collaboration two approved benchmarks for a specific military target detection task based on IR and SAR imagery have been developed. The benchmarks contain IR and SAR sensor data as well as truth data and the assessment tool ACoViS for performance measurement and analysis.

6 Acknowledgement

The authors of this contribution would like to thank the German Federal Office for Defense Technology and Procurement (BWB) for its support and approval of publication.

7 References

1. S. Fries et al., Evaluation Framework for ATR Algorithms, Proceedings of SPIE 1999, SPIE volume 3718
2. Robert C. Camp, Business Process Benchmarking, ASQ-Press 1995
3. S. Fries et al., Automatic target Detection Assessment Using ROC Curves, SPIE Defense & Security Symposium 12 – 16 April 2004, conference no. 5426-21