Defense Technical Information Center Compilation Part Notice

ADP023076

TITLE: Synthetic Infrared Data For Target Identification Training and Testing

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Proceedings of the Ground Target Modeling and Validation Conference [13th] Held in Houghton, MI on 5-8 August 2002

To order the complete compilation report, use: ADA459530

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:
ADP023075 thru ADP023108

Synthetic Infrared Data For Target Identification Training and Testing

Bruce A. Weber, Joseph A. Penn
U. S. Army Research Laboratory
AMSRL-SE-SE
2800 Powder Mill Road
Adelphi, MD 20783

ABSTRACT

The performance of infrared (IR) target identification classifiers, trained on randomly selected subsets of target chips taken from larger databases of either synthetic or measured data, is shown to improve rapidly with increasing subset size. This increase continues until the new data no longer provides additional information at which point classifier performance levels off. It will also be shown that subsets of data selected with advanced knowledge can significantly outperform randomly selected sets, suggesting that classifier training-sets must be carefully selected if optimal performance is desired.

Performance will also be shown to be dependent on the quality of data used to train the classifier. Thus while increasing training set size generally improves classifier performance, the level of classifier performance improvement will be shown to depend on the similarity between the training data and testing data. In fact, if the training data to be added to a given set of training data is unlike the testing data, performance will often not improve but may possibly diminish. Having too much data can be as bad as having too little.

Our results again [1] demonstrate that an IR target-identification classifier, trained on synthetic images of targets and tested on measured images, can perform as well as a classifier trained on measured images alone. We also demonstrate that the combination of the measured and the synthetic image databases can be used to train a classifier whose performance exceeds that of classifiers trained on either database alone.

Results suggest that it may be possible to select data subsets from image databases that can optimize target classifiers performance for specific locations and operational scenarios.

Keywords: ATR, classifier, target identification, synthetic images, infrared

INTRODUCTION

Data available to train target-identification classifiers has long been known to be insufficient to produce robust target identification against new image datasets. This problem is due to the intimate relationship of data, the limited conditions under which data is collected, and the statistical nature of the classifiers in representing the data.

To remedy this, statistical classifiers need large amounts of dissimilar data for training. Obtaining measured data from field-tests is expensive so we have produced this data synthetically. To achieve this data synthesis we have created synthetic images that not only look like measured images, but moreover perform like measured images. To this end we use a comparison of measured, versus synthetic data trained classifier performance, as a quantitative measure of synthetic data validation.

Over time the performance of our synthetically trained classifiers has improved. Attention to detail in the comparison of measured and synthetic images was crucial. However, since our databases consisted of tens of thousands of images, a direct one-to-one image comparison was impractical. Instead we made our comparisons using the image-like states of a trained K-Means classifier. Such states, called codevectors or templates, are composite images that summarize many similar individual image instances. This data compression makes practical the measured versus synthetic image comparison, and the subsequent adjustment and/or addition of new images and codevectors.

Last year we reported that synthetic-data trained classifiers could perform as well as measured-data trained classifiers [1]. However, as the subset of synthetic data was specially selected, we decided to investigate whether similar classifier performance could be achieved if randomly selected data was used to train the classifier.

Subsequently, we have expanded our synthetic database of four targets to slightly more than 90,432-files. From these images we have selected subset databases of varying size to train and test our classifiers. Two techniques were used to select subsets of synthetic image target-chips: (1) systematic unsupervised random selection to remove data bias and insure robustness to changing test conditions, and (2) biased selection using advanced knowledge optimized performance for limited conditions. The test database contained 5501-files of measured target-chips.

We next describe data creation, classifier training and testing, and data selection.

MEASURED IMAGE DATABASE EMULATION

For benchmark testing, we use the COMANCHE database of measured-world images. This database consists of approximately 30,000 image scenes containing different image instances of 10 different target types, 72 angular aspects spanning 360 degrees, three geographical locations including Yuma, Arizona, Hunter-Liggett, California, and Grayling, Michigan, both summer and winter seasons, and full diurnal time cycle. Extracted from these scenes are approximately 22,000 target chips that are divided into two databases: the SIG database of approximately 1,500 chips/target, including all 72-target aspects every 5°, for each of 10 targets in the clear, and the ROI database of approximately 1,300 chips/target, including only 8-target aspects every 45°, for each of 5 targets near clutter. All five of the ROI target types are included in the set of ten SIG target types. By any measure, the ROI target images are more difficult to recognize.

For comparison we have created synthetic data that emulates four of the ten targets in the SIG database. The four ground targets are: HMMWV, M60, T72, M113. Of these four targets, three are targets in the ROI database. We have modeled each target in identical conditions and locations, and have simulated realistic exercise routines to produce thermal signatures consistent with observed data. Similar exercise information is not available as part of the ground-truth for the measured data.

SYNTHETIC-IMAGE DATABASE GENERATION

Isothermal nodes for each target model were obtained using the PRISM [2] commercial code and IR-images were rendered using the Army Research Laboratory's (ARL) CREATION code [3]. Figure 1 shows a schematic of the algorithmic methodology [1] for generating, training, and testing of the synthetic target chip database. The dotted, blue Rhino-Muses [5] generation path provides a new methodology to be used in the coming year to replace the PRISM path used in this research for producing the isothermal target nodes.

For the purposes of this research, we quantify the validation of synthetic images by the performance to which a target identification classifier, trained on synthetic target chips, can achieve as compared to that achieved using measured target chips.

To create a synthetic database of ever increasing size we begin by selecting a small number of files from a large, parent database. This becomes database A. Next we add new files to create database B. Database B and each subsequent database of increasing size contain all the previous files selected for database A plus additional files. Each data selection is performed so as to uniformly represent the number of different targets but not necessarily uniformly represent their aspect or operating conditions. Files are randomly selected to form a database and then databases are enlarged, by the addition of more randomly selected files.

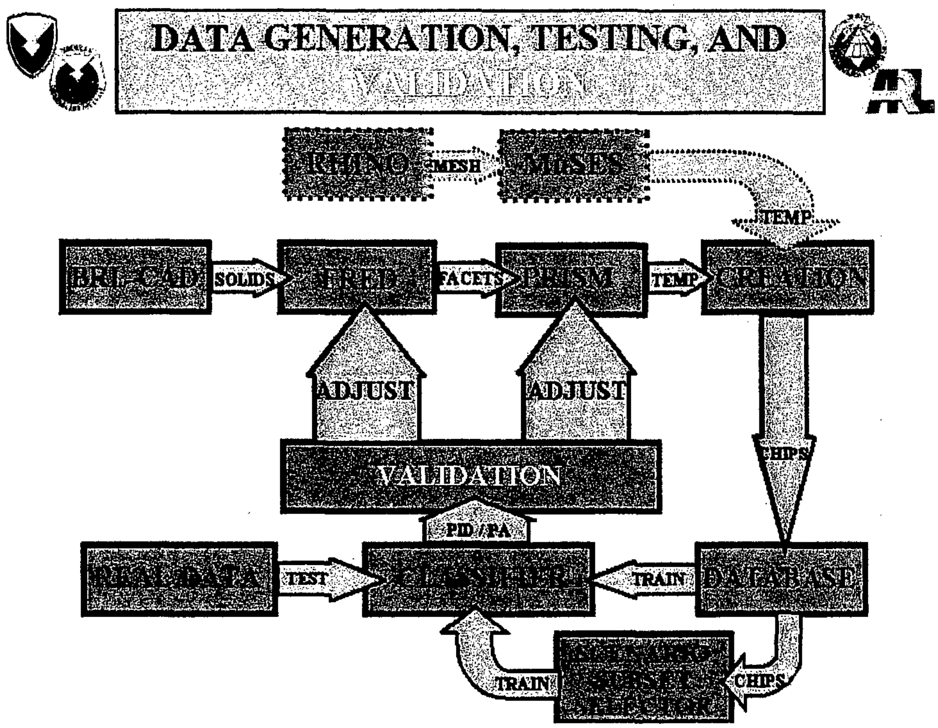

Figure 1. Iterative process for generating, training, and testing the synthetic image database.

As each database is formed a classifier is trained on the database and then the trained classifier is tested against the sequestered ROI database of measured-data.

No effort was made to adjust the ordering of the training data since the process of training averages like data into codevectors. Though training ordering does not affect a classifier's performance, database selection ordering is important with respect to database formation. (EXPLAIN THIS MORE CLEARLY!) Thus an intelligent or even lucky selection of a subset of data can outperform an unintelligently or randomly selected subset. Of course this is true only for subsets of large databases; when all of the data is used ordering does not matter.

CLASSIFIER TRAINING AND TESTING

The classifier used in this research was developed at ARL and is described as a minimized mean-squared-error (MSE) encoder [4]. All input target chips, both in the training and testing phase, are intensity scaled to zero mean and unity variance.

The classifier has two training modes: the K-Means mode, and the learning LVQ-mode. The LVQ mode is an additional mode that adjusts the results first obtained by the K-Means mode. In the K-Means mode a target-like region (identified in a previous step by a target detection algorithm) is extracted by a series of aspect dependent windows, enlarged to a fixed size, and wavelet decomposed into four (4) sub-bands. The K-means mode trains the classifier by collecting like-aspect sub-band decompositions and then creating codevectors by averaging the sum. The LVQ-mode adjusts the codevector centroids by moving the centroids to better match the training data. For the purposes of this paper we will be using the K-Means mode alone for classifier operation.

During the testing phase, an unknown target chip is extracted, sized, wavelet decomposed, and compared with each of the codevectors of each of the codebooks for each of the learned targets. The commonly used similarity measure Mean-Square-Error (MSE) is used for the comparison. The target-aspect sub-band class with the lowest MSE is declared the identity. This declaration is made independent of correct aspect identification.

DATA SELECTION

Scenarios and circumstance dictate data selection for target classification training. If a large enough database existed that contained images of targets in varying location, terrains, and seasons, then all of the data could be used to train a classifier to perform robustly in any condition. However databases are generally quite limited in the conditions they represent thus limiting the robustness of trained classifiers. Consequently scenario specific classifiers provide an alternate approach of achieving desired performance. Such classifiers are developed using advanced knowledge of a test site and/or test conditions.

Advanced knowledge might include data from recent reconnaissance, previous data collections from the test site or from sites similar in geophysical location and weather. But classifiers trained on advanced knowledge would have to be used selectively. Clearly one would not expect good performance from a classifier trained on data of targets in the Sinai in summer when the test site is Bosnia in winter.

To examine these possibilities we performed two tests: one in which synthetic data was selected randomly from a larger database to train a classifier and another in which the data was selected with advanced knowledge.

Random Data Selection

To demonstrate that synthetic data can be used to train classifiers and determine whether such training provides classifier performance comparable to classifiers trained on measured-data, we trained classifiers on either synthetic or measured-data, tested the trained classifiers on a sequestered set of measured-data, and compared the results.

Eight subset databases of increasing size were selected from the parent synthetic or measured databases. The selection was done so that any single database contained all of the data of any smaller database plus additional data. The data for each subset database, either synthetic or measured, was selected randomly. Equal percentages for each target in each subset was maintained relative to the larger, parent database, but the distribution of target aspects, seasons, times of day, and vehicle exercise states was not. The parent database for the measured-data contained 5,501-files and 90,432-files for the synthetic-data. Eight subset databases of 0.5, 1, 2.5, 5, 10, 25, 50, and 100% respectively were selected from either the synthetic and measured databases.

In addition, to testing classifiers trained on either synthetic or measured databases alone we also trained classifiers on combinations of synthetic and measured data. This was done so as to examine if overall performance could be improved by adding databases. If the data in the synthetic and measured datasets are similar then we might expect classifiers trained on any combination of data from the two sets to perform similarly. However, if the data in the datasets have some target chips that are not represented in the other set then combining data might improve the performance relative to either set alone. Of course any increase in performance would also indicate an increased similarity with the test dataset.

Randomly choosing data to form subset databases does not guarantee that the best performing groupings would be selected. In fact as increasing amounts of data dissimilar to the test database are added to a subset training-database some codevectors may be readjusted to be more unlike the test data degrading performance. This problem is addressed in the next section.

Advanced Knowledge Data Selection

Next we use a 3-step procedure to demonstrate that advanced knowledge can be used to increase classifier performance for targeted scenarios. First we divided a database of measured data into two equal sized subsets: a test dataset to be used solely for testing purposes, and a training dataset to be used for data selection. The training subset emulates our advanced knowledge in the form of previously obtained site data. This could be recent reconnaissance data from the test location, or data from another similar geophysical location. Next we trained a classifier with the training dataset, ran the entire synthetic database through the classifier and collected all files that were correctly identified. In the final step we used this selected database of synthetic target chips to train test our classifier. The dataset tested was the sequestered remaining half of the measured dataset not used to select the synthetic training dataset.

Two synthetic databases were selected in this way. One was chosen to emulate the ROI database and the other the SIG database. The ROI-Like set of synthetic data was selected using the ROI set of measured data, and the SIG-Like set selected using the SIG set of measured data.

For example the ROI-Like database was chosen by running the entire 90,432-file synthetic database through a classifier trained on the ROI training data. Selected ROI-Like images were required to be correct for both target identification and target aspect. They were also required to have a confidence value greater than 0.9. The confidence value is defined to be the difference in probability for correct identification between the first and second identification choice of the classifier. Using this process approximately 7,000 target chips were selected as the 3-target, ROI-Like database.

The SIG-Like database was chosen similarly. Using this process approximately 8,000 target chips were selected as the 4-target, SIG-Like database.

These two databases taken together with the 4-target, 5,156-image SIG database, and the 3-target, 2,200-image ROI testing database provides all of the datasets used to train and test our classifier.

RESULTS

Introduction

The two sections that follow describe our results: one addresses results obtained using randomly chosen training datasets and the other using datasets chosen using advanced knowledge.

For the testing described below, the trained classifiers are tested on the 3-target Measured-ROI-Testing set alone. Thus for performance in which Target 4 of the training set is not part of the testing set, no performance values are listed in column 4 since Target 4 can not be the correct identification. However this does not mean that Target 4 can not be declared to be the identity, only that it can not be the correct identity.

For the four-target test we are investigating, trained classifier targets 1 and 4, and targets 2 and 3 are of similar size, whereas targets 1 and 4 are smaller than targets 2 and 3. Target 1 is a HMMWV, target 2 is an M60, target 3 is a T72, and target 4 is an M113.

For all performance results reported here we used the K-Means mode of the LVQ-classifier. The K-Means parameters were chosen to optimize the number of codevectors generated for good target identification.

Results From Experiments Using Randomly Chosen Databases

In this section randomly selected training data is used for classifier training. Figure 2 shows four curves that summarize classifier performance using different sets of training data. The performance is specified as the probability of correct target identification (PID), in %/100, as a function of the percentage of the total number of training files added. For this graph a PID of 1.0 corresponds to all targets being identified correctly.

In Figure-2 the black curve (circles) shows the increase in performance of the SIG trained classifier as the percentage of the 5156-file dataset is increased. No PID saturation is observed since the PID never levels off. This indicates that each added increment of SIG files are sufficiently different from the SIG files used in smaller sets so as to add to the overall performance. The maximum PID achieved is 82%.

The blue curve (squares) shows the increase in performance for the synthetic data trained classifier as a percentage of the 90,432-file dataset is increased. PID saturation is observed as the PID begins leveling off for a training set of 9,000-files and is certainly level by about 45,000-files. Two possibilities may contribute to the PID saturation: (1) the new data being added is not sufficiently different to aid in the training, and (2) that the number of codevectors being created in the training process is not increasing. Figure-3 shows the increase in the number of codevectors as a function of the number of files per target is increased. Clearly there is no saturation in the number of codevectors being created as the database size increases. This suggests that the new data being added to the training databases is too similar to previously used files so as to not improve the performance of the classifier. The new data is redundant with other data in the overall 90,432 database. The maximum PID achieved is 70%. This suggests that randomly choosing data produces a smooth increase in performance, showing no evidence of subsets of data that outperform the total, and that using the entire synthetic database produces a level of performance 12% less than is produced by training the classifier on measured-data in the SIG database.

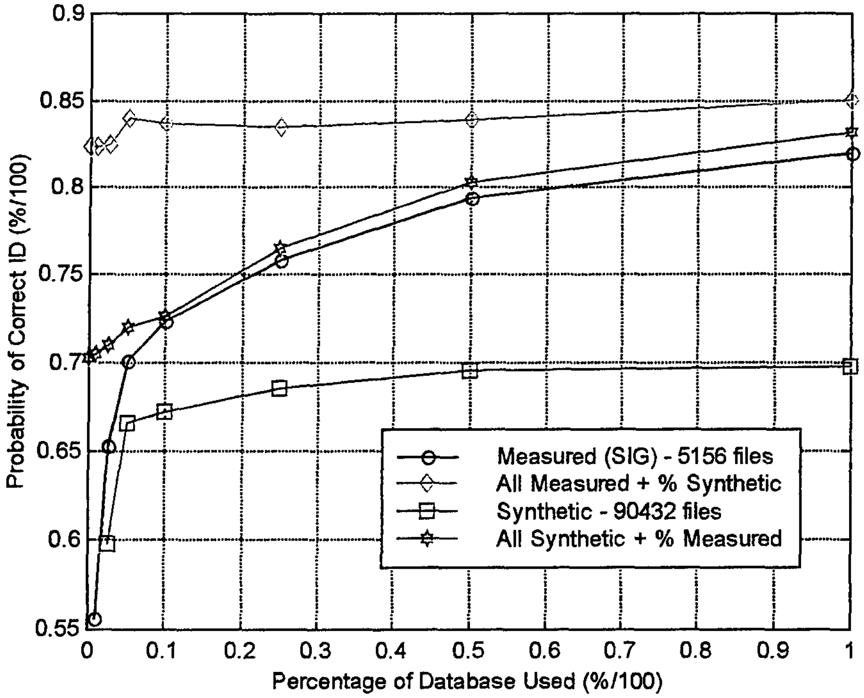

Figure 2 - Performance Comparison of K-Means Trained Classifiers Tested On Sequestered (ROI) Dataset of 5501-Files

The green curve (diamonds) shows the improvement in PID as increasing percentages of the 90,432-file synthetic database are added to the 5156-file SIG database. The performance is shown to improve to about a PID of 85%, a 3% increase over the performance observed using the SIG data alone. This difference is significant since the statistical uncertainty in recognizing files from the 5501-file ROI database is about 1.3%.

The violet curve (stars) shows the improvement in PID as increasing percentages of 5156-file SIG database is added to the 90.432-file synthetic database. The performance is shown to improve to about a PID of 83%, a 13% increase over the performance observed using the synthetic data alone and a 1% increase over the performance observed using the SIG data alone. This statistically insignificant difference shows that the SIG, and the synthetic + SIG trained classifiers perform essentially the same.

Clearly there is a difference between this end result here and that of the green curve (diamonds) for which one would expect the end points to be the same, that is, both trained on all of the SIG and all of the synthetic data. The difference here is the order in which the data used to train each subsequent classifier was selected. For the violet curve (hexagons) the classifier was first trained on all of the synthetic data before any SIG data was added. This biases the codevectors to be synthetic like. For the case of the green curve (diamonds) the classifier was first trained on the SIG data alone before any synthetic data was added. For this case the codevectors were first SIG like before they were adjusted by the addition of synthetic data. We conclude that the order in which data is added to the training process affects the final performance results achieved. We will see further evidence of this in the next section where datasets were selected using advanced knowledge.

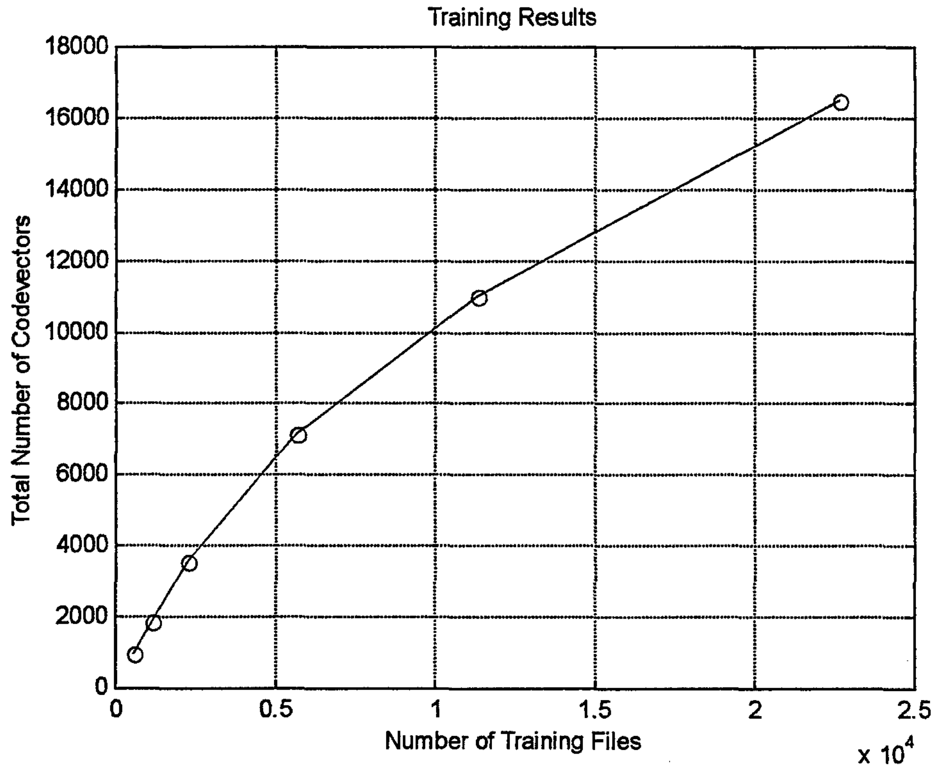

Figure 3 - This figure shows that the number of codevectors increases as the K-means classifier is trained on increasing amounts of synthetic data. The database included 22,608 target chips for each of four targets or 90,432 target chips in all.

Results From Experiments Using Advanced Knowledge

In this section advanced knowledge is used to choose training data. Tables 1 to 3 enumerate the identification probability (correct and incorrect) as a function of the target type. Such tables are termed a Confusion Matrix and the probabilities are shown in percent. The predicted target identities are listed down the left-most column and the actual identities are listed along the top-most row. The lighter, diagonal elements show the PID. A PID of 100% corresponds to all the targets of that type being identified correctly. Off diagonal elements represent the probability of misidentification.

Table 1 shows the confusion-matrix for the K-Means classifier trained on Measured-SIG data alone and tested on the sequestered Measured-ROI-Testing set. The overall PID is 82%. This is similar to the results reported by Chan et. al. [4]. Misidentifications of targets 2 and 3, both being tanks are often declared to be each other.

Probability of Correct Target Identification (%)				
Predicted/Actual	Target 1	Target 2	Target 3	Target 4
Target 1	85	4	6	-
Target 2	4	81	8	-
Target 3	2	12	81	-
Target 4	9	3	5	-

Table 1. A confusion matrix showing the probability for correct Measured-ROI target identification for a classifier trained on a database of Measured-SIG images. Overall probability for correct identification is 82%.

Table 2 shows the confusion matrix for the classifier trained on the union of the subsets of ROI-Like and SIG-Like synthetic data and tested on the sequestered Measured-ROI-Testing set. The overall PID is 81% indicating that the synthetic data trained classifier performs comparably to that of the Measured-SIG data trained classifier in Table 1. This is a different result than that obtained when the synthetic-data training sets were selected randomly. In fact, comparing this result with that of the violet curve (stars) in Figure 2, we see that at no point does the PID rise above 70%, almost 11% below the PID obtained by choosing the training data using advanced knowledge. This suggests that using too much data that is unlike the testing data to train a classifier can bias the classifier performance in a non-optimal way.

Probability of Correct Target Identification (%)				
Predicted/Actual	Target 1	Target 2	Target 3	Target 4
Target 1	90	9	14	-
Target 2	5	76	9	-
Target 3	4	14	76	-
Target 4	1	1	1	-

Table 2. A confusion matrix showing the probability for correct Measured-ROI target identification for a classifier trained on a database of the union of the ROI-Like and SIG-Like synthetic subsets. Overall probability for correct identification is 81%.

Table 3 shows the confusion matrix for the classifier trained on the union of the Measured-SIG, SIG-Like, and ROI-Like synthetic datasets. The overall probability for correct identification is 85%, an increase of over 3% from the single database results for either classifier in Tables-1 and 2 above.. This demonstrates that classifier performance can be improved when synthetic and measured databases are joined. Again misidentifications mix Target 2 for Target 3.

Probability of Correct Target Identification (%)				
Predicted/Actual	Target 1	Target 2	Target 3	Target 4
Target 1	85	2	4	-
Target 2	4	85	10	-
Target 3	2	8	80	-
Target 4	9	5	6	-

Table 3. A confusion matrix showing the probability for correct Measured-ROI target identification for a classifier trained on the union of Measured-SIG, SIG-Like, and ROI-Like synthetic datasets. Overall probability for correct identification is 85%.

For comparison, Table 4 shows the confusion matrix for the classifier trained on measured data alone. The training data consists of the union of the Measured-SIG database and the Measured-ROI-Training subset. Again the test set is the Measured-ROI-Testing subset. The overall PID is 88%. This augmentation of the Measured-SIG data with the Measured-ROI-Training data increased the classifier performance by less than 9% over the results shown in Table 1. We will benchmark this level of classifier performance since it alone uses data taken from the same database from which the testing set was chosen.

Probability of Correct Target Identification (%)				
Predicted/Actual	Target 1	Target 2	Target 3	Target 4
Target 1	97	2	2	-
Target 2	2	85	15	-
Target 3	2	13	81	-
Target 4	0	0	1	-

Table 4. A confusion matrix showing the probability for correct Measured-ROI-Testing image identification for a classifier trained on a database of the union of Measured-SIG and Measured-ROI-Training images. Overall probability for correct identification is 88%.

Using the results of Table 4 as the benchmark the relative performance for the results listed for Tables 1, 2, and 3 are 93.5, 92.4, and 96.8 percent respectively.

CONCLUSIONS

Required target identification classifier performance specifications are dependent on application. Reconnaissance performance specifications can be considerable poorer than fire-control specifications. For specialized scenarios specialized classifier development will be required.

Statistical classifiers need a lot of data to train, but our results show that choosing data must be done carefully and wisely otherwise performance can suffer. Specifically we have shown that subsets of synthetic infrared-images can be chosen randomly to train target classifiers, and that adding synthetic databases to measured databases can improve the performance of classifiers trained on either database alone. We have shown that a classifier, trained with synthetic images selected using advanced knowledge, and tested on measured images, can perform as well as a classifier trained on measured images alone. Finally we have shown that using advanced knowledge to select training data, classifiers can be trained to significantly outperform classifiers that are trained on randomly selected training data.

Finally, we have achieved these results with relatively low-resolution images, derived from extremely low-resolution target models. We have taken care to simulate physically reasonable target states commensurate with measured data and we have validated our data by comparing synthetic to measured data performance in the training and testing of target classifiers. Yet our simulations do not include target/background interactions.

New updated infrared simulators, with near real-time temperature calculations and new visualization tools are now available [5], and soon high-resolution models will also be available. These tools will make database development easier and more reliable. And soon target and background thermal interactions will be modeled also.

Our methods for selecting data demonstrate that unless care is taken when choosing data, a range of performance is possible. Yet our methods do not provide a method on how to achieve optimal performance from available databases, but the way to proceed is clear. Choose data selectively using classifiers trained with advanced, scenario specific information from both measured and synthetic databases, train and evaluate the classifiers performance on available or reasonably matching measured data, and add new training data as it becomes available.

PRODUCTS

As a product of our work we have packaged our synthetic data. Datasets include the 90,432-chip, 4-target, COMANCHE-type ground target synthetic set, the 8,000-chip training SIG-Like dataset, and the 7,000-chip, testing ROI-Like dataset. In addition, along with each synthetic image dataset we are providing individual chip ground-truth as to geophysical location, time of day, month, weather, and vehicle exercise history.

These sets are unclassified and available (Distribution-C) to qualified U.S. Government Agencies and their contractors. Qualified users must agree not to distribute the data without first obtaining prior written approval from ARL.

ACKNOWLEDGEMENTS

We wish to extend a special thanks to Sandor Der of ARL for many helpful discussions.

REFERENCES

[1] Bruce A. Weber and Joseph A. Penn, "Comparable Performance For Classifier Trained On Real Or Synthetic IR¬Images," *Proc.SPIE AeroSense/Automatic Target Recognition XI*, Vol. 4379, April 2001

[2] Curran, A.R. Bobbi Wood, Eric Martilla, Paul Rogers, "Enhancements to the Vehicle Designer / PRISM Interface," *Fifth Annual Ground Target Modeling and Validation Conference*, Vol. 5, 401-407, August 1994

[3]Hung M. Nguyen, Joseph A. Penn, Teresa A. Kipp, Giap Huynh, and Janice Colby, "The CREATION Scene Generation Program And Its Application To Development And Testing Of Automatic Target Recognizers," *ATRWG Proceedings*, Vol. 1, 56-73, (July 1996).

[4] Lipchen A. Chan, Nasser M. Nasrabadi, and Vincent Mirelli, "Wavelet-Based Learning Vector Quantization for Automatic Target Recognition", *Proc. SPIE, Aerospace/Defense Sensing and Controls,* Vol. 2755, 82–93, April 1996.

[5] Johnson, Keith, Allen Curran, David Less, Derrick Levanen, Eric Marttila, Teresa Gonda, and Jack Jones, "MuSES: A New Heat and Signature Management Design Tool for Virtual Prototyping," *Proceedings of the Ninth Annual Ground Target Modeling & Validation Conference*, Vol. 9, 76-85, August 1998.

***Correspondence**: Email: bweber@arl.army.mil, telephone: 301-394-2501; and japenn@arl.army.mil, telephone: 301-394-1039; fax: 301-394-5234