Proceedings of the IRIS Passive Sensors
Symposium, Monterey, CA, March 1996

MODELING PARAMETERS FOR TARGET IDENTIFICATION: A CRITICAL FEATURES ANALYSIS

March 1996

Barbara L. O'Kane, Ph.D.
U.S. Army CECOM RDEC NVESD
Ft. Belvoir, VA 22060-5806

Irving Biederman
University of Southern California
Los Angeles, CA

Eric E. Cooper
Iowa State University
Ames, IA

ABSTRACT

For decades, the modeling and prediction of human target identification has relied on physical parameters generalized over the whole target, such as physical size, range to the sensor, and apparent thermal contrast, defined as six cycles on the target in the ACQUIRE model. Identification performance for targets that meet this criteria are on average accurate. However, variation in the identifiability of objects meeting the ACQUIRE criterion is so wide that it suggests that some other factor, something bound up in the way people perceive and identify objects, is also influencing identification performance. The evidence for this is that some targets are easier to identify than the model would predict, while others are much more difficult; some aspect angles are more difficult, while others are more readily identified. Many perception experiments which had been performed for sensor design yielded general results of value to system design parameters but resistant variance remained in the results to baffle those demanding definitive results for specific target configurations. For the last few years, NVESD has embarked upon a strategy of understanding the human perception of thermal imagery from the standpoint of neuroscience theory, the most prominent of which are Recognition-by-Components (Biederman, 1987) and computational vision (e.g., Wilson, 1995; Fiser, et al., 1995; Lades et al., 1993).

The representation of objects in terms of an arrangement of viewpoint-invariant parts (or geons) has received considerable support from psychophysical experiments. Do these representations actually predict target identification performance in the real-world military environment under difficult conditions of identification? Two analyses were performed on the correct identification and confusions among targets result-

ing from a perception experiment conducted at NVESD in which trained military observers attempted to identify vehicles in infrared imagery. A similarity measure for all pairs of vehicles was derived from a contingency tree that expressed the similarity of the vehicle's parts according to the saliency of the parts and the degree to which viewpoint-invariant properties could be employed to distinguish among them. The confusion rate between a pair of vehicles was strongly correlated (.97) with a negative exponential function of the nodal distance (proximity of vehicles in a contingency tree) between these vehicles in the similarity trees.

In a further confirmation for the critical role of a parts-based representation, a different analysis showed that virtually all the effect of the traditional modeling variables (range, size, thermal contrast) could be predicted on the basis of whether the two most diagnostic parts were identifiable in the image. Prediction of part detection could therefore account for more variance in accuracy than global parameters.

1.0 BACKGROUND

The present range performance model (ACQUIRE) was developed by NVESD and is used by government and industry to evaluate thermal sensors and in wargames. The model predicts the range at which targets can be detected, recognized and identified. The main parameters of the model are target size (square root of the projected area), an average temperature deviation between the target and background, range, and sensor Minimum Resolvable Temperature Difference (MRTD) (Scott, 1992). There are two conditions of the present models which this research effort is addressing: 1) The model cannot predict which targets will be confused with each other; and 2) Some targets and viewing angles require more resolution than others for correct identification.

The ACQUIRE modeling methodology is applicable for an ensemble of targets but is not always accurate for particular targets due to the fact that some vehicles are relatively easy to identify while others are more difficult. It was hypothesized that some targets are less confusable with others because they may possess specific, easily detectable features which the traditional modeling methodology does not take into account. To determine whether a feature-based identification model for predicting observer performance could be developed, the Department of Army Model Improvement Program (AMIP) sponsored an experiment which was performed at NVESD in 1993. The results of this experiment formed the database for testing a similarity model to explain target confusions and predictions of correct identification based upon critical features.

The most prominent theory of human object identification states that there are specific geometric primitives (features) in an object that the human uses for identification. Biederman (1987) calls these primitives "geons". These geons are recognizable no matter what the observers' viewpoint on them, thereby allowing identification of objects at any orientation. However, empirical data upon which this theory has been developed and verified have been line drawings. The question at issue was whether this theory would also apply to grey scale, less highly resolved images from thermal systems.

1.1 RECOGNITION-BY-COMPONENTS (RBC) THEORY

The fundamental assumption underlying the RBC theory is that objects are represented as an arrangement of parts that are drawn from a vocabulary of simple shapes (geons) distinguishable by viewpoint-invariant properties. There are several advantages of parts-based representations over global shape template models (e.g., Poggio & Edelman, 1990; Lades, et al., 1993). Global shape can vary dramatically when an object is rotated in depth, has lost a part or parts, or is partially occluded. However, parts-based representations often degrade only moderately, if at all (Biederman and Cooper, 1991; Biederman and Gerhardstein, 1993). In the case of grey-scale thermal imagery, the parts-based representation is based on contrast patterns from parts of the vehicle in a more or less blurry, somewhat ambiguous image. Identifying a variation of a thermal image as a particular tank is perhaps one of the most stringent tests of a parts-based representation. This analysis attempted to determine whether observers used object parts in performing identification.

The investigation was concerned with whether an extension of the RBC model could account for discriminations among different models of tactical vehicles. At least three different perceptual bases for distinguishing among different subordinate level classes have been proposed which might be governed by very different representations (Biederman & Gerhardstein, 1993, 1995). The difference might be in 1) the shape of the largest part, 2) a relatively small part which has viewpoint invariant properties or 3) fundamentally metric differences. This latter is perhaps the most difficult case for humans to distinguish. Viewpoint invariant differences are much more readily discriminated than are metric differences, which can be produced by rotation in depth or foreshortening (Cooper, Biederman, & Hummel, 1992).

2.0 THE EXPERIMENT

A perception experiment was conducted at NVESD in which thermal images of targets at three aspect angles were presented in random order to military observers. A form of hybrid imagery was produced based upon the actual thermal signatures but manipulated with simulations to produce target signatures which might potentially occur. First, the targets were segmented manually from the background. Then, specific features were electronically "painted" to be either hot or at the temperature of the background (not visible). A range simulation was then performed on the target after which it was re-embedded in a background. Finally, sensor, atmospheric and noise simulations were added. Observers first attempted identification on all imagery and then reviewed each image a second time declaring which parts of the object were visible.

2.1 IMAGERY

Field imagery collected with a calibrated thermal system in the 8-12 micron spectral band was used in the experiment. The vehicles and their descriptions are given in Table 1. Each vehicle was positioned at three orientations (front, left front oblique, and right rear oblique). The targets were manually segmented from the background and electronically manipulated by "painting" the features to blend in with the background according to a matrix (Table 2). Then each image was processed with a sensor, range, and atmospheric simulation (Horger, 1990) to produce imagery which corresponded to the ranges and transmissions shown in Table 3. A nominal Tank Thermal Sight (TTS) sensor simulation was used for the sensor simulation. Real-time noise was created using 30 Hz frame rate with 30 noise frames added to the sensor simulation. These conditions were predicted by the ACQUIRE model to produce identification performance in the range of 20% to 75% (+/- 20%), which would include the entire range of possible performance (Table 3). An outline of the imagery generation procedure is shown in Figure 1.

2.2 PRE-TRAINING

Observers completed a self-scoring, software training package developed at NVESD with E-OIR Measurements, Inc. Images were shown in training without reference to features as cues. The trainee was required to reach a criterion level of above 90% on identification of the training set prior to participation in the experiment. Training time varied depending upon the individual and ranged from about two to four hours. Nine training images were created for familiarizing observers with the test procedures and response menus.

2.3 RESPONSE METHODOLOGY

Observers were seated before an individual PC computer and display (EO_Vision, 486DX) which displayed each scenario in random order. Subjects reviewed and responded to the complete set of imagery twice, the first time with an identification, and the second time through with a list of the parts of the vehicle which they could detect in the image. For the identification portion, observers were provided with a response menu containing the names of the vehicles and a "Don't Know" option. For the parts detection option, observers were given a list of features (gun, turret, hull, track, roadwheel, and engine) from which to select responses.

2.4 SUBJECTS

The observers were 30 tank crewmen with extensive experience in the use of thermal sights.

2.5 DESIGN

The design was a 9 (vehicles) X 2 (Range: Near = 960 m and Far = 1500 m) X 2 (atmospheric transmissions 30 and 90% for the first km) X 3 (aspects: front, left front oblique, and right rear oblique) matrix. This set was sufficient to convey all of the vehicles' major features. For the case where none of the features were deleted, the total number of trials was 9 X 2 X 2 X 3 = 108. Each target was then in ten other scenarios in which parts were variously deleted yielding a total of 1,080 trials X 30 subjects = 32,400 trials.

Table 1: Targets (APC- Armored Personnel Carrier; MBT-Main Battle Tank; RV-Reconnaissance Vehicle; FSU- Former Soviet Union).

Vehicle	Recognition Category	Tracked/ Wheeled	Country of Origin
M113	APC	Tracked	US
M3	APC	Tracked	US
Marder	APC	Tracked	German
Spahanzer Luchs	RV	Wheeled	German
Leopard1	MBT	Tracked	German

Table 1: Targets (APC- Armored Personnel Carrier; MBT-Main Battle Tank; RV-Reconnaissance Vehicle; FSU- Former Soviet Union).

Vehicle	Recognition Category	Tracked/ Wheeled	Country of Origin
T62	MBT	Tracked	FSU
BTR-60 PB	APC	Wheeled	FSU
BRDM-2	APC	Wheeled	FSU
BMP-2	APC	Tracked	FSU

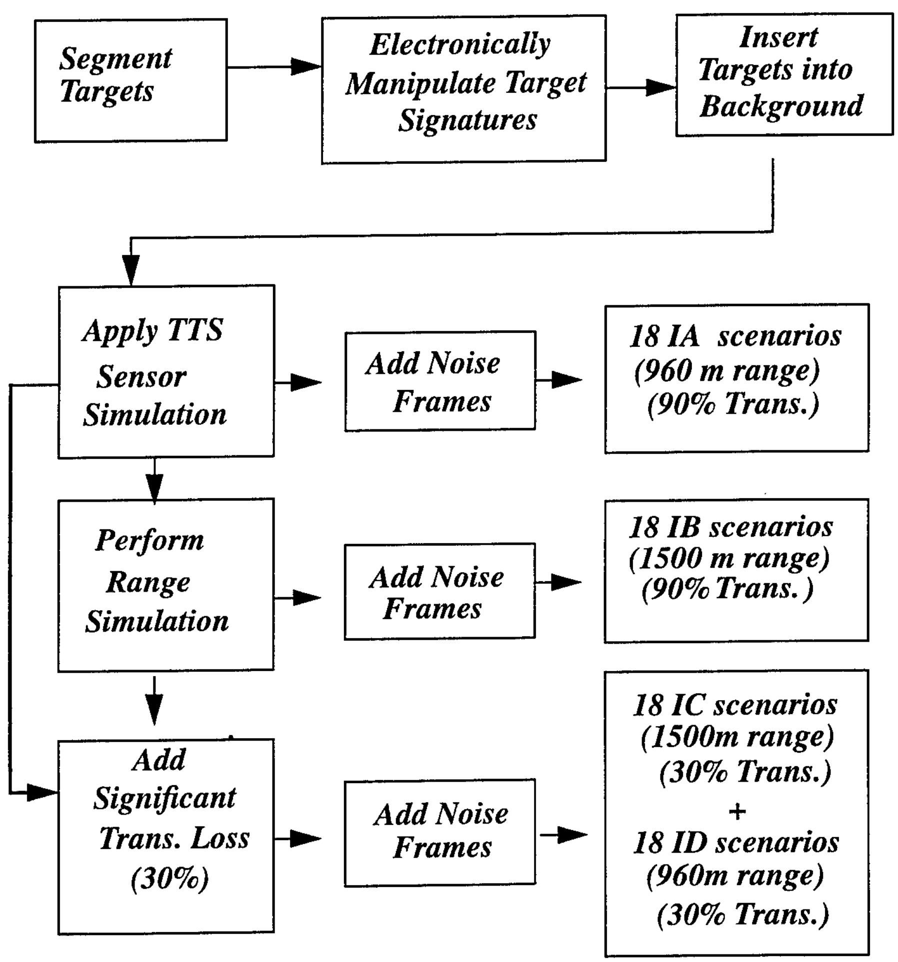
Table 2: Target Feature Variations

Scenario Label	Gun	Turret	Hull	Track	Road-wheel	Engine
A	Hot	Hot	Hot	Hot	Hot	Hot
B		Hot	Hot	Hot	Hot	Hot
C			Hot	Hot	Hot	Hot
D				Hot	Hot	Hot
E					Hot	Hot
F	Hot				Hot	Hot
G						Hot
H	Hot					Hot
I	Hot					
J		Hot				
K			Hot			
L				Hot		
M					Hot	
N			Hot	Hot		
O		Hot			Hot	
P		Hot	Hot	Hot		
Q	Hot	Hot	Hot			
R	Hot	Hot	Hot	Hot		

Table 3: Matrix of Ranges and Transmission for each Experimental Cell. Predictions are based upon the ACQUIRE model and for Scenario A.

Cell Name	Range	Atmospheric Transmission	Pred Probability of ID
IA	960m	90%	75%
IB	1500m	90%	38%
IC	1500m	30%	20%
ID	960m	30%	65%

Figure 1. Procedure for generating hybrid imagery.



3.0 SIMILARITY AMONG VEHICLES

When asked to distinguish among basic level classes (dog from elephant), people list parts as the primary basis for characterizing the classes (Tversky & Hemenway, 1984). The Hummel & Biederman (1992) model provides a basis with which to establish the similarity of pairs of objects in terms of the overlap of a structural description specifying an arrangement of *geon feature assemblies* (GFA). A GFA represents a particular geon (its cross section and axis shape [straight or curved], the parallelism of its sides, and whether it is truncated), the geon's attributes (coarse orientation [vertical, horizontal, or oblique] and coarse aspect ratio), and its relations to other geons (e.g., above, larger-than, perpendicular to). A set of GFAs corresponds to an object's structural description.

In the absence of an automatic means for extraction of viewpoint-invariant parts, the GFAs must be identified by human judgment. GFAs are available to human consciousness and descriptions of them are readily verbalized and employed in identifying objects (Biederman & Shiffrar, 1988; Biederman & Gerhardstein, 1993).

However, features are not processed in a linear fashion. Mental representations of contingencies cause some features to be part of the distinguishing characteristics of objects only in certain cases. For example, when judging which type of airplane is being observed, one would probably look at whether the plane has propellers or jets. If one were trying to tell the difference among cars, one might look at the logo for a rapidly recognized viewpoint invariant distinguishing feature.

In that study, subjects were required to use a contingency structure, noting, for example, the size when the stimulus was red or the orientation when the stimulus was green. Subjects would not process the orientation when the stimulus was red or the size when the stimulus was green. This effect was particularly evident when the head of the contingency was more discriminable than the secondary dimension, apparently so that subjects could avoid making the more difficult discriminations.

Gibson (1947) strongly advocated the inclusion of such an organizational scheme in the training of aircraft identification:

> "First an analysis should be made of the *identifying* (i.e., distinctive) *features* of the list of aircraft to be learned -- not simply of the *descriptive* features... Second, these distinguishing features should be used to make a classification of the aircraft by shapes, such as would permit a conceptual organization of their similarities and differences." (p. 152).
> "Members of a group of four-engined planes are more similar to each other than they are to any other planes on the total list. Among this group of four-engined planes some members are more apt to be confused than others. The B-17 is more like the C-54, which has a single tail, than it is like the B-24 which has twin tails." (p.153).

Discrimination among the three twin-tailed planes, for example, was based on whether the tail was oval vertical (B-24), pear-shaped (Lancaster), or triangular vertical (Halifax).

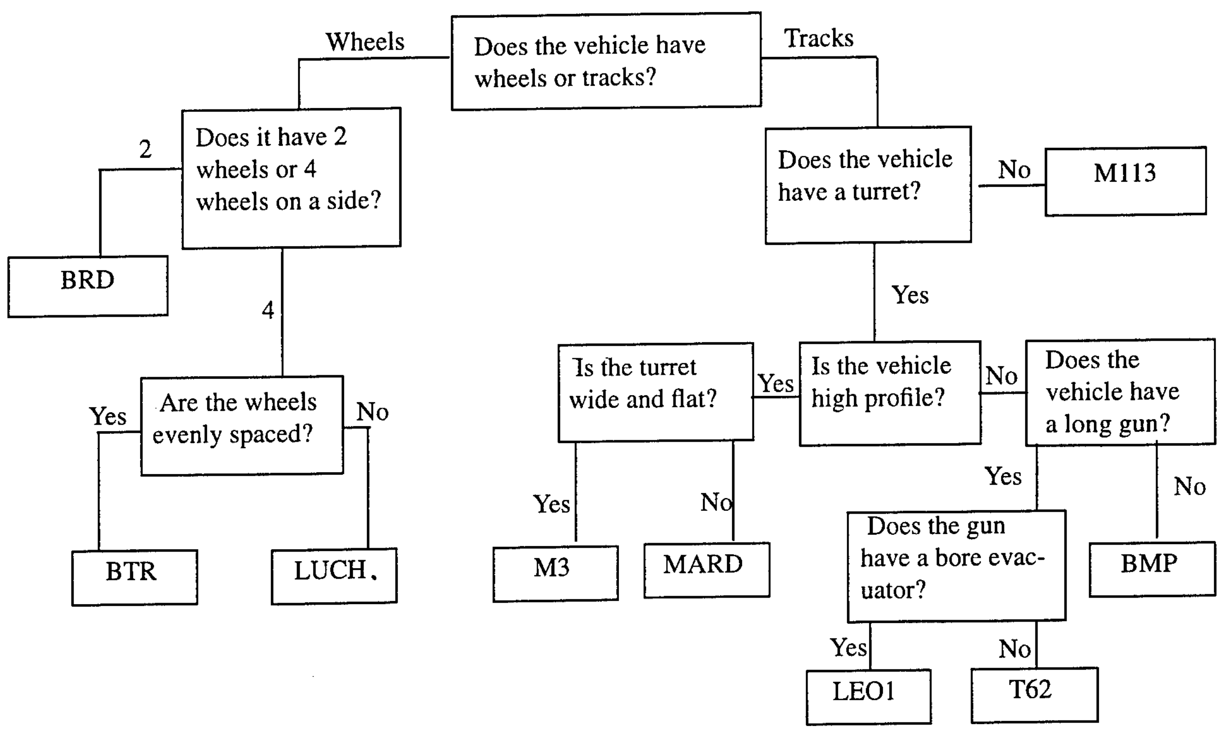
The similarity determination of the present investigation was represented by a contingency tree that, in retrospect, is reminiscent of Gibson's aircraft tree. In describing the tree, rather than refer to what would be a random structural description of the different parts, the tree will be labeled in terms of the common names of the parts, as was done by Gibson. Thus the gun will be referred to as a "gun" rather than an elongated, horizontal cylinder side-projecting, and end-to-end with a half cone, wedge, or brick.

3.1 SIMILARITY TREE

An inverted similarity tree (with the "trunk" on the top) is shown in Figure 2 as constructed by examining near range (30 m), high contrast thermal imagery similar to that used in the study portion of the training for the subjects. Salient, easily detectable features which could distinguish the vehicles from one another were noted (e.g., whether the vehicle had wheels or tracks, whether the vehicle had a gun, whether the vehicle had a cargo bed). The trees were constructed without any knowledge of the confusion probabilities among the vehicles.

The most salient features, wheeled or tracked, divide the target set into two parts with three wheeled vehicles and six tracked. Whether the tracked vehicle has a turret or not separates out one target, and as the flow chart progresses, the distinguishing parts become more metric oriented (long or short gun, high or low profile) until a final decision is made based upon the bore evacuator on the gun, a small feature located in a particular location. The use of contingencies would be especially valuable to quick identification in allowing observer to avoid making difficult discriminations of details on every trial. Naturally, these distinguishing features are not by any means the only set which could be surmised, but represents one possible scheme for a similarity tree.

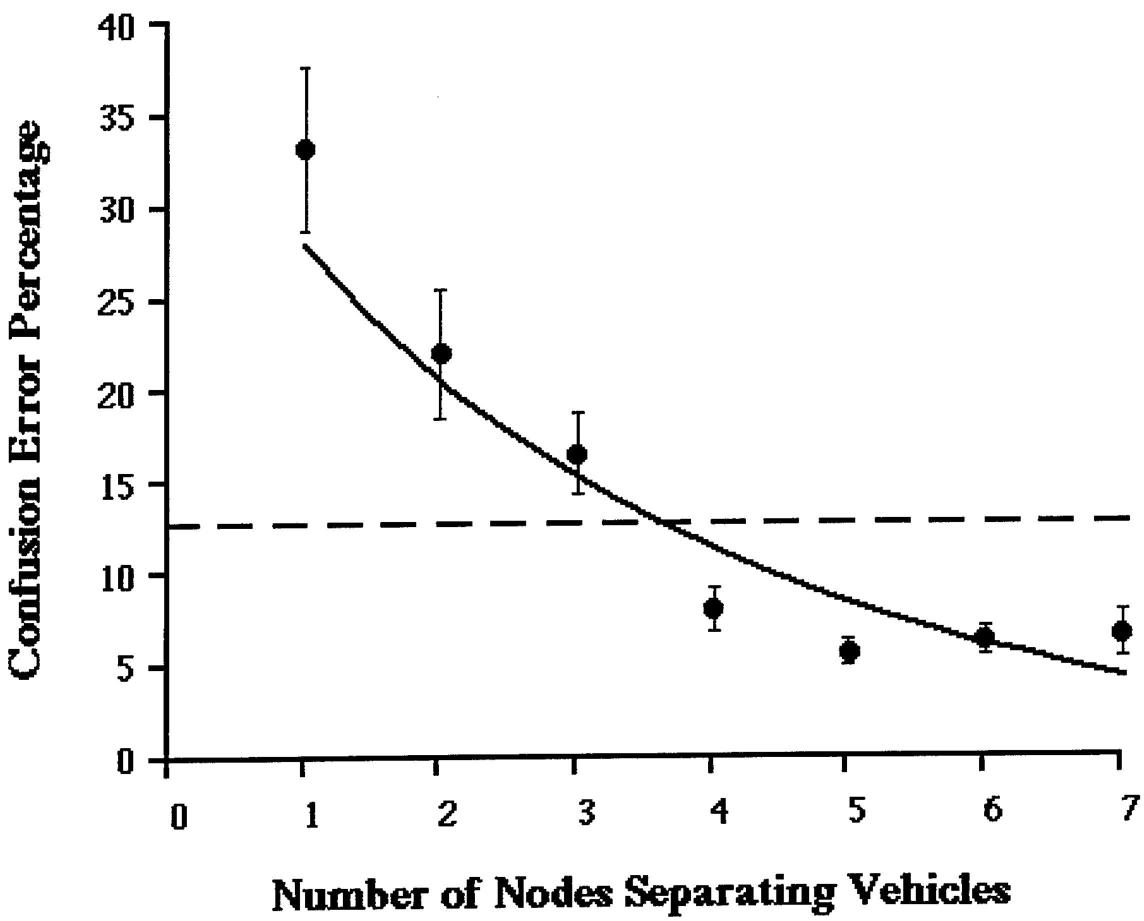
Figure 2. Similarity Decision Tree



3.2 RESULTS - CONFUSIONS AMONG VEHICLES

The first analysis of the results described here will focus on the confusions among the vehicles. Each vehicle pair was designated as a number of nodes apart in the flow chart, determined by the number of questions intervening. The maximum number for this set of targets was seven, as between those at either pole (for example, the Luchs vs the Leopard, or the BTR vs the T62). Each pair of vehicles was designated the number of nodes apart and then the average confusion for each nodal distance was plotted in Figure 3. There was an extremely high correlation between a negative exponential of the nodal distances (the solid function) and the confusion error rates: 0.974. Chance value is shown by the dashed line.

Figure 3. Nodal Distance chart



4.0 CRITICAL FEATURES ANALYSIS[1]

As discussed above, current models to predict target identification (e.g., NVESD's ACQUIRE model) incorporate global aspects of the stimulus, such as target size, range, and thermal contrast. In contrast, recent theoretical and empirical work in the psychology of object identification has emphasized the role of object parts in identification as in the RBC theory (Biederman, 1987). If, as these theories predict, detection of critical parts is essential to object identification, then prediction models for observer identification of military vehicles might be greatly enhanced by including information about whether a vehicle's critical parts are visible in the stimulus. The purpose of this second analysis was to determine if the prediction of identification performance of military vehicles in thermal imagery could be significantly enhanced by including critical part information in the prediction model.

The purpose of this second analysis was to attempt to predict the identification probabilities for vehicles in the experiment described above using both the global features that are currently employed to predict identification as well as the information about the visibility of the critical features of the vehicles. If the visibility of the critical parts determined identification performance, then such information should be included in future modeling enhancements.

The first determination to be made was whether there were parts that were associated with identification of particular vehicles. Some questions asked were: 1) Was there a good predictive value in knowing whether specific parts were visible in the image? 2) Were the critical parts the same for all vehicles? 3) Did knowing whether the parts were visible add to the prediction accuracy beyond knowing range, thermal contrast, and target size?

4.1 CRITICAL FEATURE DETERMINATION

A skilled analyst (Dr. Cooper) attempted to identify the vehicles in all 1080 scenarios. For each scenario, the analyst would record his attempted identification response, his confidence in the guess on a scale from 1-10 (with 1 being completely unsure and 10 being absolutely sure), the features of the vehicles which were illuminated, those features which were not illuminated but visible nevertheless. On those trials in which the analyst was fairly confident a correct identification had been made, the feature that was key for identification was noted as well. After performing this procedure for all 1080 scenarios (presented in random order, just as for the observers in the experiment proper), for those trials in which a correct identification was made, the number of times each feature (gun, turret, hull, tracks, roadwheels, engine) was considered the critical feature for identification was tallied. The tallying was done separately for each of the three aspects for every vehicle. This tallying led to the following list of critical features shown in Table 4.

It is possible, of course, that the features which were used by the skilled analyst were not the same as those used by the observers in the experiment. In order to better determine which features might actually have been used by the observers in the experiment, the critical features for identification were determined again in a slightly different way. For each of the three aspects of every vehicle, the correlation between the percentage of observers who detected each of the six features in a scenario and the percentage of observers

---

1. The critical features analysis was performed by Dr. Eric Cooper under an Army Research Office contract DAAl03-91-C-0034 sponsored by the U.S. Army CECOM RDEC Night Vision & Electronic Sensors Directorate. The discussion is taken partly from the unpublished final report.

Table 4: First Critical Features as Determined by A Skilled Analyst[a]

ASPECT			
Vehicle	**Front Oblique**	**Front**	**Rear Oblique**
BMP	Turret (Hull)	Turret	Engine (Hull)
BRDM	Wheels	Hull (Wheels)	Wheels
BTR	Wheels	Hull (Wheels)	Wheels
LEOPARD1	Gun (Hull)	Turret	Engine
LUCHS	Wheels	Turret	Wheels
M113	Hull	Hull (Tracks)	Hull
M3	Turret	Hull (Engine)	Turret
MARDER	Turret	Turret	Turret
T62	Gun	Turret	Gun

a. The observers' first critical features were also determined and when different from the skilled analyst are shown in parentheses.

who correctly identified the vehicle in a scenario were computed. Such an analysis was performed to determine which features' presence best predicted identification performance at each aspect. This method of determining the critical features showed fairly high (but not perfect) agreement with the skilled analyst method. In 20 out of 27 cases (76%) the same critical feature was revealed by the two methods. Potential reasons for this disparity could be that there was more than one critical feature (as will be shown below) that contributed to the identification, that the skilled analyst was more highly trained than the observers and able to detect more subtle features, or that there is a significant amount of observer dependent variance in which particular feature is most critical for any vehicle.

4.2 MODELING

After collection of the data was completed, the identification performance was modeled with linear regression using both the global variables (range, transmission, size and aspect) and the critical features variables (both as determined by the skilled analyst and as determined by the military observers in the experiment). The variables used in all the models are summarized below in Table 5.

Table 5: List of Independent Variables Used in the Regression Models

Ind. Var. Name	Values
Range	960m and 1500m
Size	Square root of the vehicle's visible area
Transmission %	30% and 90%
FO	Dummy variable 1 if aspect is Front oblique, 0 otherwise
FR	Dummy variable 1 if aspect is front, 0 otherwise
Critical Feature 1 (Skilled Analyst)	% of observers who detected the most critical feature (as determined by the skilled analyst) in a particular scenario
Critical Feature 2 (Skilled Analyst)	% of observers who detected the second most critical feature (as determined by the skilled analyst) in a particular scenario
Critical Feature 1 (Military Observers)	% of observers who detected the most critical feature (as determined by the military observers) in a particular scenario
Critical Feature 2 (Military Observers)	% of observers who detected the second most critical feature (as determined by the military observers) in a particular scenario
Random Critical Feature	% of observers who detected a feature that had been chosen randomly from the set of possible features

In all models, the dependent variable is the percentage of observers who correctly identified the target object in a particular scenario (trial). Each trial (consisting of one target at a certain range, aspect, environmental transmission, and feature manipulation) served as a single case in each of the regression models. A summary of the results are shown in Table 6. Each run of the regression model shows the variables put into the equation and the resulting adjusted R^2, which is the proportion of the variance accounted for by the model corrected for the number of predictors employed. The adjusted R^2 provides an unbiased estimator of the population R^2, and thus is the proper value to use when the proportion of variance accounted for by each model is reported.

An Analysis of Variance (ANOVA) was also run on each of the predictor models and it was found that the ACQUIRE model (Model 1 in Table 6) has reliable predictive value ($F(3.1076) = 102.277$, $p<.0001$) and accounts for 19% of the variance in the data. Reliable effects of all of the independent variables (range, size and transmission) were found. Whereas it does predict much better than chance, there is still a great deal of variance to be accounted for by other variables, such as those proposed.

Table 6: Variance accounted for (R^2) by each set of predictors.

Model Ref.	Predictors	R^2 (adj.)
1	Range, Size, Transmission	**.189**
2	Range, Size, Transmission, Aspect	**.250**
3	SA's 1st Critical Feature	**.446**
4	SA's 1st and 2nd Critical Feature	**.544**
5	SA's 1st Critical Feature, Range, Size, Transmission, Aspect	**.565**
6	SA's 1st and 2nd Critical Feature, Range, Size, Transmission, Aspect	**.627**
7	O's 1st Critical Feature	**.643**
8	O's 1st and 2nd Critical Feature	**.663**
9	O's 1st Critical Feature, Range, Size, Transmission, Aspect	**.689**
10	O's 1st and 2nd Critical Feature, Range, Size, Transmission, Aspect	**.703**
11	Random Critical Feature	**.133**
12	Random Critical Feature, Range, Size, Transmission, Aspect	**.306**

Adding aspect (front, front oblique, rear oblique) to the equation adds a small, but statistically greater amount of predictive power (Model 2). Size correlates strongly with aspect (the front aspects have small size values and the two oblique aspects have large size values), and front aspect vehicles are particularly difficult to recognize, even taking into account their size. When adding aspect to the model, size drops out as a statistically reliable predictor of recognition performance.

The third model in Table 6 predicts identification performance based only on the percentage of observers who detected the most important vehicle feature as determined by the skilled analyst. Compared to Model 2, the results are very striking. The percentage of observers detecting the critical feature when used alone as a predictor accounts for far more variance (44.6%) than all of the global variables put together (25%). (Naturally, the ACQUIRE model could theoretically be used to predict whether the observers would be able to detect the feature, if the feature size is taken into account.) If the percentage of observers who detected the second most important critical feature (as determined by the skilled analyst) is added to the above equation, the adjusted R^2 increases to .544 (Model 4).

When the model incorporated all the global variables as well as the percentage of observers who detected the critical features, the former variables added a small, but statistically significant, amount to the predictive power (Model 6) - 44.6% with critical feature alone, 56.5% with global information included as well. This may not be surprising. In some instances, observers may have been able to detect whether the critical feature was present or not, but unable (because of small target size, distant range, or poor signal to noise ratio) to see the critical feature clearly enough to discern its unique properties. For example, the MARDER has a unique turret (relative to the other vehicles in the set) which is wider at the top than at the bottom. In some scenarios, it is possible to see that the vehicle has a turret, but not possible to determine whether the turret bears the distinctive MARDER shape. The global variables might thus be expected to add some predictability to the equation because factors causing poor visibility (small size, long range, poor transmission) may hamper identification of a part's unique features even when the part can be detected.

In the above critical feature models, the feature critical for identification was determined by a skilled analyst who examined the set of vehicles carefully and determined which feature was most useful in distinguishing a particular vehicle. It is, as stated above, of course, possible that the skilled analyst was not using the same critical features as were the observers in the experiment. The purpose of the next model was to determine the upper limit of the amount of variance in identification performance that can be predicted by the percentage of observers who could detect a particular critical feature for each vehicle and aspect. In order to accomplish this, the feature that would be the best predictor of identification performance was determined by computing the correlations (for each vehicle at each of the three aspects) between the percentage of observers who identified a target in a given scenario, and the number who detected each of the vehicle features in that scenario. The feature with the highest resulting correlation will thus be considered the best predictor of identification performance.

This method of determining critical features must, necessarily, result in better predictive power than the skilled analyst method of determining critical features. The critical features chosen using the experimental observers' data are chosen using the criteria that they *are* the best predictors of identification performance.

Note from Table 6 that this model (Model 7) is the best predictor of all those which were so far tried. Using only the detection rate of one critical feature (as determined by the observers), 64.3% of the variance in identification accuracy can be predicted. Thus with a single independent variable, almost three times as much of the variability in the identification data is accounted for than by the three global variables used in the first model. If the second best critical feature (as determined by the observers) is added to the above model, the percentage of variance accounted for increases slightly to 66.3% (Model 8).

In Model 10, the purpose was to determine the maximum amount of variance in the identification data that can be accounted for by all the independent variables that have been considered. Every independent variable except size added a statistically significant amount to the predictive power of the model. (The size variable may have been less important in this experiment than in others, due to the fact that all of the vehicles were approximately the same physical size.) The additional size variables in this model do account for reliably more variance than the single variable model above, however, the absolute gain is fairly small (6% of the variance).

The above models strongly suggest that including feature detection information in the models accounts for the majority of the variance in the data. However, what all of these models fail to establish is whether it is necessary that the features used be critical features. That is, certain features of the vehicles make the vehicles distinctive relative to others in the set, and the assumption that has been made is that detection of certain features is more diagnostic of identification than others. However, it is possible that the detection of any feature (distinctive or not) might be all that is necessary for good prediction of identification performance.

Model 11 shows the prediction based upon a random feature chosen for each of the three aspects of each of the nine vehicles in the experiment. Some of the features chosen were, by chance, the critical features used in the previous models, and some were not. The percentage of observers who detected the random critical feature in a particular scenario was then used as the predictor for the percentage of observers who correctly identified the vehicle in that scenario. While detection of the random critical feature did predict a reliable amount of variance in the data (13.3%), it predicted significantly less variance than either the critical feature as determined by the skilled observer (44.6%, $Z=9.84$, $p<.01$) or as determined by the observers in the experiment (64.3%, $Z=16.73$, $p<.01$). The results suggest that detection of "any old" feature is not sufficient for identification. Certain distinctive features are present for each target, and it is detection of those features which is critical for predicting identification performance. Thus, in order to realize the predictive gains which accompany including feature detection information in the model, some method of determining which features of a vehicle are distinctive must be included.

5.0 SUMMARY

The results of the above analysis of a perception study and the modeling comparisons are quite clear. The first analysis showed that a contingency tree can be used as a basis for a similarity measure of vehicles and predict the average confusability of targets, even when the tree is very general, based upon one particular view, and applied to all views. This type of tree can be very useful in training.

The second analysis showed that models which included detection of a vehicle's critical feature(s) predicted identification performance far better than models that included only global variables. Indeed, critical feature information alone predicted performance far better than all of the global variables taken together. Further, this predictive power did not extend to detection of every vehicle feature. Detection of features which are distinctive to a vehicle predicts identification performance much better than a randomly selected feature, which is even less effective than the global variables.

6.0 CONCLUSIONS

The results of this study and modeling analysis have potential implications for automatic target recognition system development, war game predictions, and military observer identification training. Specifically, the results suggest that performance of automatic vehicle recognition systems might be greatly improved if, rather than being based on matching global properties of the stimulus (e.g., global shape or overall spatial frequency analysis), such systems were designed to detect and identify distinguishing parts of vehicles as human observers appear to do.

Further, in current war game simulations, it is necessary to be able to determine in a particular set of environmental circumstances whether an individual observer can detect and identify a particular target. Virtually all tactical decisions in the simulation must start with this piece of information. Thus, accuracy in predicting whether the observer can identify a target is vital to accurate war game simulations. The data here suggest that including information about whether critical features are detectable can greatly enhance the accuracy of predictions about whether observers can identify vehicles, and the incorporation of such information could potentially make simulations more accurate. Unless the critical features are known for a particular set of vehicles, predictions can only be made after the fact. However, after some work with this type of feature analysis, it is hoped that the critical features will become known for analysis, prediction, and training for a defined set of representative vehicles.

Finally, the data presented here have implications for how we teach soldiers to identify military vehicles. The results presented are suggestive (though not conclusive) that observers distinguish vehicles by looking for a distinctive critical feature rather than by using more global properties of a stimulus. Training performance in vehicle identification might possibly be enhanced, therefore, by instructing soldiers as to the most distinctive and easily detectable features in a particular identification set.

7.0 ACKNOWLEDGEMENTS

The authors are grateful to Luanne Obert and Dr. Rudolf Buser for their encouragement and helpful comments on this research effort. The technical support of Kathy Wolf is greatly appreciated. Beth Nystrom performed the difficult task of image preparation. Amy Settle and Angie Moulton contributed greatly to the preparation of the manuscript.

8.0 REFERENCES

1. Biederman, I. (1987) Recognition-by-Components: A theory of human image understanding. *Psychological Review,* **94**, 115-147.
2. Biederman, I., & Cooper, E. E. (1991) Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology,* **23**, 393-419.
3. Biederman, I., & Gerhardstein, P. (1993) Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology, L Human Perception and Performance*, **19**, 1162-1182.
4. Biederman, I., & Shiffrar, M. (1987) Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual learning task. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition*, **13**, 640-645.
5. Cooper, E. E., Biederman, I., & Hummel, J.E. (1992) Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology,* **46**, 191-214.
6. Fiser, J., Biederman, I., & Cooper, E. E. (1995) Test of a two-layer network as amodel of human entry-level object recognition. In J.M. Bower, ed. *The Neurobiology of Computation: Proceedings of the Third Annual Computational Neuroscience Meeting*. Boston: Kluwer.
7. Gibson, J.J. (1947) *Motion Picture Testing and Research.* Army Air Forces Aviation Psychology Program Research Report No. 7. Washington, D.C.: U.S. Government Printing Office.
8. Horger, J.D. (1990) Image generation for perception testing using computer FLIR simulation. *Proceedings of the SPIE International Symposium on Optical Engineering and Photonics in Aerospace Engineering*, **1309**, 181-189.
9. Lades, M., Vorbruggrn, J.D., Buhmann, J., Lange, J., von der Marlsburg, D., Wurtz, R.P., & Konen, W. (1993) Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, **42**, 300-311.
10. Poggio, T., & Edelman, S. (1990) A network that learns to recognize 3-D objects. *Nature*, **343**, 263-266.
11. Scott, L., & Tomkinson, D. (1992) ACQUIRE Range Performance Model. Night Vision & Electronic Sensors Directorate Report.

12. Tversky, B. & Hemeway, K. (1984) Objects, parts, and categories. *Journal of Experimental Psychology: General*, **113**, 169-193.

13. Wilson, H. (1995) Quantitative models for pattern detection and discrimination. In E. Peli, ed. *Vision Models for Target Detection and Recognition.* New Jersey: World Scientific Pub.