AD-A226 905

Feature Set Evaluation for Classifiers

Progress Report # 2

January 20, 1989

period: November 16, 1988 to January 15, 1989
contract #: DAAB07-88-C-F415
security classification: unclassified
CDRL #: A001

for: *Director*
Center for Night Vision & Electro Optics
ATTN: AMSEL - RD - NV - AC
(Martin Lahart)
Fort Belvoir, VA 22060 - 5677

by: *KAB LABORATORIES INC.*
3116 Mercer Lane
San Diego, CA 92122
(619) 455 - 5361

TABLE OF CONTENTS

I. INTRODUCTION

A. Background

On September 13, 1988 KAB LABORATORIES INC. (KAB) was awarded a Small Business Innovation Research (SBIR), Phase I contract with the Center for Night Vision & Electro Optics (CNVEO). The principal investigator for this research activity is John Konotchick of KAB, and the technical project manager for the work is Martin Lahart of CNVEO. Work on the contract commenced on September 15, 1988. The Phase I activity is to conduct research aimed at developing techniques for improving automatic recognition/classification of targets. This is the second progress report under that activity, and covers the second two months of a six month contract.

B. Objectives

Automatic Target Recognizers (ATRs) have tried a wide variety of feature set classifiers in attempting to improve the quality of their classification of targets. The selection of these feature set classifiers to date has largely been based upon subjective intuition of the analyst. The analyst typically approaches the problem by starting with a proposed feature set which is derived somewhat heuristically based on an analyst's understanding of the underlying physical phenomena which differentiate a target from any background "clutter" or "noise" which may exist. This underlying phenomenology can be exceedingly complex in the case of real military targets, in real clutter filled backgrounds, imaged by electro-optical sensors under the less-than-ideal circumstances which may exist in a battle field environment.

The feature set for ATR applications could easily contain a large number of individual features or measurements (e.g., location of hot spots, geometric ratios, areas, perimeters, texture mixture, etc.). For real time systems, these features must be extracted quickly and processed to determine the target identification (classification). To minimize computations and keep ATR processor requirements at a reasonable level, the ATR algorithms should be efficient and extract only those features which are most useful to the identification process. The selection of this set of reduced features which possess the most powerful discriminating capability is the subject of this study.

KAB has proposed to use an existing software package, developed by PAR Government Systems Corporation (PGSC), called the On-Line Pattern Analysis and Recognition System (OLPARS) as a tool for feature set analysis. Under this contract CNVEO will be furnished with an OLPARS licence, software, and documentation. The OLPARS will also be enhanced by our research to include a new promising feature set evaluation algorithm aimed at meeting specific CNVEO needs.

The Phase I SBIR activity has proposed meeting the following five technical objectives:

1. identify and propose a collection of feature set evaluation algorithmic tools which address unique characteristics of feature sets used in ATR applications.
2. implement at least one new promising feature set evaluation algorithm in FORTRAN and integrate it within the On-Line Pattern Analysis and Recognition System (OLPARS), which is an existing commercial software system which provides general purpose feature set evaluation and classifier design capabilities.
3. demonstrate the performance of the new feature set evaluation algorithms already within OLPARS using feature sets derived form both real and simulated E/O imagery.
4. provide DoD with a licenced VAX-compatible copy of the augmented OLPARS software package.

5. document the proposed new set of feature set evaluation algorithms and the test results obtained with the newly implemented algorithm within a final technical report.

By using the OLPARS in our research we will be taking advantage of considerable previous work on this subject. The OLPARS was initially developed in the early 1970's as a pattern analysis support tool. Since that time it has been enhanced to increase its capability for analysis and display and to make it user friendly. It also comes with full supporting documentation. Upon completion of the Phase I activities CNVEO would possess an independent capability to analyze, select and test feature sets and to evaluate their relative discriminating power for target classification. This capability should provide a means for both improving and testing their own ATR approaches and for evaluating the approaches suggested by industry.

C. Scope

This report covers the second two months of a six month study. The Phase I activity calls for $25,000 of material cost for the purchase of OLPARS, computer time, and a subcontract to PGSC for 75 man-hours of support on the OLPARS program. The remaining $25,000 is spread over 6 months for KAB manpower to support research, and for incidental costs such as travel.

In the sections which follow the progress and plans of the activity will be reported. Section II. RESULTS, will present the work to date and accomplishments. Section III. STATUS, will summarize manpower expenditures to date and relationships to program milestones. Section IV. PLANS, will present the major planned activities for the remaining two months. Finally, Section V. CONCLUSIONS, will summarize the findings to date.

II. RESULTS

The program schedule given in the SBIR proposal called for completing Task 1 (Feature Set Evaluation) and starting on Task 2 (Algorithm Implementation) in the first two months. KAB has kept to that schedule, as reported in the first progress report. During the second two months (see Figure 2.) the algorithm was to be characterized, implemented, and evaluation of the algorithm was to begin. This section will describe that activity.

The feature set evaluation algorithm chosen for implementation is the Bhattacharyya distance measure. The Bhattacharyya coefficient is defined as $b = \int [p(\mathbf{x}:W_1)p(\mathbf{x}:W_2)]^{1/2}dx$, and the Bhattacharyya distance as[1][2]

$$B = -\ln b \;=\; -\ln \int [p(\mathbf{x}:W_1)p(\mathbf{x}:W_2)]^{1/2}dx,$$

where $p(\mathbf{x}:W_i)$ is the multivariate probability density function when pattern vector $\mathbf{x}$ $(x_1,x_2,\ldots.,x_n)$ belongs to class W_i $(i=1,2)$. If our class density functions are assumed to be Gaussian distributed, i.e.,

$$p(\mathbf{x}:W_i) = [1/[(2\pi)\{\det C_i\}^{1/2}]]\exp{-1/2}[(\mathbf{x}-\mathbf{m}_i)^T\{C_i\}^{-1}(\mathbf{x}-\mathbf{m}_i)],$$

where $\mathbf{m}_i$ is the mean of class i and $\{C_i\}$ is the covariance matrix of class i, then the Bhattacharyya distance between class E and class F will be given by,[1][3][4][5]

$$B = 1/8(\mathbf{m}_E-\mathbf{m}_F)^T\{(C_E + C_F)/2\}^{-1}(\mathbf{m}_E-\mathbf{m}_F) - (1/2)\ln[\det\{(C_E + C_F)/2\}/[\det\{C_E\}^{1/2}\det\{C_F\}^{1/2}]],$$ where

$\det\{C_E\}$ is the determinant of the covariance matrix of class E. This expression for the Bhattacharyya distance can be used to obtain a ranking of various combinations of features,(i.e., where 1, 2, ..., n features are used) for their ability to discriminate between any two classes E and F. The larger the B distance, the better will be our discrimination. It is also possible to use the Bhattacharyya distance measure to obtain a measure of the error expected from our feature selection.

The conditional Bayes error probability for a two class problem is given by,[3]

$$e^*(\mathbf{x}) = \min[P(W_1:\mathbf{x}), P(W_2:\mathbf{x})],$$

where $\mathbf{x}$ = unknown pattern vector, W_i = class (1 or 2), and $P(W_i:\mathbf{x})$ = the a posteriori probability of $\mathbf{x}$ belonging to class W_i.

Using a geometric mean inequality

$$e^*(\mathbf{x}) \leq [P(W_1:\mathbf{x})P(W_2:\mathbf{x})]^{1/2}.$$

Taking the expectation of this yields,

$$E^* = \int e^*(\mathbf{x})p(\mathbf{x})dx \leq \int \{P(W_1:\mathbf{x})P(W_2:\mathbf{x})\}^{1/2}p(\mathbf{x})dx$$

$$\leq [P_1P_2]^{1/2} \int \{p(\mathbf{x}:W_1)p(\mathbf{x}:W_2)\}^{1/2}dx = [P_1P_2]^{1/2}b,$$

where P_1 is the a priori probability of class 1, $p(\mathbf{x}:W_1)$ is the multivariate probability density function (Gaussian in our case) of pattern vector $\mathbf{x}$ given class 1, and
<u>b = the Bhattacharyya coefficient</u> $= \int \{p(\mathbf{x}:W_1)p(\mathbf{x}:W_2)\}^{1/2}dx.$

The expectation can also be written as,

$$E^* \leq \{P_1P_2\}^{1/2} \exp(-B) \quad \text{where}$$

<u>B = Bhattacharyya distance</u> $= -\ln b.$

This gives the upper bound on error. Similar reasoning can derive a lower error bound for the Bhattacharyya distance measure of,

$$(1/2,[1-\{1-4P_1P_2\exp(-2B)\}^{1/2}],$$

so that we can bracket an upper and lower bound on expected error of,[3][4]

$$(1/2)[1-\{1-4P_1P_2\exp(-2B)\}^{1/2}] \le E^* \le [P_1P_2]\exp(-B).$$

This simple error bounding provides one of the advantages of the Bhattacharyya distance measure. Through a simple analytical computation the bounds on average Bayes risk can be determined (or alternatively, 1-E*, the probability of correct classification).

The current Bhattacharyya implementation planned for the CNVEO OLPARS will compute B only for Gaussian distributed classes. It can be worked out for other distributions, and in general it has been worked out for exponential density distributions, (e.g., Poisson, Gaussian, etc.)[1]. The difficulty of reprogramming OLPARS, however, does not make this a good testbed to experiment directly with various algorithms. Our assumption of Gaussian distributions is probably a fair one, however, given the current knowledge of the features to be investigated.

There were some difficulties encountered during this two month period, which caused the implementation of the Bhattacharyya distance measure to be delayed. The OLPARS, while a mature and capable analysis system does not permit easy modification of its software. The system, moreover, is protected by licencing agreements so that configuration management of the software is important. KAB's subcontractor, PGSC, was required under the subcontract to program the Bhattacharyya distance algorithm into their OLPARS. The limited number of individuals with this skill in PGSC, became a problem. Mike Koligman is the PGSC expert on

OLPARS in San Diego, but his demand on other PGSC commitments in November and December made him unavailable for support of this program. The same rationale similarly delayed the reading-in of the CNVEO data into OLPARS. By the end of December the additional commitments of Mike and other PGSC personnel began to abate and work commenced on the Bhattacharyya programming.

At the end of this reporting period (January 15, 1989), PGSC had made significant progress in installing the Bhattacharyya distance measure into OLPARS, but was still not finished. To help insure that this delay can be made up, KAB plans two actions. KAB will evaluate the Bhattacharyya distance measure using a data set currently on the OLPARS, the NASA test data set. This will save further delay in having the data set programmed for reading into OLPARS. KAB will also stand-by, prepared to begin the evaluation as soon as PGSC has completed their programming. With these two actions, KAB is hopeful it can return the program milestone events to the schedule as planned.

As mentioned in the previous progress report, the OLPARS was delivered to the CNVEO three months early. While this had little impact on our schedule planning, it has occasioned some additional support. In early January, Mike Koligman (PGSC) spent a day at CNVEO to provide some OLPARS support to personnel there. It appears that the OLPARS is being used, and CNVEO data sets have been read into the system.

references

[1] Kailath, T., "The Divergence and Bhattacharyya Distance Measures in Signal Selection", IEEE Transactions on Communications Technology, COM-15, 52, February 1967, pp 52-60

[2] Andrews, H.C., Mathematical Techniques in Pattern Recognition, John Wiley & Sons, New York, 1972, pp 37-43

[3] Devijver, P.A. and Kittler, J., Pattern Recognition: A Statistical Approach, Prentice-Hall Intl., London, 1982, pp 57-

58, 97-99

[4] Chen, Chi-hau, Statistical Pattern Recognition, Spartan Books, Rochelle Park, New Jersey, 1973, pp 57-60, 71

[5] Fu, K.S., Applications of Pattern Recognition, CRC Press, Boca Raton, FL, 1981, pp 88

III. STATUS

While striving to meet schedule goals the KAB team has remained flexible and responsive to CNVEO specific desires. The OLPARS is installed and running at CNVEO, well ahead of the original schedule. The Bhattacharyya programming into OLPARS, conversely, has slipped. PGSC should soon be completing this effort. In the remaining two months, the algorithm will be evaluated and the study will be documented. This activity is expected to be completed within the time and labor constraints remaining. Figure 1., on the next page, presents a plot of manpower expenditures for the first four months, overlaid on the total Phase I allotment. It should be noted that some hours for the first two months were not received until month three, when they were entered. Table 1 presents the data in tabular form.

Figure 2., graphically presents the schedule for the program.

IV. PLANS

The major near term activity of the KAB team will be to complete the installation and evaluation of the Bhattacharyya distance algorithm on OLPARS. When it has proven to be effective it will be presented to CNVEO at the final briefing meeting and installed on their OLPARS. Finally, it will be documented in the final technical report of Phase I.

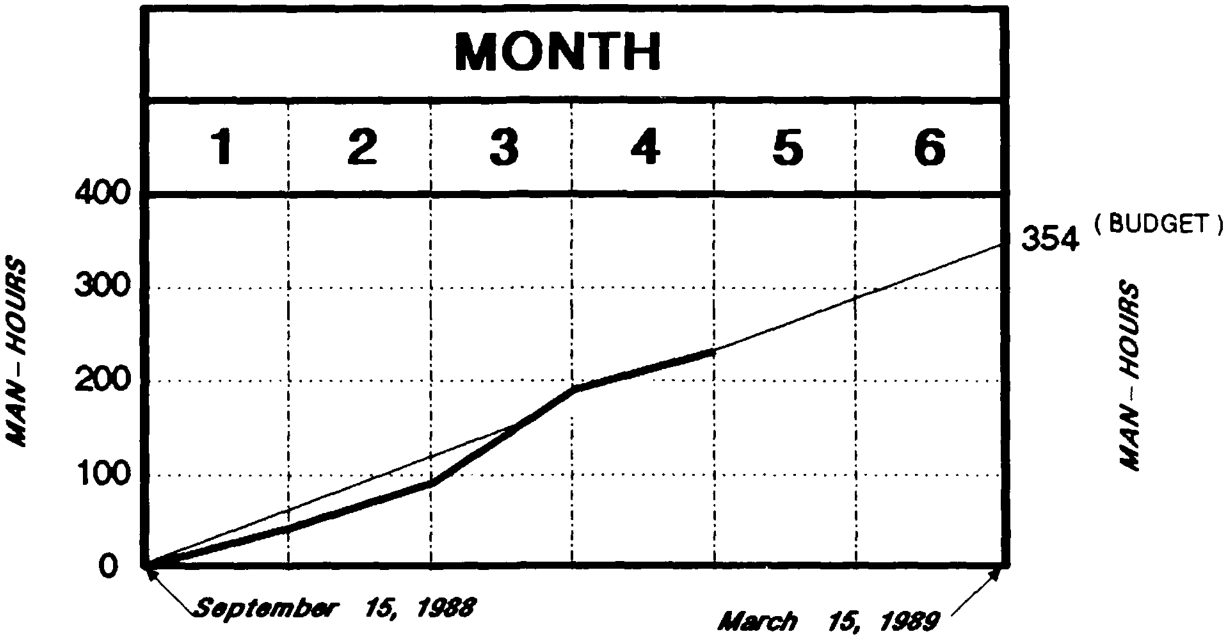

Figure 1. MAN-HOURS DELIVERED

PERSONNEL BUDGET/ACTUAL	MONTH 1	2	3	4	5	6
KABL principal investigator	36/40	36/45	36/40	36/20		
KABL research assistant	6/0	6/0	6/20	6/4		
PGSC senior staff member	15/0	15/0	15/45	15/15		
CUM. TOTAL	57/40	114/85	171/190	228/229		

Table 1. MAN-HOURS DELIVERED

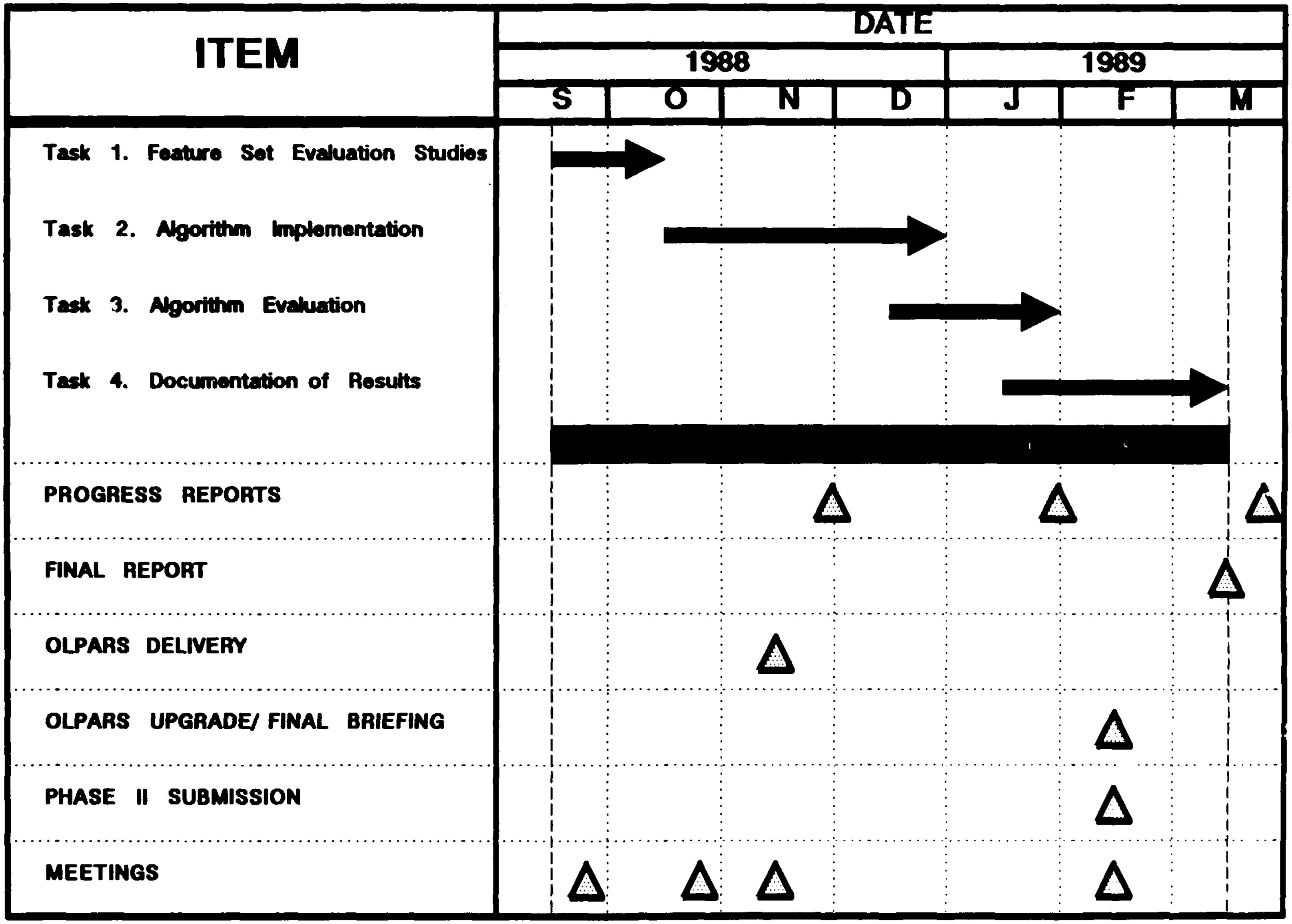

Figure 2. PHASE I SCHEDULE

During the remaining two months of Phase I, the KAB team will also plan to perform initial analysis of data sets provided by CNVEO. Time permitting, the team will also attempt to provide another OLPARS upgrade which will permit data vector identification for displayed data vectors. It is expected that close contact with the CNVEO sponsor, Mr. Lahart, will be maintained throughout these final two months.

V. CONCLUSIONS

The Bhattacharyya distance measure will provide a significant enhancement to the OLPARS. The OLPARS does not currently provide evaluation of multiple combined features, nor does it provide error bounding on its discriminent measures. The addition of this new tool on OLPARS will provide both a means to evaluate best groups of features, and will also provide error bounds on the ability to classify using those features. As a mature and complex system, however, OLPARS does not permit easy modification by users. PGSC has had some schedule slippage in their installation of the Bhattacharyya distance measures, but we are hopeful that the final schedule dates will still be met.