Small Business Innovation Research (SBIR) Program
Proposal Cover Sheet

Proposal Number: F041-115-1205 Agency: Air Force DUNS: 102385056
Topic Number: AF04-115 CAGE:
Proposal Title:

Multivariate Statistical Analysis Techniques for Data Fusion 2+

Firm:
Firm Name: RBI
Mail Address: 9511 W. Waneta Lake Rd.
Hammondsport, NY 14840-9511
Website Address: linkny.com/edrice4/engnr
Proposed Cost: 87633 Phase: I Duration: 9 Months
Business Certification: (Check all that apply)
Are you a small business as described in paragraph 2.2 (note: wholly owned subsidiaries are not eligible)? YES
Number of employees including all affiliates (average for preceding 12 months): 1
Are you a socially or economically disadvantaged business as defined in paragraph 2.3 ? NO
Are you a woman-owned small business as described in paragraph 2.4 ? NO
Are you a certified HUBZone small business concern as described in paragraph 2.9 ? NO
Are you a service-disabled veteran-owned small business as described in paragraph 2.11 ? NO
Are you a veteran-owned small business as described in paragraph 2.12 ? YES
Are you proposing to use a Federal facility or FFRDC? NO
Has a proposal for essentially equivalent work been submitted to other US government agencies or DoD components? NO
If yes, list the name(s) of the agency, DoD component or other SBIR office and Topic Number in the space below.

Project Manager/Principal Investigator
Name: Mr. Edward G. Rice
Title: Principal Investigator
Phone: (607) 292-6639
Fax: (607) 292-6639
E-Mail: edrice4@linkny.com

Corporate Official (Business)
Name: Mr. Edward G. Rice
Title: Sole Proprietor
Phone: (607) 292-6639
Fax: (607) 292-6639
E-Mail: edrice4@linkny.com

_____ _____

_____ _____
Signature of Principal Investigator Date Signature of Corporate Business Official Date

Technical Abstract

(Limit your abstract to 200 words with no classified or proprietary information)

The proposed effort entitled "Multivariate Statistical Analysis Techniques For Data Fusion 2+" will exploit little used multivariate statistical analysis techniques to develop an data fusion capability. The union of multivariate statistics and the AI techniques already being exploited for data fusion 2+ will provide superior probabilistic based decision making methods. Data fusion involves continuous refinement of estimates and evaluation of assessments. The probabilistic approach involved in multivariate analysis is well equip to refine data factors and measure likelihood of assessments. Multivariate statistics is used to evaluate factors used for an hypothesis and determine adequacy of data for conclusions. Data fusion involves evaluation of sources and possible modification of processes, chores directly related to this factor analysis techniques exploited in multivariate statistics. Multivariate Statistical Analysis techniques enable powerful data groupings and factor identifications. They supply dynamic probability values that confirm very high fidelity discriminations and make probabilistic estimates of fuzzy areas. An level 2 data fusion capability using multivariate statistic analysis techniques will be developed in an open architecture which will be prototyped and tested in a Phase II development.

Anticipated Benefits/Potential Commercial Applications of the Research or Development.

(No classified or proprietary information)

The commercialization of the multivariate statistical methods is pending the open architecture development of powerful data analysis modules. RBI expects to drive this R&D to that end. The tools used by the statistician to perform the correlations, factor analysis and clustering of diverse data require expert handling and understanding. The less diverse data used in the engineering environments enable the honing of these statistical methodologies into automated data processing modules which perform specific correlations and factor analysis to enable powerful understandings of raw data. These modules can be interconnected to perform multiple functions in the same way that business modules are interconnected in the open architecture of Java Beans. Thereby the engineer can use these statistical methods to do whatever he wants with myriad's of data.

RBI expects to use this multivariate statistical methodology to 1) find powerful solutionis to voice and signal identification challenges, 2) to perform geopositioning, tracking and fusion applications in the missile defense arena , 3) to perform data correlations and factor analysis for homeland defense functions and 4) to perform level 2+ data fusion advancements. As a new start company operating in a recession RBI has no material resources to invest in these four development areas. However, RBI has found an ample human resource of engineers operating out of their areas of expertise who are eager to invest their talents in the possibilities found in this vision. The opportunity for patents and sales of specific modules is expected to result from the Phase II development. The synergism from the addition of programmers and talented system engineers for a Phase II development will generate more opportunities for spin-offs of this technology. These diverse applications of such a powerful set of modules is expected to reap financial benefits within two years of their availability. The advancements to be retrofit into the modules for even more diverse applications will generate more benefits as well as advanced R&D investments.

List a maximum of 8 Key Words or phrases, separated by commas, that describe the Project.

Information Fusion, Fusion 2+, Multivariate Statistics, Statistical Analysis, Situation Assessment, Threat Assessment, Correlation, Clustering, Factor Analysis

Last Updated on: 1/8/2004 11:40:05 PM

2. IDENTIFICATION & SIGNIFICANCE OF THE OPPORTUNITY

In just the past few years of processing advancements multivariate statistics has emerged to enable powerful statistical analysis of diverse data sources. As cluster analysis methods have turned multidimensional and the interrelationships of various data sourses and data types have pointed to new insights and more decisive conclusions. As factor analysis has advanced into Eigen vector and multidimensional correlation analysis the links between the decisions factors and the drawn conclusions has matured and stratified. Probabilistic analysis has left the single dependent variable techniques of the past and is launched into analysis of multiple independent and dependent variables which brings statistical analysis into vogue with decisions making done with artificial intelligence (AI).

The union of multivariate statistics and the AI techniques already being exploited for data fusion 2+ will provide superior probabilistic based decision making methods. Data fusion involves continuous refinement of estimates and evaluation of assessments. The probabilistic approach involved in multivariate analysis is well equip to refine data factors and measure likelihood of assessments. Multivariate statistics is used to evaluate factors used for an hypothesis and determine adequacy of data for conclusions. Data fusion involves evaluation of sources and possible modification of processes, chores directly related to this factor analysis techniques exploited in multivariate statistics. Multivariate Statistical Analysis techniques enable powerful data groupings and factor identifications. They supply dynamic probability values that confirm very high fidelity discriminations and make probabilistic estimates of fuzzy areas.

The marrying of Multivariate Statistical Analysis with current AI data fusion techniques deserves a systematic evaluation. The benefits of such a marriage loom just out of reach of the statistician and just beyond the AI expert. A system engineer with training in statistics and experience in data fusion can unite the two powerful fields.

2.1 Background. Information fusion to provide enhances situational awareness is accomplished daily by the human intellect. Man can process multiple types of information; sight, sound, and touch; distance, time, and danger to stay keenly aware of his situation, or not. Artificial intelligence was intended to mimic this human process without fatigue. It was able to process multiple types of data and even reach some conclusions, but had massive downfalls when it came to situational awareness. The advent of fuzzy logic helped but the goals of intelligent situational awareness would need greater strides than the probabilistic inclusions of fuzzy logic. The infusion of statistical analysis techniques into the data fusion arena adds powerful correlation and factor analysis methods which are missing in AI and fuzzy logic approaches.

Multivariate Statistical analysis has come a very long way in the powerful computational environment of the past 5 years. The superb processing capabilities of the last few years has fielded powerful tools for the statistician to perform market analysis, political estimation and social/medical groupings based on multiple variables and variations. These probabilistic multivariate analysis techniques are finding powerful applications in these abstract fields of reasoning and will find equally powerful application in the data fusion arena. Multivariate Statistical Analysis techniques enable powerful data groupings and factor identifications. They supply dynamic probability values that can dynamically feed AI decision makers already developed.

The marrying of Multivariate Statistical Analysis with data fusion deserves a systematic evaluation. The benefits of such a marriage loom just out of reach of the statistician and just beyond the expert in AI.

2.2 Multivariate Statistics Background. Multivariate statistics are used extensively to :

1) develop taxonomies (or system of classifications.)
2) to investigate useful ways to conceptualize or group items.
3) to generate hypothesis, and

4) to test hypothesis.

Many mathematical and statistical computer tools are available to accomplish theses purposes. The massive computer power unleashed with multivariate analysis tools is revolutionizing the analysis capability of marketers, politicians, sociologists and graduate statistics students. These powerful tools unleash faster analysis methods, multidimensional scaling techniques and cluster analysis advances to look at and understand interrelationships among multiple variables. Politicians, Sociologists, and Business Marketers are on the leading edge of this technology breakthrough. Engineers doing signal analysis, geopositioning, target ID and level 1 fusion are on the trailing edge when it comes to utilizing these powerful multivariate analysis techniques.

Multivariate statistics can do two major functions to edify the fusion or signal processing tasks of the systems engineer. First it looks at multiple variables related to the task and through the use of correlation's and Eigen vector analysis it determines the interdependence of the various variables. One method of revealing the interdependence of variables is through multidimensional scaling (MDS). We can define MDS as "a set of multivariate statistical methods for estimating the parameters in and assessing the fit of various spatial distance models for proximity data" (cf. Davison, M.L. (1992)) The more powerful tool for this interdependence is with factor analysis. (cf. Kim, J., and Mueller, Charles W. (1978)) When making threshold decisions in either signal processing or in data fusion the interrelationship of the factors in the decision process provide vital information. With a clear understanding of the interrelationships the key factors for a particular decision may be separated from the myriad of data and be used to clarify and validate the decision.

Secondly the multivariate statistical analysis enables groupings of data into categories of similarities. These groupings can be as wide or as narrow as desired. The categorization can use the fidelity of multiple variables to single out an individual entry and even perform ID or fingerprint type classifications. This type of grouping is done via cluster analysis (Aldenderfer, M.S., and Blashenfield, R.K. (1984)) or MDS methods previously mentioned. Previous statistical analysis used dependence methods whereby multiple regression and analysis of variance used several variables but one dependent variable was predicted by means of several independent variables (Hair, J.F. et al. (1992)). These multivariate statistical methods allow examination of interrelationships among variables thereby enabling powerful new abilities to enhance signal processing, identifications or 'fingerprinting' of signals or data sources, and the fusion of related data into situational awareness. Multivariate statistics has powerful applications being utilized by statisticians yet still untapped by system developers.

3. PHASE I TECHNICAL OBJECTIVES

The phase I objectives are to structure several multivariate statistic analysis tools into a front end processor which acts on fusion data in such a way that effective level 2 data fusion can be performed by conventional AI tools. The tools methodology will need to be recoded into open and transportable Java code. These modules will be used to automate performance of correlations, factor analysis then cluster analysis on applicable data sources. The pertinent and grouped factors will then be used to perform situational awareness type activities. This phase will enable the first measure of the performance of multivariate statistics analysis methods in level 2 data fusion. This architecture and the measure of its effectiveness will then be used to advance a data fusion prototype which can be coded and tested in an open architecture environment during Phase II of the effort.

4. PHASE I - WORK PLAN

The Phase I work will define and test the applicable architecture for the integration of multivariate statistics analysis technology into existing AI fusion methods. These tasks assume that the multivariate statistics will operate as a front end processor and AI methods will invoke situational awareness activities. This is a tentative architecture which will be reviewed in development.

The Phase I work plan will include the following tasks for achieving the stated objectives:

4.1 Task I - Investigate and/or Develop The Data Fusion Test Scenario (100 hr.)

The information necessary to perform level 2 data fusion will be diverse and abundant. Several scenarios that have multiple data sources will need to be defined which will challenge a data fusion methodology. Fusion algorithms can be tailored to react well to a scenario or two. In this task a diverse set of scenarios shall be investigated which will challenge a multivariate statistics driven fusion methodology.

4.2 Task II - Investigate and Develop Applicable Multivariate Analysis Methods (300 hr.)

Multivariate statistical analysis tools are generally made to analyze dichotomous data from survey or polling inputs. These tools are robust and powerful but are not automated nor versatile. The data to be input to these multivariate statistic methodologies will need to be handled by a front end before factor analysis, cluster analysis, or cross correlation can be effective. The multivariate methods will be implemented in Java to expedite their implementation into an open architecture. These tools are not yet automated, and the results to be fed into the decision makers must be handled by a back end before AI tools can complete the data fusion. The front end and the back-end of the multivariate statistic analysis tools will be defined and developed in Java in this task.

4.3 Investigate AI and Fuzzy Logic Methods (150 hr.)

Artificial Intelligence (AI) tools which evaluate data to perform situation awareness will be investigated and selected in this task. The AI tools will be investigated and then modified to take full advantage of the correlations and factor analysis aspects of the multivariate statistics.

4.4 Evaluate Union of Multivariate and AI Methods in Realistic Scenario (200 hr.)

In this task the multivariate statistics output will be optimized to benefit fusion and the AI data handlers will be optimized to handle this optimal data for effective level 2 data fusion. The scenarios will be exercised to evaluate the effectiveness and advantages of the multivariate statistic package. The overall architecture of the inclusion of the statistics tools in this approach will be examined and the architecture and statistics modules will be tweaked to finalize the multivariate statistical approach to level 2+ data fusion.

4.5 Reporting. (50 hr.)

A. Bimonthly Progress reports: In addition to day-to-day informal contacts with the program monitor, technical progress reports will be submitted every two months.

B. A Final Report will be generated. A final report detailing all of the development and evaluation will be submitted at the end of Phase I.

The Phase I work would follow the tentative schedule below:

Task	Days	Hrs	Start	Stop
SBIR Phase I MVSAT for Data Fusion	280	800	03/01/04	12/06/04
Task 4.1 Data Fusion Test Scenario	35	100	03/01/04	04/05/04
Task 4.2 Multivariate Analysis Methods	105	300	04/05/04	07/19/04
Task 4.3 Logic Methods	52.5	150	07/19/04	09/09/04
Task 4.4 Evaluate Union	70	200	09/09/04	11/18/04
Task 4.5 Reporting	17.5	50	11/18/04	12/06/04

Waterfall Chart of Tasks

```
SBIR Phase I        ...A.....M.....J.....J.....A.....S.....O.....N.....D.
4.1 Scenario        XXXXXXX.....................................................
4.2 Multivariate    .......XXXXXXXXXXXXXXXXXXXXX.....................
```

```
4.3 Logic Methods    ..........................XXXXXXXXXX..............
4.4 Eval Union       ....................................XXXXXXXXXX.....
4.5 Reporting        ..............................................XXX.
```

5. RELATED WORK

The Principal Investigator is the sole proprietor of RBI and has been involved in pursuing categorizing arbitrary bit streams for the past several years. While in the USAF he was responsible for the programs that enhance the performance of numerous intelligence collection platforms. With minimal funding and resources RBI is staying abreast of the current technologies and intelligence mission requirements. RBI has used multivariate statistical analysis methods to perform classifications and identification of arbitrary bit streams. The multivariate methods as used in that task will have direct application to several system engineering challenges found in the military intelligence R&D incentives. RBI maintains connections with both DOD military intelligence R&D efforts and multivariate statistics R&D advancements to stay on the leading edge of this breaking technology.

6. Relationship With Future Research Or R&D

The successful use of multivariate statistics methods in the data fusion arena will open doors of opportunity for both level 2+ data fusion and for this powerful factor analysis methodology. If these statistical methods mark notable improvements for data fusion in Phase 1, the follow on phases of prototyping will field open architecture Java modules which will find application in multiple fields of both fusion and digital processing. RBI will advance it's system engineering arm to lead in other R&D efforts which will exploit these little used methodologies in areas of military intelligence collections, missile defense missions, homeland security intelligence collections and level 2+ data fusion areas that are currently stalemated without adequate correlation and clustering techniques to sort through their myriad of mitigating factors.

7. Commercialization Strategy

The commercialization of the multivariate statistical methods is pending the open architecture development of powerful data analysis modules. RBI expects to drive this R&D to that end. Multivariate statistics is currently in vogue and application in college, in social/political polling and in marketing positions. The tools used by the statistician to perform the correlations, factor analysis and clustering of diverse data require expert handling and understanding. Thereby the statistician can make the statistics say whatever he wants. The less diverse data used in the engineering environments enable the honing of these statistical methodologies into automated data processing modules which perform specific correlations and factor analysis to enable powerful understandings of raw data. These modules can be interconnected to perform multiple functions in the same way that business modules are interconnected in the open architecture of Java Beans. Thereby the engineer can use these statistical methods to do whatever he wants with myriad's of data.

RBI expects to use this multivariate statistical methodology to 1) find powerful solutionis to voice and signal identification challenges, 2) to perform geopositioning, tracking and fusion applications in the missile defense arena , 3) to perform data correlations and factor analysis for homeland defense functions and 4) to perform level 2+ data fusion advancements. As a new start company operating in a recession RBI has no material resources to invest in these four development areas. However, RBI has found an ample human resource of engineers operating out of their areas of expertise who are eager to invest their talents in the possibilities found in this vision. The opportunity for patents and sales of specific modules is expected to result from the Phase II development. The synergism from

the addition of programmers and talented system engineers for a Phase II development will generate more opportunities for spin-offs of this technology. These diverse applications of such a powerful set of modules is expected to reap financial benefits within two years of their availability. The advancements to be retrofit into the modules for even more diverse applications will generate more benefits as well as advanced R&D investments.

8. KEY PERSONNEL

Edward G. Rice, Senior Engineer

EDUCATION:
M.S., Electrical Engineering, Air Force Institute of Technology, Wright Patterson AFB Ohio, March 1992.
B.S. Electrical Engineering, Ohio State University, Columbus Ohio, March 1982.

CURRENT POSITION AND RESEARCH:

Edward Rice is a retired USAF officer and the Sole Proprietor of RBI. He has more than 18 years of experience in USAF intelligence systems. RELEVANT EXPERIENCE: Prior to starting RBI Ed Rice was the consultant that developed and tested the basic Euclidean Distance Vector Fitting Technique for Categorizing Arbitrary Bit Streams. From 1989 to 1995 he was assigned to Rome Laboratories where he managed numerous intelligence collection technology initiatives. From 1984 to 1989 as a USAF Electrical Engineer he worked with numerous data collection and weapon delivery systems furthering and rounding out his background in intelligence systems. He is currently Pastoring a Baptist Church and teaching high school math and science part-time, while pursuing a M.Div. Degree. He is available to work on RBI efforts up to 40 hours per week.

9. FACILITIES/EQUIPMENT

Due to cutbacks and limited resources RBI currently has no facilities and minimal equipment. The economic recession in the New York area makes available ample office space and computer resources which can be in place with minimal investment and time. Currently algorithm development and statistical analysis is performed on a home computer in an office in the basement of my home. A Java development environment , Excel Spreadsheets utilizing Visual Basic, tools in Microsoft Office and Lotus Millennium have proven very adequate on a PC to perform all research and module developments. Should RBI be funded for a multivariate statistical development ample facilities and equipment could be in place within 30 days.

10. CONSULTANTS

Several engineers with backgrounds in statistics, intelligence and computers are available to act as consultants for this effort. RBI is close to Cornell University, Rochester and Syracuse NY where unemployed engineers await a call and opportunity to get involved in a visionary development for RBI. Additionally the internett makes available a world of such expertise which can be utilized to perform this multivariate statistical development. Any consultants utilized by RBI for this effort will be paid by RBI, are not reflected in the cost proposal and shall not alter the cost proposal of this effort.

11. PRIOR, CURRENT OR PENDING SUPPORT

RBI has 3 SBIR proposals submitted which utilize the powerful advantages of multivariate statistical analysis techniques. There is great potential in any one of these proposals, if two or three of them are initiated the synergism between them will benefit all. RBI can pull together the resources necessary to perform all three of these efforts simultaneously with advantage to each participant. The three proposals are titled as follows:

Multivariate Statistical Analysis Techniques For Data Fusion 2 +

SBIR Topic Num: AF04-115
SBIR Title: Innovative Approaches to Fusion 2 +
SBIR Research & Technical Areas: Information Systems

SBIR Topic Author: Michael Hinman,
Phone: (315) 330-3175 Fax: (315) 330-4380
Email: hinmanm@rl.af.mil

Multivariate Statistical Analysis Techniques For Arbitrary Bit Stream Pattern Recognition

SBIR Topic Num: MDA04-020
SBIR Title: Innovative Techniques for Missile Defense
SBIR Research & Technical Areas: Weapons
SBIR Topic Author:Mr. Vincent Nguyen
Phone: (703) 695-0285 Fax: (703) 695-6222
Email: Vincent. Nguyen@mda.osd.mil

Multivariate Statistical Analysis Techniques For Automatic Speech Identification

SBIR Topic Num: AF04-062
SBIR Title: Expanded Speech Recognition to Include Foreign Accents
SBIR Research & Technical Areas: Human Systems
SBIR Topic Author: Mr. David Williamson
Phone: (937) 255-7593
Email: david.williamson@wpafb.af.mil

12. COMPANY COMMERCIALIZATION REPORT (SEE ADDITIONAL ELECTRONIC SUBMITTAL)

13. COST PROPOSAL (SEE ADDITIONAL ELECTRONIC SUBMITTAL) (See Last Page)

14. REFERENCES and FOOTNOTES

Aldenderfer, M.S., and Blashenfield, R.K. (1984)
Cluster analysis. Newbury Park, CA: Sage Publications.

Davison, M.L. (1992)
Multidimensional scaling. Malabar, Fl, CA: Krieger Publishing.

Hair, J.F. et al. (1992)
Multivariate data analysis (3rd ed.). New York: Macmillan.

Kim, J., and Mueller, Charles W. (1978)
Introduction to factor analysis: What it is and how to do it. Newbury Park, CA: Sage Publications.

Romesburg, H.C. (1984)
Cluster analysis for researchers. Belmont, CA: Lifetime Learning Publications.

Rummel, R.J. (1984)
Applied factor analysis. Evanston, IL: Northwestern University Press.

.

RBI Cost Proposal

RBI 9511 W.Waneta Lake Rd, Hammondsport NY 14840
Date: **1-Jan-04**
Phone: **(607) 292-6639** CAGE Code: _____
Title: **Multivariate Statistical Analysis Techniques For Data Fusion 2** +
Topic: **AF04-115 Innovative Approaches to Fusion 2 +**
Total Proposal Amount **$87,673.97**

Direct Material	EA	PER COST	EST COST	TOTAL
a. PURCHASED PARTS	NA		0	
b. SUBCONTRACTED ITEMS	NA		0	
c. OTHER	NA		0	
TOTAL DIRECT MATERIAL			.	0
	9	MO		
Direct Labor	EST MM	RATE/HR	EST COST	TOTAL
(1728 MH/MY)	20	hr/wk		
Principle Engineer	806	45	36,270	
Jr Engineer	0	24	0	
Programmer	0	24	0	
Publications	0	13	0	
TOTAL DIRECT LABOR	5.6	0	ENGRS	36,270
LABOR OVERHEAD	OH RATE	XBASE	EST COST	
a. IN PLANT	0.7	36,270	25,389	
b. ON SITE	0.55	0	0	
TOTAL LABOR OVERHEAD	.	.	.	25,389
SPECIAL TESTING	.	.	.	0
SPECIAL E UIPMENT	.	.	.	0
TRAVEL	EA	RATE	EST COST	
a. TRANSPORTATION	3	979	2,937	
b. PERDIEM	10	85	850	
c. LOCAL TRANSPORTATION	10	32	320	
TOTAL TRAVEL	.	.	.	4,107
CONSULTANT	Hrs	Rate	Cost	
	0	75	0	
TOTAL CONSULTANT	.	.	.	0
OTHER DIRECT COST	.	.	.	0
TOTAL DIRECT COST AND OVERHEAD		.	.	65,766
GENERAL ADMIN EXPENCE	0.25	OF COST	.	16,442
COST OF MONEY	0	OF COST	.	0
TOTAL ESTIMATED COST	.	.	.	82,208
FEE OR PROFIT	0.07	OF COST	.	5,426
TOTAL ESTIMATE AND FEE OR PROFIT		.	.	87,633.20