

UNCLASSIFIED

Defense Technical Information Center
Compilation Part Notice

ADP021704

TITLE: Information Fusion for Situational Awareness

DISTRIBUTION: Approved for public release, distribution unlimited

This paper is part of the following report:

TITLE: Proceedings of the International Conference on International Fusion [6th]. Held in Cairns, Queensland, Australia on 8-11 July 2003. Volume 1: FUSION 2003

To order the complete compilation report, use: ADA442007

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:

ADP021634 thru ADP021736

UNCLASSIFIED

Information Fusion for Situational Awareness

Dr. John Salerno, Mr. Mike Hinman, Mr. Doug Boulware, Mr. Paul Bello

AFRL/IFEA, Air Force Research Laboratory, Rome Research Site Rome, NY, USA

John.Salerno@rl.af.mil, Michael.Hinman@rl.af.mil, Douglas.Boulware@rl.af.mil, Paul.Bello@rl.af.mil

Abstract - Information Fusion is beginning to receive increased attention not only within the military, but also within the civilian sector. The notion of maintaining constant awareness of one's surroundings is not a new military goal, but to provide this capability through computers has been a long standing and challenging problem. In this paper we will explore various techniques that we believe are necessary to provide situational awareness and how they can be tied together into an overall system architecture. We will also explore various sources of information, their contribution to the problem and how we propose to test and evaluate the proposed architecture.

Keywords: Higher Level Fusion, Situational Awareness, Situation Assessment, Indications & Warning, Area Assessment

1 Introduction

In an article written for the NY Times, Robert Steele [10], writes "Despite the fact that U.S. taxpayers have been paying more than \$30 billion a year for a national intelligence and counter intelligence community to protect it from both traditional state-based threats and unconventional non-state actors, the events of 9-11 demonstrated our inability to detect and prevent bold asymmetric attacks that used our own airliners as precision missiles. Armed with new concepts, money, and suicidal pilots, Osama bin Laden has cost us at least \$20 billion in damages." From this statement, Steele continues, "We need a 'new craft of intelligence' that can access and digest the broad historical, cultural, and current events knowledge that is available openly in over twenty-nine languages — by exploiting these open sources we can create open source intelligence suitable for informing our public as well as our state and local authorities and our international partners, as to the threats to our nation." This is a focus of information fusion and how it can be applied to address Situational Awareness - the goal of this paper. After this section we present a short discussion of the two most popular fusion models and how we applied them to a process. We present a high level overview of the components as

defined in our process. We conclude with a discussion on metrics.

1.1 Background

Over the years, more than thirty fusion models have been proposed and countless research initiatives and personnel have attempted to define them. No model has become as influential in the Data Fusion area as the Joint Director's of Laboratories (JDL). As shown in Figure 1, and described by Steinberg, Bowman and White [11], the JDL model has five levels: Level 0 - Sub-Object Data Assessment; Level 1 - Object Assessment; Level 2 - Situation Assessment; Level 3 - Impact Assessment; and Level 4 - Process Refinement.

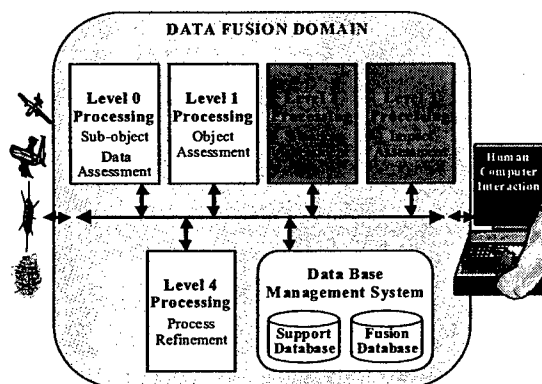


Figure 1. JDL Fusion Model.

A stream of data enters the model at level 0, Sub-Object Data Assessment. Level 0 provides physical access to the raw bits or signal. In addition, estimation and prediction of the existence of an object is performed based on pixel or signal level data association and characterization.

Objects are correlated and tagged over time in an attempt to build tracks and to perform object identification during level 1 processing, or Object Assessment. During Situation Assessment, or level 2 processing, the knowledge of objects, their characteristics, relationships with each other and cross force relations are aggregated. This aggregation is performed in an attempt to understand the current situation. Previously discovered or learned models

generally drive this assessment. After Situation Assessment, the impact of the given situation must be assessed. The impact estimate can include likelihood estimates and cost/utility measures associated with the potential outcomes of a player's planned actions. The final level, Process Refinement, provides a feedback mechanism back to each of the other layers, including the sensor itself. Over time the JDL model has evolved. The original model actually consisted of levels 1, 2 and 3. The model was first altered with the addition of level 4, and level 0 was just recently added. To date, those that use the model have concentrated their efforts on sensor level (0 and 1) object identification and tracking algorithms and in developing algorithms to perform model assessment.

While the JDL provides a functional model for the data fusion process, it does not model it from a human perspective. Endsley [3] provides an alternative to the JDL model that addresses Situational Awareness from this viewpoint (i.e., Mental Model). Her model has two main parts: the core Situation Awareness portion and the various sets of factors affecting Situation Awareness. The core portion follows Endsley's [4] proposition that Situation Awareness has three levels of mental representation: perception, comprehension, and projection. The second and much more elaborate part describes in detail the various factors affecting Situation Awareness. Endsley defines Situation Awareness as a state of knowledge that results from a process. This process, which may vary widely among individuals and contexts, is referred to as Situation Assessment, or as the process of achieving, acquiring, or maintaining Situation Awareness. The three levels of Situation Awareness as proposed by Endsley, shown in Figure 2, must be described in greater detail because Endsley's model fuels much of the current research in Situation Awareness.

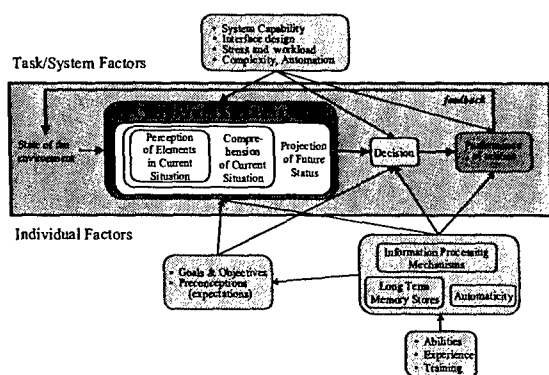


Figure 2. Endsley's Situation Awareness Model.

The first step in achieving Situation Awareness is what Endsley has labeled as Perception (or Level 1). Perception provides information about the status,

attributes and dynamics of the relevant elements in the environment. It also includes the classification of information into understood representations and provides the basic building blocks for the higher levels. Without a basic perception of important information, the odds of forming an incorrect picture of the situation increase dramatically [5].

Comprehension of the situation, or level 2, encompasses how people combine, interpret, store, and retain information. Thus, it includes more than perceiving or attending to information; it includes the integration of multiple pieces of information and a determination of their relevance to the underlying goals. Comprehension provides an organized picture of the elements with an understanding of the significance of objects and events. It also combines new information with already existing knowledge to produce a composite picture of the developing situation. The ability to forecast future events mark decision-makers that have the highest level of Situation Awareness. Endsley calls this level Projection (Level 3). Situation Awareness refers to the knowledge of the status and dynamics of the situational elements and the ability to make predictions based on that knowledge. McGuinness and Foy [8] extended Endsley's Model by adding a fourth level, which they called Resolution. This level provides awareness of the best path to follow to achieve the desired outcome to the situation. Resolution results from drawing a single course of action from a subset of available actions. McGuinness and Foy believe that for any fusion system to be successful, it must be resilient and dynamic. It must also address the entire process; from data acquisition to awareness, prediction and the ability to request elaboration or additional data. McGuinness and Foy put Endsley's model and their model into perspective with an excellent analogy. They state that Perception is the attempt to answer the question "What are the current facts?"; Comprehension asks "What is actually going on?"; Projection asks "What is most likely to happen if...?" and Resolution asks "What exactly shall I do?" Another point to be made is that any proposed model should not promote a serial process, but rather a parallel one. Neither the JDL Model nor Endsley's suggest otherwise. Each function (for example in Endsley's model: Perception, Comprehension, Projection and Resolution) happens in parallel with continuous updates provided to and from each other function. Table 1 summarizes the four functions or levels. McGuinness' extension is presented here only to show the parallels between the JDL and Endsley model.

PERCEPTION	<u>Contents:</u> Explicit objects, events, states, values <u>Processing:</u> Sensing, detection, identification
COMPREHENSION	<u>Contents:</u> Implicit meanings, situation types <u>Processing:</u> Interpretation, synthesis
PROJECTION	<u>Contents:</u> Future scenarios, possible outcomes, <u>Processing:</u> Prediction, simulation
RESOLUTION	<u>Contents:</u> Intentions, courses of action <u>Processing:</u> Decision-making, planning

Table 1. Summary of the Four Functions (as defined in McGuiness and Foy).

In the next section we begin a discussion of our concept from the bottom up - the data. From there we present various components that we feel are necessary in building a Situation Awareness framework.

2 A Conceptual Framework

The majority of the differences in the proposed models deal with syntax rather than semantics. Instead of arguing or debating the differences, we would like to move beyond these debates and provide a roadmap as to how we can build a fusion system for Situation Awareness and

Assessment. We will, however, discuss our concept in terms of Endsley's model for reference. A second point to make before we begin our discussion is that typically concepts or systems are built from the top down. The problem that we have seen is that the majority of the time non-real world assumptions are made. Specifically, information needed or assumed to be available often does not exist, only exists at higher security levels (and is thus not available at the lower level), or lacks the granularity required. For this reason, we begin our discussion from the data up. The data is what will restrict or confine our comprehension and thus projection. In our concept we classify data based on its perishability: real time, near-real time and historical or non-real time. Real-time data contains streaming bits or signals from various types of sensors. This data can range from un-exploited sensor data to still and moving imagery. Near-real time data includes reports generated by analysts exploiting sensor data or analyzing imagery. Near-real time also includes other sources such as news wire services (e.g., CNN, Reuters, API, etc.) Non-real time or historical data has been previously collected or archived (e.g., AltaVista, Yahoo, Lycos, etc.) Historical data can also include data from existing databases. Figure 3 provides a view of our concept that is based on both the type of data (as previously discussed) and Endsley's three lower levels.

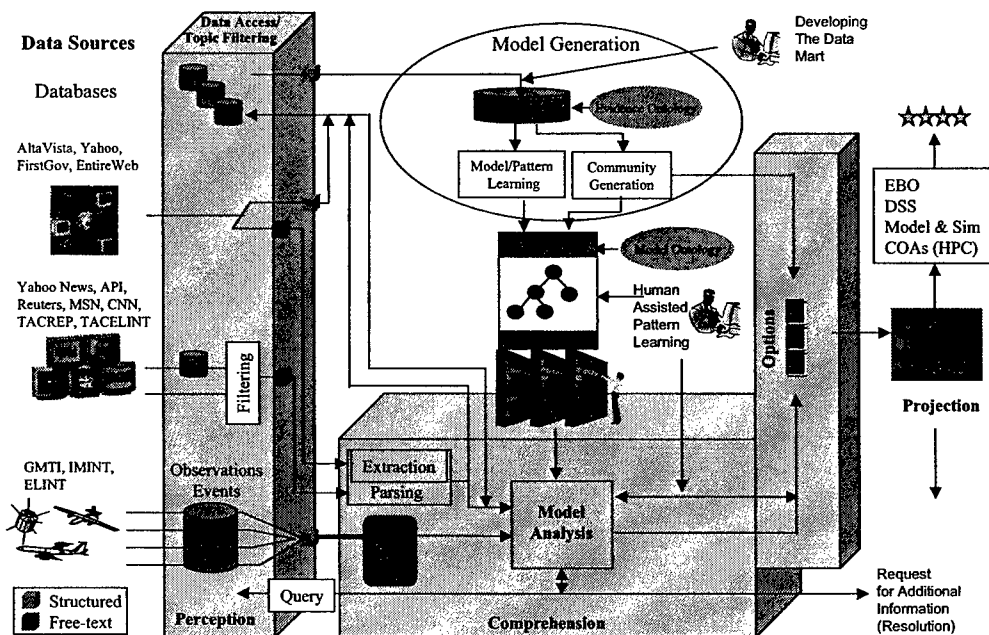


Figure 3. Conceptual Functional Flow.

2.1 Acquiring Perception

The best decisions are made with the best and most accurate data. The challenge in any system is in acquiring data - the right data and all the data. This data can exist anywhere and across multiple sources. How do we find this data? How do we know that we have all the available data? How do we know the data is correct? These are a few of the major issues that are being examined under a number of research activities. The bottom line is that currently we don't have an answer to many of these issues.

2.2 Obtaining Comprehension

With the technological evolution in Command and Control (C2) environments, it is possible to create more powerful and sophisticated systems that make available to the decision-maker a huge repository of data and information about the environment. However, the problem with today's systems is not a lack of information, but finding what is needed when it is needed [5]. Providing as much data and information as possible about the environment is not necessarily an adequate way to support the decision-maker's performance [2]. In fact, in some situations, most of the data can be seen as distracters and noise for the decision-maker.

In order to comprehend the current situation and its relevancy one must have some knowledge of similar situations that occurred in the past and relevant events currently occurring to develop some basis. If this basis or foundation does not exist, an individual will resort to basic instinct and reaction. In our environment this knowledge from the past is captured in what we refer to as models and the comparison or similarity measures of which models are unfolding or active is performed through model analysis. Thus for us to maintain a "correct" situation awareness we must be able to capture the past in order to be aware of the present.

Models can exist in many forms and are either learned or developed. Models would be learned by deriving them through data sets and would include such concepts as activities and capabilities. Developed models would be based on factual data obtained from databases, e.g. equipment and units. We will discuss models and model generation techniques briefly in the next section.

Model Analysis may be a key technique in reducing the amount of data to present to the analyst and increasing the analyst's performance. The goal of Model Analysis is to search vast quantities of data and/or monitor incoming streams of data for specified

patterns, or models, of interest. Model Analysis, however, may be performed in two distinct ways; retrospectively or prospectively. In a retrospective manner, an analyst uses a Model Analysis tool to search a given data set for models, or patterns, of interest. The results of each search are the instances within the data that conform to the specified pattern along with a measurement of how closely the instance data matches the model of interest. While retrospective Model Analysis tools exist, they struggle to overcome a number of key limiting factors. Model Analysis that relies on textual input must either have a secondary source that can resolve questions about what certain words mean, or the input must be normalized before it is fed to the Model Analysis tool. Time and visualization also become limiting factors as the size of the data or models increases. New techniques must be developed to drastically reduce the search space in order for Model Analysis to be applicable in certain key domains. Similarly, it is very difficult to view large amounts of data, complex models, and information that are neither spatial nor temporal.

In the prospective approach, events and observations are being processed and analyzed against various models in real time. The objective here is to alert the analyst as to possible activities unfolding. The majority of the work in this area has been done on the data fusion side. Hinman [6] described a series of innovative approaches of traditional fusion algorithms and heuristic reasoning techniques to improve situational assessment and threat prediction. Provided below are some examples of techniques and challenge problems addressed. It should not be implied from this list that the discussed technique is the only solution to the given problem or that the technique only applies to the problem.

- Bayesian techniques have been utilized for the successful implementation of a force aggregation capability that permits the identification of military units.
- Knowledge Based approaches are being utilized to identify vehicles based primarily on vehicle movement information.
- Artificial Neural Systems (Neural Networks) are being utilized in a couple of different applications. The first application utilizes a multi-layer network that has been trained using Back Propagation to identify pairwise preferences of analysts to support situation assessment. The second application describes a prototype implementation based on Linear Vector Quantization (LVQ) and Ellipsoidal

Basis Functions that postulates threat (Attack, Retreat, Feint, or Hold).

- Fuzzy Logic techniques are being evaluated for the development of a fuzzy logic event detector that performs a fuzzy logic-based analysis of predicted courses of action to infer enemy intent and objectives.
- Genetic Algorithms are being utilized in various applications relative to situation assessment. The application that will be discussed in this paper involves using Genetic Algorithms for determining plausible courses of action.

2.2.1 Building Long Term Memory

Predictive analysis requires information about past events and their outcomes. Much of the work in this area requires a predefined model built by subject matter experts, or substantial amounts of data to train model generation software to recognize patterns of activity. To date these models are manually intensive to construct, validate, and interpret. Algorithms are needed to provide efficient inference, reasoning, and machine learning procedures. Learning applications range from data mining programs that can discover general rules from large data sets to "knowledge assisted" hybrid approaches aimed at accomplishing deeper levels of reasoning and pattern identification as well as information filtering systems that automatically learn users' interests.

Witten, Frank & Gray [12] defined data mining as the extraction of implicit, previously unknown, and potentially useful information from data. The idea is to build computer programs that sift through databases automatically, seeking regularities or patterns. They go on to state that strong patterns, if found, will likely generalize to make accurate predictions on future data. We divide Data mining techniques into two activities: (1) identifying patterns based on event associations which we refer to as pattern learning and (2) identifying groups based on similar activities which we refer to as community generation.

It is crucial that we thoroughly sift through archived data to look for the associations between entities at multiple levels of resolution. Pattern learning technologies serve to address this task by providing techniques that mine *relational* data. Pattern learning can be roughly described as the process of examining the relationships between entities in a database; the end-product of which are predictive models (statistical extrapolations), capable of describing what has been examined in terms of an abstract mathematical formalism (usually, a graph-

theoretic construct). Relational data presents several interesting challenges:

- Relational learning must consider the neighborhood of a particular entity, and not just a singular record.
- Most learning is predicated on (usually false) assumptions of independent samples. Relational data does not meet this criterion.
- Data must be semi-structured to make learning possible. A query language must be developed to support the retrieval of data.

The biggest concern in developing a pattern learner for situation awareness is the relatively low number of so-called "positive instances", turning the pattern learning process into an anomaly detection process. Problems such as these are often considered "ill-posed" in the computational learning community, and more often than not, partially invalid assumptions about the data must be made to correct for these conditions. If improperly handled, low rates of positive instances will completely confound the learning process, resulting in low-fidelity models, which produce high numbers of false positives/negatives.

Missing and corrupted data are also a prime source of error. Numerical data is naturally a bit easier to work with, given the fact that we can interpolate. The lack of numerical descriptors for the type of archived data with which we often deal exacerbates the issue of missing items. Luckily, there has been a recent surge of research activity in the domain of relational learning, addressing all of these issues. For a more complete description of these problems, and their respective solutions, we direct the sedulous reader to [7].

Imagery, from the perspective of situation awareness, presents its own unique set of issues. Many researchers have been involved in activities which aim to locate and identify objects in imagery by using novel classification techniques. Waxman, etal [12], for example, enumerates the combinations of features from different imagery data (SAR, Ikonos, etc.), and trains a specific type of neural network with positive instances in the picture verified by the user. The feature sets are mined, and the network is trained using the verified instances. This network is then used to classify objects with similar features in any subsequently presented image.

Community generation and the class of problems it is trying to solve can be categorized as one of discerning group membership and structure. Under this topic two types of paradigms are being investigated: one where two parties and the activity type are given and one where only one party and one

associated event is given. Zhang [14] describes the first class as bi-party and the later as uni-party.

Community generation algorithms will typically take events and relationships between individuals (whether implicit or explicit) and develop some correlation between them. This correlation value defines the strength of the link. Why are these models important to us? The models derived provide us insights into organizational structure and people of interest. Let us consider the first instance – organizational structure. Suppose that we have identified two groups whose structures are shown in Figure 4.

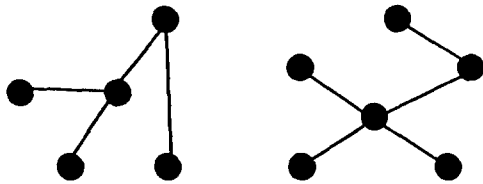


Figure 4. Community Generated Models.

We can easily see from the above models, that there is a key node in the model, which if removed or identified could have major impacts on the community. In this case, it could be a key individual within an organization. A second use of this information is the development of a behavioral model for the group. Knowing the individuals in charge of the group and “understanding” their behaviors or could facilitate more advanced modeling and simulation capabilities as well as direct surveillance efforts.

2.3 Metrics and Evaluation Environment

Measures of Performance (MOP's) and Measures of Effectiveness (MOE's) are central to the development of a capability for Situation Awareness. Many difficult questions must be addressed during the acquisition and processing of the information which is vital to achieving awareness. Some of the more challenging obstacles to overcome in this process include the following:

- Measuring the quality of information.
- Qualitative measures of uncertainty over heterogeneous information sources.
- Developing consistent metrics between the multiple interacting technologies which enable Situational Awareness.
- Managing the bias/variance tradeoff during model selection.

Quality of information lies at the very heart and soul of the Situation Awareness process. Without reliable, consistent (in the logical sense) information, the other technologies which drive the process are effectively rendered useless.

There are formalisms that exist for measuring quality of information, effectiveness of information extraction (in certain contexts), effectiveness of learning, and confidence measures for hypotheses. The challenge facing researchers developing large-scale systems that provide Situational Awareness is to find the common tie that make these measures meaningful to one another, providing both local (component-oriented) and global (process-oriented) frames of reference to characterize system performance.

Situation Awareness requires that our architecture be knowledgeable of past experiences to enable awareness of the current state of the world. An implicit assumption involved in the acquisition of this “experience” is that the vast majority of the data to be mined is in free text format. Issues with the processing of natural language present us with a set of challenges, which seem to attach themselves to almost every component of the process. The English language contains many anomalies and exceptions, which serve to confound attempts to statistically characterize it. Developing semantics linking free-text and numeric information is of vital importance to overcoming this obstacle.

Theoretically, the notion of model selection is crucial to the development of effective fusion architectures for situation awareness. Following the “maxim of parsimony” (Occam's Razor), the aim is to develop a system which learns precisely enough information to perform effectively (minimize number of false positives and false negatives).

3 Putting The Pieces Together

Thus far we have discussed many pieces of a large puzzle. To bring things back into perspective, we present a simple flow of the concept as shown in Figure 3. In the concept presented, there are two major flows – a background process and a “real” time process. It should be obvious by now that the concept that we have presented in this paper is model driven. The problem with many such concepts in the past is the existence and construction of such models. The primary focus of the background process is to build and nominate potential models that can be activated. We believe collection, information extraction and data mining to be key technologies in this portion of the concept. For if we cannot build such models, the concept will quickly fall apart. The “real” time

portion of the concept is triggered as new data or information is provided. As this new information enters the system, it is examined for relevancy based on standing profiles. Information that passes this stage is then parsed to extract relevant attributes which are sent on to determine whether the new information is an item of interest. Models are compared and prioritized based probability of activation. Based on this prioritization, a list of possible models can then be provided to such disciplines as Effect-Based Operations (EBO), Predictive Battlespace Awareness (PBA), Model and Simulation to derive Courses of Actions (COAs) and Decision Support Tools.

4 Summary

Today, Situation Awareness is focused on the tactical picture and is reactive, instead of strategic and pre-emptive. Research under the higher levels of Fusion will enable rapid understanding of strategic intent and impact assessment by future strategic planners and thus support Information Dominance. In this paper we have presented an initial approach to acquiring Situation Awareness. We hope that we have introduced some issues related to fusion and clarified what role fusion may play in the future. What is presented here is only a starting point. Work will proceed to bring components together to demonstrate a process that supports Situation Awareness.

References

- [1] Adam, E. C., *Fighter Cockpits of the Future*, Proceedings of 12th DASC, the 1993 IEEE/AIAA Digital Avionics Systems Conference, 318-323.
- [2] Breton, R., Roy, J. and Paradis, S., *A Command Decision Support Interface for Human Factors and Display Concept Validation*, CRDV-TR-2001-218.
- [3] Endsley, M. R., *Toward a Theory of Situation Awareness in Dynamic Systems*, Human Factors Journal, 37(1), pages 32-64, March 1995.
- [4] Endsley, M. R., *Theoretical underpinnings of Situation Awareness: A Critical Review*, In M. R. Endsley, & D. J. Garland (Eds), *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- [5] Endsley, M. R. & Garland, D.J., *Situation Awareness Analysis and Measurement*, Lawrence Erlbaum Associates, Mahawah, New Jersey, USA, 2000.
- [6] Hinman, M. L., *Some Computational Approaches for Situation Assessment and Impact Assessment*, Fusion2002, Annapolis MD, July 8-11, 2002.
- [7] Jensen, D., *Statistical Challenges to Inductive Inference in Linked Data*, Preliminary Papers of the Seventh International Workshop on Artificial Intelligence and Statistics, 1999.
- [8] McGuinness, B., & Foy, J. L., *A subjective measure of SA: The Crew Awareness Rating Scale (CARS)* Proceedings of the first human performance, situation awareness, and automation conference, Savannah, Georgia, October 2000.
- [9] Schank, R. C., and Abelson, R. P., *Scripts, Plans, Goals and Understanding*, Erlbaum, Hillsdale, NJ, 1977.
- [10] Steele, R. D., *The New Craft of Intelligence Making the Most of Open Private Sector Knowledge*, Time Web Site, 2002.
- [11] Steinburg, Alan N., Bowman, Christopher L., and White, Franklin E., *Revisions to the JDL Data Fusion Model*, presented at the Joint NATO/IRIS Conference, Quebec, October 1998.
- [12] Waxman, A.M., Fay, D.A., Rhodes, B.J., McKenna, T.S., Ivey, R.T., Bomberger, N.A., Bykoski, V.K., and Carpenter, G.A., *Information fusion for image analysis: Geospatial foundations for higher-level fusion*, 5th International Conference on Information Fusion, Annapolis, 2002.
- [13] Witten, Ian H., Frank, Eibe and Gray, Jim, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufman Publishers, Oct 1999.
- [14] Zhang, Zhongfei, Salerno, John, Regan, Maureen, Cutler, Debra, *Using Data Mining Techniques for Building Fusion Models*, Proceedings of SPIE: Data Mining and Knowledge Discovery: Theory, Tools, and Technology V, Orlando, FL, Apr 2003, pp. 174-180.