Evaluating Algorithmic Techniques in Supporting Situation Awareness

Dr. John Salerno[a], Dr. Erik Blasch[b], Mr. Michael Hinman[a], Mr. Douglas Boulware[a]
[a]AFRL/IFEA, Air Force Research Laboratory, Rome Research Site, Rome NY USA
[b]AFRL/SNAA, Air Force Research Laboratory, WPAFB, OH USA
John.Salerno@rl.af.mil; Erik.Blasch@wpafb.af.mil; Michael.Hinman@rl.af.mil; Douglas.Boulware@rl.af.mil

ABSTRACT

How well does an algorithm support its purpose and user base? Has automation provided the user with the ability to augment their production, quality or responsiveness? In a number of systems today these questions can be answered by either Measures of Performance (MOP) or Measures of Effectiveness (MOE). However, the fusion community has not yet developed sufficient measures and has only recently devoted a concerted effort to address this deficiency. In this paper, we will summarize work in metrics for the lower levels of fusion (object ID, tracking, etc) and discuss whether these same metrics still apply to the higher levels (Situation Awareness), or if other approaches are necessary. We conclude this paper with a set of future activities and direction.

Keywords: Situation Awareness, Metrics, Evaluation, Precision, Recall, Misassignment Rate, Evidence Recall, Cost Utility, Timeliness, Data to Information Ratio

1. INTRODUCTION

Over the course of the last two decades two separate communities have investigated the challenges in Situation Awareness (SA). Each group has developed its share of models and each of these models has its advantages and disadvantages. Using the Joint Directors of Laboratories (JDL) model [13], the Fusion community has had considerable success developing sensors, sensor management systems, and tracking and identification systems. However, the current tracking and identification systems rely on the user making sense of a collection of vehicle tracks plotted on a map. While this bottom-up model works well for the lower level fusion community, it does little in the way of actually defining SA and Threat Assessment. On the other hand, the Situation Awareness community has been addressing this very issue, but only from a cognitive viewpoint. The majority of the activity in this area has been accomplished by Endsley [5] and has been focused on a pilot's awareness of their environment. Numerous other problem domains have also been identified including asymmetric threat, cyber and homeland security. These domains share certain themes. Within each domain the overarching objective is to make sense out of a glut of data. In addition, observations must be evaluated to determine their (1) importance as well as how they relate to one another and the evolving situation and (2) responsiveness to quickly and accurately recognize the situation. But each domain also has its own unique problems. These problems include the amount of data available, the format of the data, and the amount of time available. Figure 1 provides an overview of the overall SA process.

Figure 1: Summary of SA Goal

As a step in developing the goal of SA, we presented [2, 12] a detailed discussion of the JDL and Endsley's models as well as the motivation for combining the two by using Endsley's work to further define Level 2 (Situation Assessment) of the JDL Model. In Section 2, we introduce two views on how we can measure the effectiveness of SA. We provide a number of examples and conclude this paper with future directions.

2. MEASURING THE EFFECTIVENESS OF SITUATION AWARENESS

Situation Awareness provides the transformation from data to information – it provides *context*. There are a number of ways in which we can measure how effective SA is. Endsley [8] has introduced a technique which she has called "Situation Awareness Global Assessment Technique" or SAGAT, which is a subjective approach. The objective of SAGAT is to measure the effectiveness of the system to provide or display the necessary data in the most effective way in which an operator can develop awareness. Her technique involves immersing the human into a simulated controlled environment. At specified time intervals, the simulation is stopped and the person is asked a series of questions to test their awareness of their environment. While this is a viable approach to measure how well the data has been presented to an operator (visualization), it does not provide us insight into how well various automated SA algorithms and/or processes work.

In this section we propose two additional views. The first attempts to quantify the benefits of SA from a purely data visualization/overload view. We will argue here that SA can significantly reduce data overload by aggregation of objects and/or connecting events to known situations. To make our case, we will present two examples: tactical and cyber and show a significant reduction in data. In the second part of this section we attempt to measure how well our SA system is performing and how well it meets mission requirements. To answer these questions, we will introduce four specific metrics: confidence, purity, cost utility, and timeliness. We provide a detailed discussion, definitions for each metric and an example of its use. However before we begin our discussion, let us review a number of level 1 metrics and see how applicable they are to level 2.

The goal in developing an overarching vision in SA is to assist in the development of an architecture that can provide us with a set of solutions to the data to information problem. As such we are interested in how effective we are in providing a process which takes single objects or events and make sense of them by connecting the dots, not in how well any of the individual components are working. Sensemaking includes comprehending a situation and its impending meaning that affords action. Klein's Recognition-primed decision making model [9] is based on command decisions made by firefighters, which has features of sensemaking in its reliance on past experience. Sensemaking emphasizes that users realize their reality by making sensed information rationally accountable to themselves and others by determining salient meaning of situational patterns. The goal of such a system is to take in evidence of "real" world activity and to map it to the correct situation (as defined through models). In such a system, evidence can be (1) assigned to the correct model, (2) incorrectly assigned to a model, (3) not assigned to any model when it should be (4) or not associated with any model. Figure 2 provides a summary. In addition, the system should provide alerts when any of the models are found in the data. Therefore, the system may also correctly provide an alert, provide a false alert, or fail to provide an alert.

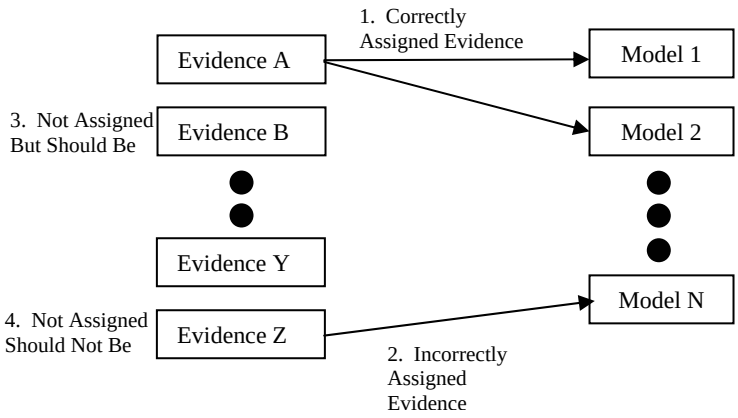

Figure 2: Summary of Possible Assignments of Evidence

The major difference between Data Fusion and SA is that Data Fusion is a bottom up, data driven methodology while SA is a top down, categorization methodology. While most of the data fusion metrics are specific to tracking, they provide a foundation upon which SA metrics may be developed. Under level 1, individual objects are identified and tracked. The primary job of the tracker is then to associate or stitch individual object reports together in an attempt to track single objects. Thus, metrics developed for use in level 1 try to measure how well the tracker can perform this association. Such metrics as Correct Correlation Percentage, Track Fragmentation Percentage, Miscorrelation Percentage, and Ambiguity Percentage have been proposed. These metrics attempt to characterize the ability of the tracker to associate the given contact reports to the correct object/track. Other measures include: Constructed Track Ratio, Track Purity, Track Continuity, Track Length, Track Fidelity, and Collective Score [4].

In SA we have developed a number of models or situations that we are looking for. As such, the primary objective of an SA capability is to associate events or observables to specific situations. An SA capability cannot, on its own, identify new situations [15]. However, if a model is thought of as a track, and evidence is thought of as a contact, then some of the metrics above, or slight variations of them, may be applicable to an SA system. Metrics such as Correct Correlation Percentage, Miscorrelation Percentage, Track Purity and Track Fidelity are such examples. We will come back to specific metrics after we introduce a measure which we call the Data to Information Ratio or DIR.

2.1 The Data to Information Ratio

Can we measure the benefits of SA? One way to look at SA is in the amount of reduction in data that an operator needs to consider. To address this idea, we introduce a measure which we call the Data to Information Ratio or DIR. The purpose of introducing such a metric is an attempt to quantify the value of SA from a purely data point of view. What we are trying to quantify is the reduction of the data required for understanding by associating the individual pieces into a higher level entity or situation. Therefore we define the DIR as:

$$\text{Data to Information Ratio} = \frac{\text{Number of Observations}}{\text{Number of Complex Entities}} \tag{1}$$

The higher this ratio, the greater the reduction and the quicker the user can develop an understanding of the situation. Let us consider two examples; one tactical scenario and one cyber scenario. Consider the situation just after Iraq invaded Kuwait. Figure 3 shows a typical map (left) as a result of sensor reports from such sensors as JSTARS, the map on the right provides what we call a situational map (located online at: http://www.globalsecurity.org/military/world/iraq/orbat-ground-91.htm).

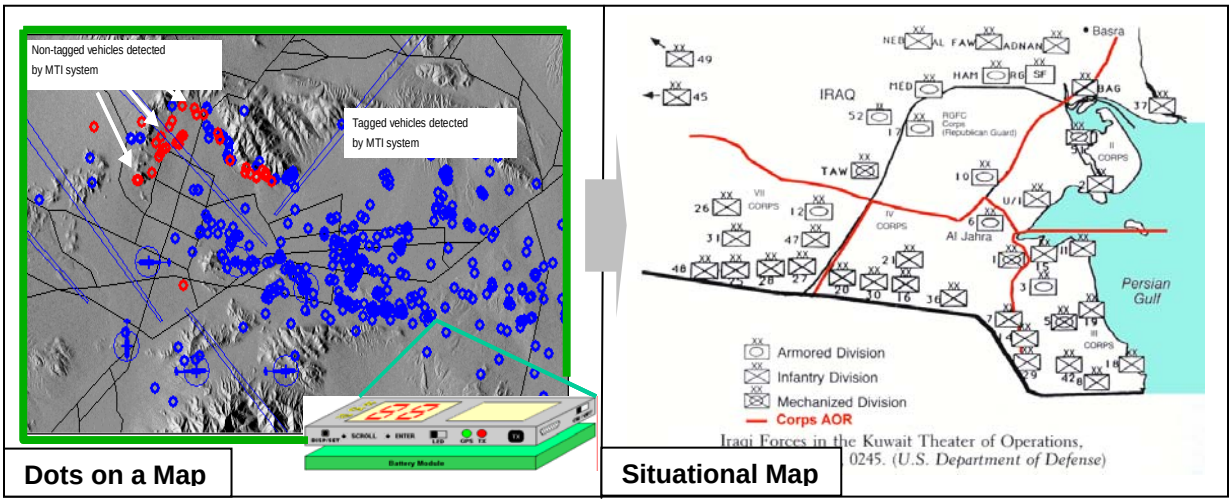

Figure 3: Two Map Views

This situational map shows the approximate location of the units/divisions and provides the operator with a high level view of where the Iraqi military is located. If the operator then desires more details with respect to the individual pieces

of a specific unit they can drill down on a specific area. In this manner, the operator is not overwhelmed with all the dots and can quickly see what is going on. In order to provide this high level view, a number of techniques/algorithms must yet be brought together. For example, clustering techniques may group individual vehicles into units. Template matching techniques may then be used to label each cluster. These templates are formed from known information about the composition of a unit or division within a specified country's military organization.

So how can we compute the DIR in this case? We began with the data from War Online – Middle East Balance (located online at http://www.waronline.org/en/mideast/iraq_army.htm). This source provided us with a breakout of an Armored and Mechanized Division. We could not readily find a similar breakout of the Infantry or Motorized Infantry or Special Forces so in each we took an educated guess. This data is provided in Table 1.

	Infantry Division		**Motorized Infantry Division**	
	Army*	Republican Guard	Army*	Republican Guard
Total Personnel	6,000 men	-	-	14,000 men
Tanks	22	-	-	44
IFV and APC	12	-	-	818
Artillery Pieces	57	-	-	114
Air Defense Artillery	45	-	-	90
MANPADS	25	-	-	50
TOTAL EQUIPMENT	161	-	-	1116

	Armored Division		**Mechanized Division**	
	Army*	Republican Guard	Army*	Republican Guard
Total Personnel	6,000 men	14,000 men	6,000 men	14,000 men
Tanks	122	308	87	220
IFV and APC	236	538	272	622
Artillery Pieces	57	114	57	114
Air Defense Artillery	45	90	45	90
MANPADS	25	50	25	50
TOTAL EQUIPMENT	485	1100	486	1096

*NOTE: Data for the Army Divisions reflects only a 50% staffing

Table 1: Composition of Iraqi Military

Based on the values provided in Table 1 and given the situational map as shown in Figure 3 we can compute the total approximate equipment that should be in place. These values are provided in Table 2.

TOTAL INVASION FORCE SIZE	**Army**		**Republican Guard**	
	Divisions	Equipment	Divisions	Equipment
25 Infantry	25	4025	0	-
3 Motorized Infantry	0	-	3	3348
8 Armored	6	2910	2	2200
5 Mechanized	3	1458	2	2192
1 Special Forces	0	-	1	90
TOTALS (42)	34	8393	8	7770

Table 2: Summary of Iraqi Divisions involved in invasion of Kuwait

Based on Table 2, if we have identified all the equipment correctly (assuming no noise or clutter) we would have a DIR of 16,203/42 = 386. The DIR tells us that if such a capability exists that could automatically cluster individual objects into groups and label these groups based on their makeup of equipment types, it could significantly reduce the amount of clutter on the map and provide better overall awareness. In reality there could be a significant amount of noise or clutter, either making the clustering or identification more difficult. In other cases, SA can also overcome missing, incorrect or incomplete data. We realize that identifying the unit solely on the equipment is in many cases very naive and will not work. At best this approach can only get us down to 1 in 25 Infantry Divisions, 1 in 3 Motorized Infantry Divisions, 1 in

8 Armored Divisions, and 1 in 5 Mechanized Divisions. More data about the group must be ascertained from other sources.

Our next example comes from the cyber domain. Here we can consider two scenarios. The first is where we can perform simple aggregation. For example a typical first step in a potential attack would be the reconnaissance phase. In this phase, the attacker is looking for ways in which they can exploit a machine. One way in which this is done is by launching a port scan. A port scan attempts to locate live hosts and open ports of which the attacker can exploit. As a port scan is being performed, a significant number of attempts are made to a large number of hosts within a network. It would not be uncommon to see 1,000 – 10,000 port scans in a given time period. Simple aggregation in this case can reduce the data overload to the operator. It would also provide them with the idea that an attacker is attempting to learn ways in which the network is vulnerable. In this scenario a DIR of 1,000 to 10,000 would not be uncommon and is even larger than our tactical scenario.

In our first two examples we chose scenarios where we have a large number of objects and few groups. Because of this ratio we expected a large DIR. Will this always be the case? No. Let's look briefly at another example (still in the cyber domain). What happens if the port scan was only one part of a multi-stage attack? For example, other events such as phishing attacks (e.g., a query against a DNS Server or Firewall) or FTP Bounce attacks (e.g., FTP 'Port' Non-secure IP, etc.) can occur. In this case the DIR would be much less. Simple attack models seem to be between 26 and 176 events in size. If this was the case then our DIR would range from 176 to 26 reduction in data.

What we have presented here is a concept in which we have attempted to measure the benefits of SA. In all cases SA will provide a reduction in the amount of data. It is our hope that SA can provide a high level view of what is happening with the ability to drill down into the specific data as needed.

2.2 SA System Metrics

Blasch [1] identified a total of five areas including accuracy, confidence, throughput, timeliness, and cost. Since we believe that throughput is addressed by timeliness we will only address four dimensions. Confidence metrics attempt to measure the level to which the system may be trusted to detect a problematic situation. Measurements of purity attempt to characterize the quality of the information provided in the alerts. Cost metrics may measure the cost of the system, as well as the cost savings generated by the system. Finally, timeliness metrics attempt to measure the temporal aspects of the system's performance with regard to the mission objectives. These metrics were explicitly chosen because of their generic nature and their ability to remain applicable across differing technological approaches. At the most basic level, they should apply to a system that generates alerts for problematic situations in vast amounts of data.

2.2.1 Confidence

The first two performance dimensions capture the system's ability to detect problematic situations. This task is essentially one of classification or retrieval and we therefore defer to the metrics that have been used for years in these areas: precision and recall. Recall indicates the number of problematic situations that were detected and precision indicates the percentage of correctly raised alerts.

$$\text{Recall} = \frac{\text{Number of Correct Situations Detected}}{\text{Number of Situations in Ground Truth}} \tag{2}$$

$$\text{Precision} = \frac{\text{Number of Correct Detections}}{\text{Number of Detections}} \tag{3}$$

Consider the following example. An imaginary system detects three suicide bombing plots and generates three false alarms for suicide bombing plots. The same system also correctly identifies a plot for the biological attack with one false alarm, but fails to detect the dirty bomb plot and generates one false alarm. In this scenario, the system would achieve a 4/7 or 57 % recall rate with 4/9 or 44% precision.

Attack	Ground Truth	Detected	False Alarms
Suicide Bomber	5	3	3
Bio Attack	1	1	1
Dirty Bomb	1	0	1

Table 3: Data for Computing Confidence Example

$$\text{Recall} = \frac{3+1}{5+1+1} = \frac{4}{7}$$

$$\text{Precision} = \frac{3+1}{(3+1)+(3+1+1)} = \frac{4}{9}$$

We can also compute recall and precision for each attack model. For example, the recall for "Suicide Bomber" is 3/6 or 50% and the precision is 3/5 or 60%. We must state explicitly that we have intentionally avoided "accuracy." As many others have noted, the traditional accuracy measure, $\frac{tp+tn}{total}$ assumes an equal cost for false negatives and false positives and becomes increasingly misleading as the class distribution becomes less uniform. We note here a difference between level 1 and 2. In the case for level 1 sensor track measurements have equal cost. Therefore in this case it makes sense to talk about accuracy.

2.2.2 Purity

However, detection and false alarm rates are not enough when it comes to Situation Awareness. Ultimately, an SA system will drive preemptive and responsive actions, and those actions will depend heavily on what the SA system provides. If we envision a system that provides alerts to a user based on evidence then we must consider that the system may correctly alert the user based on some evidence it finds, but it may include irrelevant evidence and/or it may not include relevant evidence. Costs may be associated with either shortfall, and therefore we must include a measure of the quality of the detections as well as the rate of detection. The misassignment rate, shown below, is essentially the same idea as the tracker miscorrelation percentage, and represents precision at the evidence assignment level. Evidence recall is simply the recall at the evidence level.

$$\text{Misassignment Rate} = \frac{\sum_{i=1}^{M} I(m_i)}{\sum_{i=1}^{M} S(m_i)} \tag{4}$$

$$\text{EvidenceRecall} = \frac{\sum_{i=1}^{M} C(m_i)}{\sum_{i=1}^{M} E(m_i)} \tag{5}$$

where M is the number of alerts, or matched models, $I(m_i)$ is the amount of incorrectly associated evidence, $S(m_i)$ is the size or amount of evidence in the match, $C(m_i)$ is the amount of correctly associated evidence for model i, and $E(m_i)$ is the expected amount of evidence for model i. In the case of a false alarm, $E(m_i)$ is 0. Note that these measures are not complements. The evidence recall measures the amount of missing evidence, while the misassignment rate measures the amount of extraneous and incorrect evidence provided. These measures could also be weighted to reflect the importance of certain pieces of evidence.

Consider the example:

	Expected # facts	Total # Facts in Alert	Correct Facts in Alert
Suicide Bombing	5	3	3
Bio Attack	7	6	2
Bridge Attack	0	4	0

Table 4: Data for Computing Purity Example

$$\text{Misassignment Rate} = \frac{0+4+4}{3+6+4} = \frac{8}{13}$$

$$\text{Evidence Recall} = \frac{3+2+0}{5+7+0} = \frac{5}{12}$$

If the dataset contained a single suicide bombing and a biological attack, the above system would have correctly raised alerts for the suicide bombing and the biological attack and would have raised a false alert for an attack on a bridge. This would result in 100% recall and 66% precision. However, if evaluated further one can see that in addition to the poor precision, the system also has very poor alert quality with a missassignment rate of 8/13 or 61.5% and an evidence recall of 5/12 or 41.7%. Imagine if more than half of the evidence in an alert was unrelated and more than half of the pertinent information was missing. Such a system would provide little actionable intelligence and ultimately be distrusted.

2.2.3 Cost Utility

In the real world, however, certain alerts may be more valuable than others. When looked at in this light, detection rate and the quality of the detections may not be enough. Another important measure of an SA system is cost utility. Note that while it would be useful to know the amount of money and time saved by the introduction of automation that is not the intention of this metric. Instead, we define cost utility as a weighted combination of precision and recall based on a notion of cost that is an abstraction of negative consequences. Each problematic scenario in a test could have a positive cost associated with the failure to detect it. Similarly, a negative cost would be associated with each type of false alarm. Within this construct, we can calculate the cost utility of a system with the following formula:

$$\text{Cost Utility} = \frac{\sum_{i=0}^{M} Cost(m_i)}{\sum_{i=0}^{N} Cost(n_i)} \tag{6}$$

where $Cost(m_i)$ is the cost (positive or negative) of alert i, M is the number of alerts, N is the number of situations in the ground truth that should generate alerts, and $Cost(n_i)$ is the positive cost of not generating an alert for situation n_i. This score will actually range from an unbounded negative number to 1. A score of 1 is the optimal score in which no false alarms were generated and an alert was generated for every problematic situation. A score of 0 indicates no savings and while it could be achieved through a combination of false alarms and correct detections it indicates that the overall cost is the same as if nothing was done. A negative number indicates that using the system is actually worse than doing nothing at all. Consider the example below. Detecting a suicide bomber is worth five, but generating a false alarm of a suicide bomber costs 1. Similarly, detecting a biological attack or a dirty bomb is worth forty while false alarms cost ten. A particular dataset contains five suicide bombers, one biological attack and a dirty bomb. An imaginary system detects three suicide bombing plots and generates three false alarms for suicide bombing plots. The same system also correctly identifies a plot for a biological attack with one false alarm, but fails to detect the dirty bomb plot and generates one false alarm. Given the associated costs, this system is shown to save roughly thirty percent of the costs.

Attack	Detection Value	False Alarm Cost	Ground Truth	Detected	False Alarms
Suicide Bomber	5	-1	5	3	3
Bio Attack	40	-10	1	1	1
Dirty Bomb	40	-10	1	0	1

Table 5: Data for Computing Cost Utility Example

$$\text{Cost Utility} = \frac{(3*5)+(3*-1)+(1*40)+(1*-10)+(1*-10)}{(5*5)+(1*40)+(1*40)} = \frac{34}{105} \text{ or } 30.5\%$$

2.2.4 Timeliness

Another important factor for any system in the real world will be time. Certain techniques may require more time than others and thus we would like a measurement of the delay in the system due to the processing time. Presumably, data is continually fed to the system and therefore we must be careful not to penalize the system for not raising an alert when it did not have enough evidence to generate an alert. One solution is to only measure the difference between the time at which the alert was generated and the time stamp of the most recent item of evidence in the alert. However, the time requirements of the mission must still be related to the systems performance.

2.3 Data Set Characterization

Alone, these metrics are not enough. Knowing that a particular system achieved eighty percent recall and ninety-five percent precision, or that a system achieved a cost utility of fifty percent on some dataset may mean nothing. Another class of metrics must be developed to characterize the datasets across multiple dimensions. Message Understanding Conferences demonstrated the need for dataset characterization. In spite of focusing on metrics for information extraction, the conferences missed this important aspect year after year. While participants may have reported slight improvements over the years, these improvements were not grounded in any control condition. The systems were evaluated every year with different datasets that lacked a formal comparison to the previous datasets. With a thorough dataset characterization, the metrics discussed above begin to take on more meaning as they are used with respect to dataset complexity. Capturing the complexity of a dataset across multiple dimensions would not only help in benchmarking progress, but it would also allow the community to more accurately characterize the real world requirements. This may become increasingly important as we continue to rely on classified datasets and researchers without security clearances. In addition, such a characterization could potentially focus research efforts on the specific problem areas. A limited amount of work has been conducted in the area of data set characterization. In signal processing, for example, evaluations are generally done with respect to the signal to noise ratio in the data. In addition, in data mining the problem sets are often characterized with respect to the size of the database and perhaps some information related to the connectedness of items within the database. Unfortunately, dataset characterization for SA systems will be even more difficult because of the breadth of tasks, techniques, and data types the systems require.

3. CONCLUSION

In summary we have proposed a number of measures that can be used to evaluate SA systems. The first, which we called the Data Information Ratio (DIR) attempts to quantify the benefits of SA from a purely data visualization/overload view. We argued here that SA can significantly reduce data overload by aggregation of objects and/or connecting events to known situations. To make our case, we presented two examples: tactical and cyber and showed a significant reduction in data. We also demonstrated that this reduction is highly dependent on the domain.

A second set of metrics was also introduced. In this second set, we attempted to measure how well is our SA system performing and how well does it meet the mission requirements. To answer these questions we introduced four dimensions: confidence (precision and recall), purity (misassignment rate and evidence recall), cost utility, and timeliness. We provide a detailed discussion, definitions for each metric and an example of its use. Table 6 provides a summary of these metrics.

Dimension	Metric	Definition/Purpose
Confidence	Precision	Percentage of correct alerts
	Recall	Probability of detection
Purity	Misassignment Rate	Percentage of evidence incorrectly associated.
	Evidence Recall	Percentage of found.
Cost Utility	Cost Utility	Percentage of cost savings achieved by the system.
Timeliness		Time between event and alert

Table 6: SA System Metrics

In addition, we also emphasized that data characterization metrics must be developed in order to accurately comprehend the systems performance with respect to the difficulty of the task. This paper is only the beginning in developing standardized metrics for SA that affords baselining and improvement evaluation. Much work still needs to be done including large, complex data analysis, human-in-the-loop testing, and subjective and objective SA metric refinement. Other planned activities include the implementation of these metrics to evaluate our existing SA processes in an environment in which we can control and perturb the difficulty of the tasks.

REFERENCES

1. Erik P. Blasch, M. Pribilski, B. Daughtery, B. Roscoe, and J. Gunsett, "Fusion Metrics for Dynamic Situation Analysis", In Proc SPIE 5429, April 2004, pp. 428 – 438.
2. Erik P. Blasch and Susan Plano, "JDL Level 5 Fusion Model "User Refinement" Issues and Applications in Group Tracking", In Proc SPIE Vol 4729, Aerosense, 2002, pp. 270 – 279.
3. J. Brandstadt, M. Kozak, J. Capanna and J. Jones, "Context Metrics for Quantifying Scenario Difficulty when Evaluation Track and Fusion Performance", *Proc NSSDF*, May 1999.
4. J. Brandstadt, M. Kozak, P. Redmond, J. Jones, "A UAV Flight and Sensor Simulator for Generating RADAR and Infrared Detections as Stimuli for Track and Fusion Processors", *Proc NSSDF*, May 1999.
5. Martin Dowd, Final Scientific and Technical Report for Consistent Operational Picture Via Distributed Fusion (COPVDF), AFRL Contract # F30602-97-C-0341, August 14, 1999
6. Mica R. Endsley. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors Journal*, Volume 37(1), pages 32-64, March 1995.
7. Mica R. Endsley. Theoretical underpinnings of Situation Awareness: A Critical Review. Mica. R. Endsley, and D. J. Garland (editors), In *Situation Awareness Analysis and Measurement* (pp. 3-32). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
8. M. R. Endsley, "Measurement of Situation Awareness in Dynamic Systems", Human Factors, 37(1), pgs. 65-84.
9. Klein, G. A. "A recognition-primed decision (RPD) model of rapid decision making," in *Decision Making in Action: Models and Methods*, 138-147, Norwood NJ, Ablex.
10. Klein, G. A. "Using Stories to strengthen our Intuition," presentation, August 2003.
11. R. Macior, M. Kozak and J. Brandstadt, "Multi-Platform GMTI Tracking Exploitation", *Proc NSSDF*, June 2001.
12. Barry McGuinness and Louise Foy, A subjective measure of SA: The Crew Awareness Rating Scale (CARS). In *Proc of the First Human Performance, Situation Awareness, and Automation Conference*, Savannah, Georgia, October 2000.
13. John J. Salerno, Michael Hinman, Douglas Boulware, "Building A Framework for Situation Awareness", In *Proc of the International Society on Information Fusion (ISIF) Conference*, Stockholm, Sweden, June 2004.
14. Alan N. Steinberg, Christopher L. Bowman, and Franklin E. White. Revisions to the JDL Data Fusion Model, *presented at the Joint NATO/IRIS Confer*ence, Quebec, October 1998.
15. C. Taylor, J. Jones, S. Scott, T. Grogan, N. Collins, J. Brandstadt, and M. Kozak, "Long-Term Track Maintenance in AMSTE: Multisensor Fusion, Feature-Aided Tracking, Sensor and Resource Management", *Proc NSSDF*, May 2003.
16. E. L. Waltz and J. Llinas, *Multi-Sensor Data Fusion*, Artech House, Norwood, MA, 1990.
17. E. L. Waltz, "Integrating the Data Fusion and Data Mining Processes", NSSDF04.