World Modeling for Advanced Surveillance Systems

Yvonne Fischer

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
yvonne.fischer@kit.edu

Technical Report IES-20xx-xx

Abstract: The objective of advanced surveillance systems is not only to collect as much sensor data as possible, but also to process and represent it in a meaningful way for supporting situation awareness of a decision maker. However, in today's surveillance systems, there is still a need for information processing methods that meet these higher-level objectives. In this article, the information flow inside of an advanced surveillance system is highlighted and the term situation is discussed with respect to different abstraction levels. Furthermore, several challenges are identified that an advanced system has to address. Therefore, methods selected for information processing should meet these challenges in order to provide a high-level functionality for situation awareness support.

1 Introduction

During the operation of complex systems that include human decision making, acquiring and interpreting information from the environment forms the basis for the state of knowledge of a decision maker. This state is often referred to as situation awareness. The most commonly used definition of situation awareness was provided by Endsley in [End95]:

"Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."

Due to this definition, situation awareness consists of three levels, namely perception, comprehension, and projection, as depicted in Figure 1.1. The first level of situation awareness includes the detection of relevant elements and its characteristics in the environment. These elements are of course domain specific and

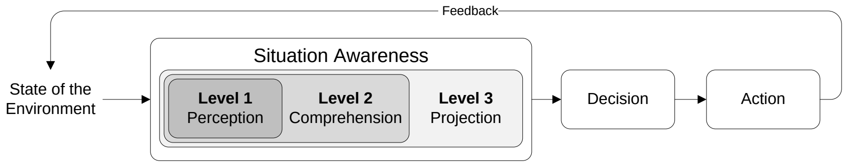

Figure 1.1: The process of dynamic decision making (adopted from [End95]).

their status, their attributes and their dynamics have to be observable by sensorial means. The second level of situation awareness is based on the relevant elements that have been detected on the first level and includes the understanding of the significance of these elements in relation to the operator's goals. The third and highest level of situation awareness is again based on the lower levels and deals with the ability to project future actions of elements in the environment.

Thus, a high level of situation awareness consists of much more than simply collecting information about elements in the environment. It is furthermore a result from the comprehension of its meaning and the projection of future states in order to make decisions on the most favorable actions. Situation awareness is therefore referred to as a mental state or a state of knowledge, whereas the processes to achieve and maintain that state are referred to as situation assessment. As a high level of situation awareness provides the complete knowledge which is necessary for effective decision making, the decision process itself and the performance of actions are separate stages of the dynamic decision making process as illustrated in Figure 1.1.

Endsley described several factors that have a major influence on the decision making process. First, individual factors influence the situation assessment process, for example the operator's abilities, experience, and training. But individuals do not only vary in their information processing mechanisms but also in their expectations and objectives. Other influencing factors can be summarized as system factors which include the system capabilities or the interface design, and also some features of the task environment like workload, stress or complexity.

The concept of situation awareness established by Endsley is applicable in many different domains and it can also be used for advanced surveillance systems. Especially in security-related tasks, like the surveillance of specific areas, decision makers should always have a high level of situation awareness. Situations of interest that take place in surveyed areas are often of a high complexity and dynamic, because they consist of multiple different objects that interact with each other and their activities evolve over time. In such a complex and dynamic environment, the limited capacity of a person's attention is quickly exhausted. The focus of attention is therefore a major limit on situation awareness.

In today's surveillance systems, level 1 situation awareness is highly supported through various heterogeneous sensors and appropriate signal-processing methods for extracting as much information as possible about the surveyed environment and its elements. The challenge of advanced surveillance systems is therefore not only to collect as much sensor data as possible, but also to process and represent them in an intelligent and meaningful way to give a sufficient information support to a decision maker. Or, in other words, to detect and assess complex situations that evolve over time as an automatic support to an operator's situation assessment process. The information overload is then reduced by providing only relevant or task-oriented information, which can be used to guide the focus of attention of a decision maker and allows him to decide and react in a timely and effective manner.

Working with heterogeneous sensors, the theories of multi-sensor data fusion [HM04], [Mit07] offer a powerful technique for supporting situation awareness. A lot of data fusion models have been developed and compared to Endsley's situation awareness model [lBRW07], whereas the most dominant model is the JDL model [SBW99]. However, there is still a need for concepts and methods supporting higher level situation awareness (level 2 and 3) that are able to infer real situations from observed elements in the environment and to project their status in the near future.

The paper is structured as follows. The next section gives an overview of the information flow inside an advanced surveillance system. Section 3.1 deals with a discussion on situational abstraction levels and tries to give a definition of the term situation. Section 3.2 is a first attempt of formalizing situations. In Section 4, several problems concerning automatic situation assessment in surveillance systems are identified. The paper finishes with a conclusion and outlook in Section 5.

2 Information Flow

Regarding data fusion in surveillance systems, the object-oriented world model (OOWM) is an approach to represent the relevant information extracted from sensor signals, fused into a single comprehensive, dynamic model of the monitored area. It was developed in [BEVB09], whereas the basic ideas have been published in [EGB08]. A detailed description of the architecture can be found in [MRV10] and an application of the OOWM for wide area maritime surveillance is proposed in [FB10].

In Figure 2.1, the information flow inside an advanced surveillance system is illustrated, whereas the real world is depicted on the top and the world model, i.e.

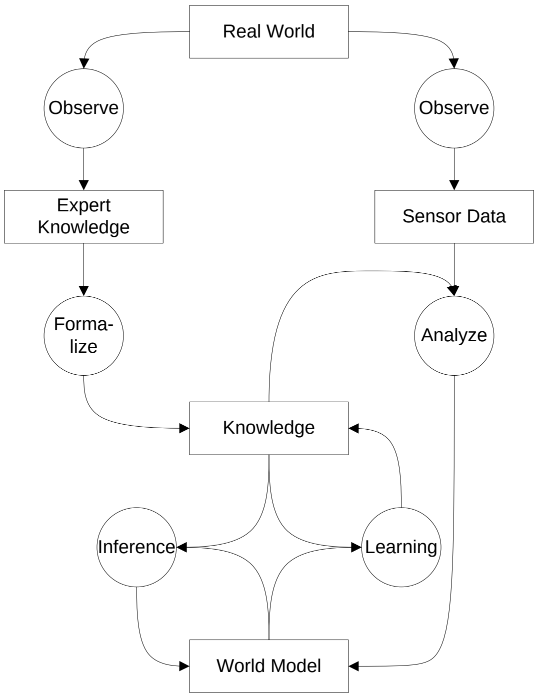

Figure 2.1: Information flow in surveillance systems.

the OOWM, is depicted on the bottom. Rectangular boxes represent aggregates of information and round boxes represent processes. The real world is defined as a spatio-temporal section of the physical world. Relevant parts of it can be observed by humans (or experts) and the result of this process is called expert knowledge. By formalizing the expert knowledge, knowledge-or more precisely-machine readable knowledge is generated.

The physical conditions of the real world can also be observed by appropriate sensors. The sensor data itself represents a spatio-temporal section of the real world and can be analyzed by using knowledge. The analyzed sensor data is transferred as information to the world model. The world model can be interpreted as a representation of the real world and its history, which is generated by using knowledge for analyzing sensor data, or by inference methods. Analyzing sensor data with knowledge includes for example data association and tracking methods, or consistency checks of the world model. Updating the world model with new sensor information is conducted by the inference process.

The knowledge includes all information that is needed for updating the world model by inference or for analyzing sensor data. It builds the basis for the description of relevant aspects of the real world in the world model. The knowledge is of course strongly dependent of the application domain, the context and the task. Furthermore, the knowledge is not static because it can be changed by new information coming from the world model or from expert knowledge. This dynamic aspect of the knowledge is also visualized by the learning process.

3 Higher-Level World Modeling

Regarding situational modeling, several concepts exist in literature. Roy proposed in [Roy01] the concept of situation analysis as a process to provide and maintain a state of situation awareness. He also proposed definitions of situational elements like entities, events and activities. Another refinement of the situational terminology with respect to the JDL data fusion model is given in [Sal07]. The concept of situation management in dynamic systems proposed by Jakobson [JBL07] includes not only the processes of perceiving and recognizing situations, but also the analysis of past situations and the prediction of future situations. In [Ste08], a rough taxonomy of functions related to situation assessment is proposed and a general overview of current approaches to automating this process is given.

3.1 Situational Abstraction Levels

In the revised version of the JDL data fusion model [SBW99], situation assessment (JDL-level 2) is defined as the estimation and prediction of relations among entities. The resulting network of relations among its elements is then referred to as the state of aggregation or the estimated situation. However, there is no formal representation of a situation, as the JDL definition admits any variety of relations to be considered. Types of relations exist at many different levels of abstraction, ranging from quantitative to highly abstract qualitative statements. Therefore, a formal representation of a situation, which fulfills the essential requirements in various application areas, is not easy to define. Situations are characterized mainly by their respective qualitative statements and their representation is therefore strongly dependent on the application domain.

Figure 3.1 shows a general decomposition of a situational description with respect to different abstraction levels. The level of abstraction is determined by the quantity of context information added to the observed element, whereas only relevant

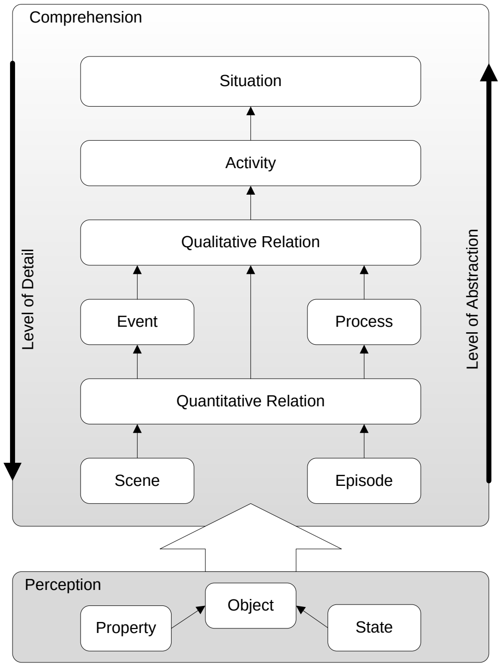

Figure 3.1: Situational Abstraction Levels.

context information is used. The context information consists of knowledge that is not directly observable by sensors, for example expert knowledge. Its content and relevance is determined by the application domain and the task that an operator has to solve. The higher the level of abstraction, the lower is the level of detailed knowledge of a single, observed element. In the following, we will explain the decomposition in detail and give examples for each level of abstraction.

With the focus on surveillance systems, the perception stage includes the acquisition of object information by means of various sensors. Time invariant attributes about an object are summarized as properties and time variant attributes are summarized as the object's state. When observing for example human beings, the results from the perception stage are therefore the person's position and velocity as states and the height as property. This information is the input for the next stage, the comprehension of a situation.

At the lowest level of abstraction, a scene includes all observed objects at a certain point in time. A scene can therefore be interpreted as a snapshot or as a spatial subset of the world's observable objects at a point in time, whereas an episode includes also the time-dimension. An episode is the recording of all observed objects in a period of time (either discrete or continuous) and can therefore be defined as a spatio-temporal subset of the world's observable objects. Note, that at this level of abstraction, no relational aspects between objects are regarded.

The next level of abstraction deals with the description of quantitative relations that can be extracted directly from the information content of a scene or of an episode. Quantitative relations are quantitative statements about the connection between two or more relevant information values, mostly about the attribute values of some objects. The spatial distance measured in meter between two objects is for example a quantitative relation. Note that quantitative relations do not assume that the information values are derived from different objects. Another example of a quantitative relation is therefore the distance that an object has passed between two time points.

In Figure 3.1, special placements between quantitative and qualitative relations are given to events and processes. They can be interpreted as special cases of quantitative relations. An event is defined as the change of relevant object information at a point of time and a process describes the behavior of relevant object information during a time period. For example, the disappearance of an observed person could be tagged as an event or that a person's attribute value, indicating its speed, has changed to zero. A process would be the person's speed value or the direction of its movement over a time period. Events and processes are not limited to a single object. A process between two objects could be the decreasing distance between them and an event between them could be that the distance value of the respective quantitative relation changed to zero.

On the next higher level of abstraction, events, processes and quantitative relations can be summarized to qualitative relations. Detailed knowledge of attribute values of the observed objects goes lost at this level. A qualitative relation is an interpretation of the underlying events, processes and quantitative relations. Examples for qualitative relations are a person that is walking, a person that stops its movement, a person that is moving towards another object, or a person that meets another person.

Qualitative relations are strongly connected to activities. However, we state that activities take place in a longer period of time and are more complex in their construction. As qualitative relations can be interpreted as single and non-decomposable structures, an activity includes also the temporal relationships between them. An activity is a sequence of qualitative events, processes and relations. Temporal relationships of overlapping processes can for example be expressed by Allen's temporal interval logic [All83]. An example for this level of abstraction is a fighting activity between two human beings. The term behavior is often used if the focus is on activities conducted by humans or only by a single object. However, we will use the term activity for this level because it has a broader meaning.

At the highest level of abstraction, there is the situation itself. The human comprehension of a situation can be interpreted as the knowledge of everything of relevance that is going on. Therefore, based on our discussion so far, we give the following definition of the term situation:

A situation at time t is defined as a world state, which is characterized by the collection of relevant activities up to the time t and their interpretation with respect to the context knowledge.

As the world evolves over time, it changes from one state to another. Therefore, the change from one situation to another is due to the change of any activity that is going on or due to a change of the context. As an example, we assume a fighting activity between two humans that is going on so far. Regarding the context, the situation is completely different if the fighting takes place on the street or inside a boxing ring, although the underlying activity is the same.

The situation assessment process can therefore be described as the estimation and interpretation of the relevant state of the real world, which however does not only consist of the recognition of all activities that are going on. Moreover, it also includes contextual conditions like the environment in which an activity is taking place and its aim is to reduce the quantity of information with respect to its relevance.

3.2 Situational Configuration Spaces

In this subsection it is assumed that there are objects in the real world that have been observed. As described above, the objects have properties and states, which we will summarize as attributes. Attributes are for example the existence, the type, the position, the size or the color of an object. Objects can be divided into several classes, based on their type, whereas the number and style of attributes are determined by the object's type. Relations can be temporal or attributive and they can consider several objects. Most of the time only relations between two objects, binary relations are considered. As objects, also relations have attributes like the existence, the type or for example the relative distance between two objects. They can also be divided into classes based on their type.

If we interpret a scene as a snapshot of all observed objects at a certain point in time including their attributes, a scene can be formalized as depicted in Figure 3.2.

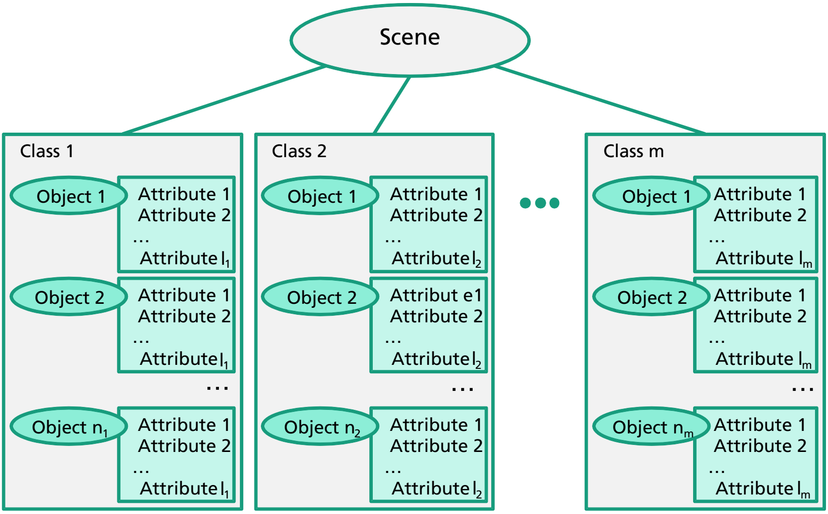

Figure 3.2: Formalization of a Scene.

Therefore, a configuration space K_{Scene} of a scene consists of

- the possible object classes $C_1, \dots, C_m$,
- the number n_i of observed objects of class C_i, $i = 1, \dots, m$, and
- the attributes $A_1^i, \dots, A_{l_i}^i$ of Class C_i, $i = 1, \dots, m$.

The configuration space of a scene can then be defined as

$$\begin{aligned} K_{Scene} &= \prod_{i=1}^{m} \left(\prod_{j=1}^{n_i} \left(\prod_{k=1}^{l_i} A_k^i \right) \right) \\ &= \prod_{i=1}^{m} \left(\prod_{j=1}^{n_i} C_i \right) \\ &= \underbrace{(C_1 \times \dots C_1)}_{n_1} \times \underbrace{(C_2 \times \dots C_2)}_{n_2} \times \dots \times \underbrace{(C_m \times \dots C_m)}_{n_m} \end{aligned}$$

The dimension of the configuration space K_{Scene} is therefore dependent on the

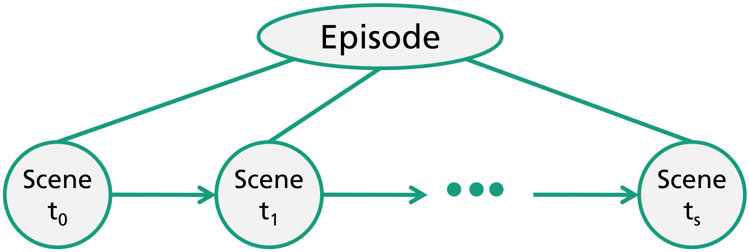

Figure 3.3: Scenes and Episodes.

number of observed objects per class:

$$\dim K_{Scene} = \sum_{i=1}^{m} n_i \cdot l_i.$$

Based on the definition of the configuration space of a scene, we can add the time-dimension for defining the configuration space of an episode, see Figure 3.3. The configuration space $K_{Episode}$ of an episode consists therefore of

- the configuration space of a scene K_{Scene}, and
- the time-sequence $\{t_0, t_1, \ldots, t_s\} := T$.

The configuration space of an episode can then be defined as

$$K_{Episode} = K_{Scene} \times T.$$

The dimension of the configuration space $K_{Episode}$ is therefore

$$\dim K_{Episode} = 1 + \dim K_{Scene}.$$

Situations can be interpreted as episodes enriched with relations and they include therefore higher-level information which is not included in a scene or an episode, see Figure 3.4. The configuration space $K_{Situation}$ of a situation consists therefore of

- the possible relational classes $R_1, \ldots, R_p$,
- the number q_i of relations of the relational class R_i, $i = 1, \ldots, p$,

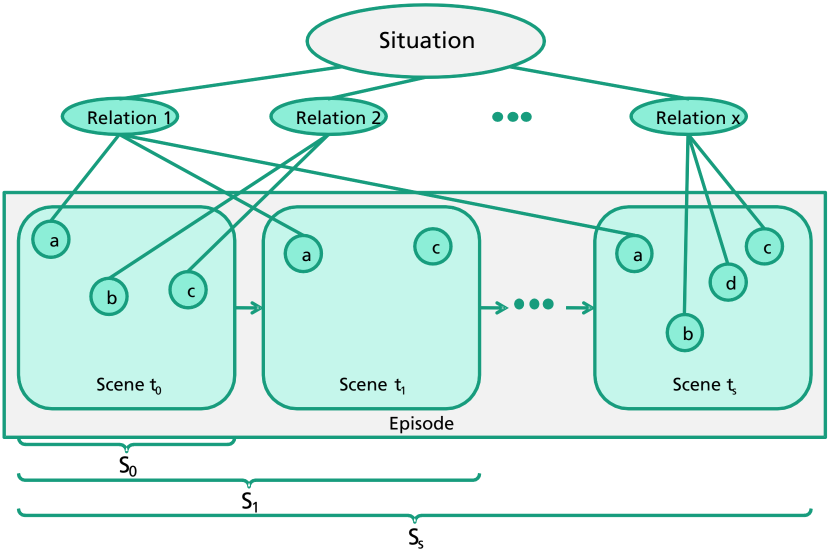

Figure 3.4: Formalization of a Situation.

- the attributes $B_1^i, \ldots, B_{r_i}^i$ of the relational class, and
- the configuration space of an episode $K_{Episode}$.

The configuration space of a situation can then be defined as

$$\begin{aligned} K_{Situation} &= K_{Episode} \times \prod_{i=1}^{p} \left(\prod_{j=1}^{q_i} \left(\prod_{k=1}^{r_i} B_k^i \right) \right) \\ &= K_{Episode} \times \prod_{i=1}^{p} \left(\prod_{j=1}^{q_i} R_i \right) \\ &= K_{Episode} \times \underbrace{(R_1 \times \ldots R_1)}_{q_1} \times \ldots \times \underbrace{(R_p \times \ldots R_p)}_{q_p}. \end{aligned}$$

The dimension of the configuration space $K_{Situation}$ is therefore

$$\dim K_{Situation} = \dim K_{Episode} + \sum_{i=1}^{p} q_i \cdot r_i$$

$$= 1 + \sum_{i=1}^{m} n_i \cdot l_i + \sum_{i=1}^{p} q_i \cdot r_i.$$

The definitions in this chapter are a first attempt of formalizing the term situation by the introduction of a configuration space. The formalizations are straightforward, based on the objects that have been observed in the environment. Obviously, the dimension of the configuration space is quite high, even if only few observed objects are present. The formalizations also show the complexity of defining situations.

4 Challenges of Advanced Surveillance Systems

In this Section we will identify the main problems and challenges of situation assessment in advanced surveillance systems. More advanced systems also support such high-level functions as described in general in [Das08]. Probabilistic methods like hidden Markov models can be used for situation recognition [MDPB09], but are strongly dependent on training data. Also several other approaches have been proposed, for example grammar-parsing detection of abnormal behavior of a person's movement in indoor surveillance [BF10] or logic based approaches for the recognition of human activities [IS10]. In [SNSS10], a heuristic graph matching approach for the identification of meaningful patterns in large volumes of data have been proposed as an enhancement to existing situation assessment methods. In [GGS06], Markov random fields are used to model contextual relationships and maximum a posteriori labeling is used to infer intentions of observed elements.

Mostly, there is a lack of training data, especially for critical situations that an operator wants to detect. For interventional reasons, critical situations have to be detected timely, which means during their development and not only when they are finished. Moreover, the system should be able to deal with uncertain observations, as signal processing methods usually provide estimated feature values and also false detections. The system should be able to deal with incomplete observations, whereas the incompleteness can be of spatial and of temporal type. Spatial incompleteness follows from incomplete sensor coverage, as for example in wide areas it is not possible to continuously observe every part of the environment. Temporal incompleteness follows from spatial incompleteness in the past. As situations evolve over time, it is possible that the beginning of a situation was not observed. Furthermore, the system should be able to predict the situation state in the near future and give a clue to the question: What might happen next?

The main challenges of situation assessment functions are therefore:

- dealing with no training data,
- detection of situations during their development,
- dealing with uncertainties,
- dealing with incomplete observations, as a result of
  - spatial incompleteness due to incomplete sensor coverage,
  - temporal incompleteness due to missing observations in the past, and
- predicttion of developments in the near future.

Due to these challenges, the result of the situation recognition should not be a binary decision if a certain situation is recognized or not. The result should be a Degree of Belief for each template situation, indicating the existence of the underlying and ongoing situation in the real world.

5 Conclusion and Outlook

In this article, the information flow inside of an advanced surveillance system has been highlighted and the term situation has been discussed with respect to different abstraction levels. A first attempt of formalizing the term situation via the introduction of a configuration space has been provided. Furthermore, several challenges has been identified that an advanced surveillance system has to address. Methods selected for situational assessment should meet these challenges in order to provide a high-level functionality for situation awareness support. Further research directions involve the generation of example situations in surveillance applications and the practical realization of various situation assessment methods. The objective is to support the situation awareness of a decision maker as best as possible because in today's surveillance systems, there is still a need for information processing methods that meet the higher-level objectives.

Bibliography

[All83] James F. Allen. Maintaining Knowledge about Temporal Intervals. In *Communications of the ACM*, volume 26, 1983.

[BEVB09] Alexander Bauer, Thomas Emter, Hauke Vagts, and Jürgen Beyerer. Object oriented world model for surveillance systems. In *Future Security: 4th Security Research Conference, Fraunhofer Press*, pages 339–345, 2009.

[BF10] Alexander Bauer and Yvonne Fischer. Task-oriented situation recognition. In *Proceedings of SPIE*, volume 7709, 2010.

[Das08] Subrata Das. *High-Level Data Fusion*. Artech House, Inc, 2008.

[EGB08] Thomas Emter, Ioana Gheta, and Jürgen Beyerer. Object oriented world model for video surveillance systems. In *Future Security: 3rd Security Research Conference, Fraunhofer Press*, pages 315–320, 2008.

[End95] Mica R. Endsley. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[FB10] Yvonne Fischer and Alexander Bauer. Object-oriented sensor data fusion for wide maritime surveillance. In *2nd International Conference on Waterside Security, WSS*, 2010.

[GGS06] Robin Glinton, Joseph Giampapa, and Katia Sycara. A Markov Random Field Model of Context for High-Level Information Fusion. In *9th International Conference on Information Fusion*, 2006.

[HM04] David L. Hall and Sonya A. H. McMullen. *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Inc., 2004.

[IS10] Joris Ijsselmuiden and Rainer Stiefelhagen. Towards high-level human activity recognition through computer vision and temporal logic. In *KI 2010: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, volume 6359, pages 426–435, 2010.

[JBL07] Gabriel Jakobson, John Buford, and Lundy Lewis. Situation Management: Basic Concepts and Approaches. In *Information Fusion and Geographic Information Systems*, pages 18–33, 2007.

[lBRW07] Éloi Bossé, Jean Roy, and Steve Wark. *Concepts, Models, and Tools for Information Fusion*. Artech House, Inc, 2007.

[MDPB09] Daniel Meyer-Delius, Christian Plagemann, and Wolfram Burgard. Probabilistic Situation Recognition for Vehicular Traffic Scenarios. In *Proceedings of the 2009 IEEE international conference on Robotics and Automation*, pages 4161–4166, 2009.

[Mit07] Harvey B. Mitchell. *Multi-Sensor Data Fusion: An Introduction*. Springer, 2007.

[MRV10] Jürgen Moßgraber, Frank Reinert, and Hauke Vagts. An architecture for a task-oriented surveillance system: a service and event based approach. In *Proc. Fifth International Conference on Systems, ICONS*, pages 146–151, 2010.

[Roy01] Jean Roy. From Data Fusion to Situation Analysis. In *Proceedings of the Fourth International Conference on Information Fusion*, 2001.

[Sal07] John J. Salerno. Where's level 2/3 fusion - a look back over the past 10 years. In *10th International Conference on Information Fusion, IEEE*, 2007.

[SBW99] Alan N. Steinberg, Christopher L. Bowman, and Franklin E. White. Revisions to the JDL Data Fusion Model. In *Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE*, volume 3719, pages 430–441, 1999.

[SNSS10] Kedar Sambhoos, Rakesh Nagi, Moises Sudit, and Adam Stotz. Enhancements to high level data fusion using graph matching and state space search. *Information Fusion*, 11(4, October):351–364, October 2010.

[Ste08] Alan N. Steinberg. Foundations of Situation and Threat Assessment. In Martin E. Liggins, David L. Hall, and James Llinas, editors, *Handbook of Multisensor Data Fusion: Theory and Practice*, pages 437–501. CRC Press, 2nd edition, 2008.