A Top-Down-View on Intelligent Surveillance Systems

Yvonne Fischer* and Jürgen Beyerer*†
*Vision and Fusion Laboratory, Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
†Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)
Karlsruhe, Germany
Email: yvonne.fischer@kit.edu, juergen.beyerer@iosb.fraunhofer.de

***Abstract*—In today's surveillance systems, there is a need for enhancing the situation awareness of an operator. Supporting the situation assessment process can be done by extending the system with a module for automatic interpretation of the observed environment. In this article the information flow in an intelligent surveillance system is described and the separation of the real world and the world model, which is used for the representation of the real world in the system, is clarified. The focus of this article is on modeling situations of interest in a human-understandable way and how to infer them from sensor observations. For the representation in the system, concepts of objects, scenes, relations, and situations are introduced. Situations are modeled as nodes in a dynamic Bayesian network, in which the evidences are based on the content of the world model. Several methods for inferring situations of interest are suggested. Following this approach, even high-level situations of interest can be modeled by using different abstraction levels. Finally, an example of a situation of interest in the maritime domain is given.**

***Keywords*-surveillance system; data fusion; situation awareness; situation assessment; probabilistic reasoning.**

I. Introduction

During the operation of complex systems that include human decision making, the processes of acquiring and interpreting information from the environment forms the basis for the state of knowledge of a decision maker. This mental state is often referred to as situation awareness [1], whereas the processes to achieve and maintain that state is referred to as situation assessment. In today's surveillance system, the situation assessment process 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. Using these methods is, of course, an essential capability for every surveillance system in order to be able to observe a designated area and to detect and track objects inside this area. The approach of collecting as much sensor data as possible and extracting as much information as possible from it is termed bottom-up, or also data-driven processing.

However, this approach is not useful for the situation awareness of an operator, because his workload in interpreting all this information will be too high. The challenge of intelligent surveillance systems is therefore not only to collect as much sensor data as possible, but also to detect and assess complex situations that evolve over time as an automatic support to an operator's situation assessment process, and therefore enhancing his situation awareness. The approach of defining and presenting only relevant information about events and activities is termed top-down processing. However, there is a need for concepts and methrfid ods supporting higher level situation awareness, i.e., methods 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. In Section II, an overview of related work is given. As this article follows the top-down approach, the information flow in an intelligent surveillance system is highlighted in Section III. In Section IV, the methods of modeling situations of interest and inferring their existence are explained. In Section V, an example in the maritime domain is given.

II. Related Work

Working with heterogeneous sensors, the theories of multi-sensor data fusion [2] offer a powerful technique for supporting the situation assessment process. A lot of research has been done in combining object observations coming from different sensors [3], and also in the development of real-time methods for tracking moving objects [4]. Regarding data fusion in surveillance systems, the *object-oriented world model (OOWM)* is an approach to represent relevant information extracted from sensor signals, fused into a single comprehensive, dynamic model of the monitored area. It was developed in [5] and is a data fusion architecture based on the JDL (Joint Directors of Laboratories) data fusion process model [6]. Detailed description of the architecture and an example of an indoor surveillance application has been published in [7]. The OOWM has also been applied for wide area maritime surveillance [8].

First ideas of modeling situations in surveillance applications have been presented in our previous work in [9]. For the situation assessment process, probabilistic methods like hidden Markov models can be used, see for example [10]. In [11], Markov random fields are used to model contextual relationships and maximum a posteriori labeling is used to infer intentions of observed elements.

 ISBN: 978-1-61208-184-7

However, most of the methods used for situation assessment are based on machine learning algorithms and they result in models that humans are not able to understand. They are also strongly dependend on training data, which are not always available, especially not for critical situations. The contribution of this work is the modeling approach from a top-down perspective, which tries to model situations from a human perspective, i.e., what an operator wants to detect, and how to link them to methods for automatic interpretation.

III. Information Flow in Surveillance Systems

In surveillance applications, a spatio-temporal section of the real world, a so-called *world of interest*, is considered. The general information flow for intelligent surveillance systems is visualized in Figure 1, wherein information aggregates are represented by boxes, and processes are represented by circles. The information flow is as follows.

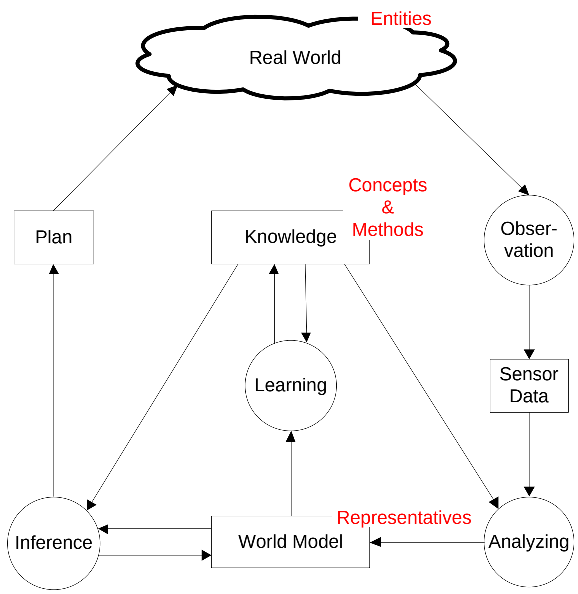

Figure 1. Information flow in a surveillance system represented by information aggregates *(boxes)* and processes *(circles)*.

First of all, all elements in the real world are termed *entities*. By the term entity, not only physical objects are meant, as entities can also be non-physical elements in the real world like relations or the name of a vessel. Thus, entities can represent observable or unobservable elements.

Sensor systems for observing the real world can be of extremely heterogeneous types, e.g., video cameras, infrared cameras, radar equipment, or radio-frequency identification (RFID) chips. Even human beings can act like a sensor by observing entities of the real world. Observing the world of interest with sensors results in sensor data, for example a radar image or a video stream. Sensor data is then analyzed by means of knowledge and the resulting information is transferred to the world model. Analyzing sensor data includes for example the detection and localization of moving vessels at sea from a video stream. Knowledge contains all information that is necessary for analyzing sensor data, for example specific signal-processing methods and algorithms used for the detection, localization and tracking of vessels in video streams.

The world model is a representation of entities in the world of interest and consists therefore of *representatives*. Every representative has a corresponding entity in the real world. The mapping between entities in the world of interest and representatives in the world model is structure-preserving and can therefore be interpreted as a homomorphism. Specific mappings are defined by *concepts* and are part of the knowledge. Concepts are for example used in the analyzing process by defining how an observed vessel is represented in the world model. As the world of interest is highly dynamic and changes over time, the history of the representatives is also stored in the world model. However, as mentioned before, some entities can't be observed directly. Therefore an inference process is reasoning about unobservable (and also unobserved) entities by means of knowledge. A simple inference process is for example to calculate an object's velocity from the last and current position. A more complex inference process would be to estimate if the intention of an observed vessel is benign or adversarial. Doing this way, the world model is always being updated and supplemented with new information by predefined inference processes.

Summing up, knowledge contains all information for analyzing sensor data, updating the world model and supplementing it with new information. Concepts are used for the representation of real-world entities in the world model. Characteristics of the knowledge are of course extremely dependent on the application domain. Additionally, knowledge is not static. The content of the world model can be used for acquiring new knowledge by a learning process, for example structure or parameter learning in graphical models.

To close the loop of the information flow, the result of an inference process could also include a plan of how to act further in the real world. This could be an action plan for an agent, for example to call the police, or a sensor management plan, for example a request for more detailed information from a special sensor.

IV. Modeling and Inferring Situations of Interest

Two problems are faced in this section: First, several concepts have to be defined, which means to define how the real-world entities can be represented in the world model. Second, the inference process has to be defined, which means to define how to reason about non-observable entities like situations or intentions from observed entities.

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-184-7

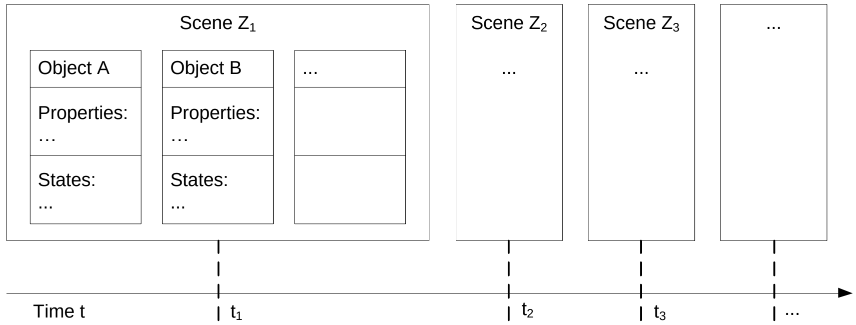

Figure 2. The concept of an object and a scene.

A. Concepts of world modeling

In this section, some basic concepts that can easily be used for the representation of real-world entities are defined. Addressed concepts here are objects, scenes, attributive relations, and situations. However, the world model can easily be extended by defining new concepts, e.g., for activities and events.

The concept of an *object* is defined as a physical entity of the real world. Regarding its spatial position, an object can be mobile, e.g., a vessel, or stationary, e.g., a land border. An object has several attributes, which can be divided into properties and states. Properties are time-invariant attributes, e.g., the length or the name of a vessel. State values can change over time and are therefore time-variant, e.g., the position or the velocity of a vessel. As the representation in the world model also has a memory, which means that the past states of an object are stored, the complete history of the observed object is always available. Furthermore, the representation of an object in the world model does not only include observed attributes, but also inferred ones. For example, based on observed positions of a vessel, the velocity can be inferred. Furthermore, attribute values can be quantitative or qualitative. For example, the absolute position and velocity of a vessel are quantitative attributes, and the attribute value that a vessel is made of wood is a qualitative one.

The concept of a *scene* is defined as the set of all observed and inferred object information at a point in time. A scene can therefore be interpreted as a time-slice, consisting of all objects and their attributes. To include the time aspect, a sequence of scenes can be defined, when the scenes are considered at several discrete points in time. However, a scene does not include any type of relations in an explicit way. This means, that it is for example not explicitly modeled that two vessels are close to each other. But implicitly, of course, this relation can be inferred by the positions of the two vessels. The concept of an object and a scene is visualized in Figure 2.

The *configuration space* is defined by all possibly occuring objects and their attributes. Thus, a scene, which is represented in the world model, can be identified by exactly one point in the configuration space. A sequence of scenes can be interpreted as a trajectory through the configuration space defined by a series of points in time.

The concept of *attributive relations* is defined as a statement about dependencies between at least two different attribute values of one or more objects. Similar to the attribute values of an object, relational values can be quantitative, e.g., the distance of two objects, or they can be qualitative, e.g., two objects are close to each other. Mostly, relational values are inferred, but some can also be observed, e.g., a measured distance by a laser. A relation can also exist between representatives of the same object in different scenes, e.g., the distance an object has covered between the two scenes.

The concept of a *group* is defined as set of object representatives that have the same values for a specific attribute. It is therefore a special case of an attributive relation and can also be interpreted as an equivalence-relation on a specific attribute value. Examples for groups are vessels that have the same size or vessels that are all in a certain area.

The concept of a *situation* is defined as a statement about a subset of the configuration space, which is either true or false. A specific situation of interest exists, if its statement was inferred to be true. Situations are therefore characterized by qualitative attribute values and their truth is inferred based on information in the world model. This means that situations have a higher level of abstraction and the level of detail included in the quantitative attribute values of objects and relations is getting lost. The simplest situation is a statement about qualitative attribute value of an object, e.g., that a vessel is made of wood. There are also situations, which can only be inferred by observing the real world over a period of time, e.g., the situation that a vessel is taking a straight course.

But although situations are also characterized by information collected over a time-period, they only exist at a special point in time. Their existence in the next time-point has to be verified again. However, there are a lot of dependencies between different situations. First of all, situations can be

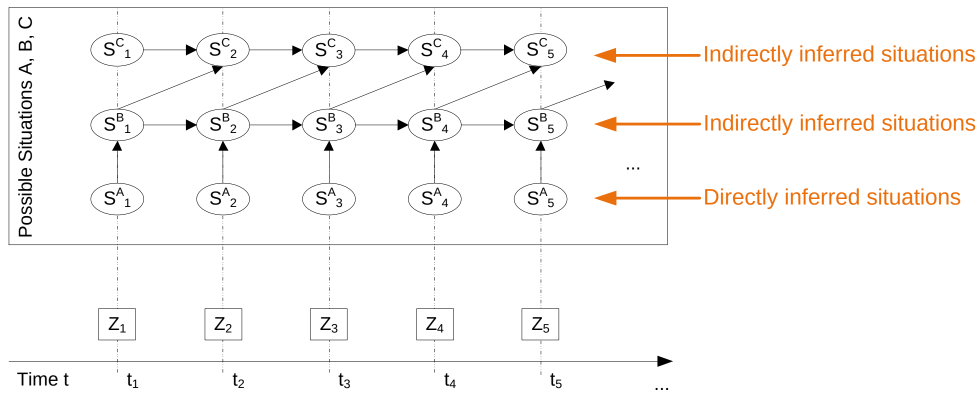

Figure 3. A network of situations, divided into directly and indirectly inferred situations.

inferred from other situations, e.g., if a vessel is heading in a certain direction and has a lot of peaple on board, the inferred situation could be that the vessel is carrying refugees on board. Furthermore, several situations can exist in parallel or the existence of one situation can exclude the existence of another situation. Mathematically, a situation at a time t can be modeled as a binary random variable S_t, such that

$$S_t(\omega) = \begin{cases} 1 & \text{if } \omega \text{ is true,} \\ 0 & \text{if } \omega \text{ is false,} \end{cases} \tag{1}$$

and ω is the statement of the situation of interest. Then, we are interested in the probability, that ω is true, and thus that the situation S_t exists at time t. We write this existence probability as $P(S_t = 1)$, or $P(S_t)$ in short.

For calculating this probability, the aforementioned dependencies between other situations have to be modeled. The following two cases can be distinguished:

- Directly inferred situations: the existence probability $P(S_t)$ can be inferred directly from the information content of a scene (or other concepts like relations or groups)
- Indirectly inferred situations: the existence probability $P(S_t)$ depends on the existence probability of other situations.

This also includes, that the existence probability of an indirectly inferred situation in future can for example be supported by the earlier existence of the situation itself, and the existence probability of a directly inferred situation cannot be supported over time. This concept of a network of situations is visualized in Figure 3.

B. Inferring Situations of Interest

Due to this modeling, the network of situations can be interpreted as a probabilistic graphical model, namely a Dynamic Bayesian network (DBN). In a simple Bayesian network, the basic idea is to decompose the joint probability of various random variables into a factorized form. Random variables are depicted as nodes and conditional probabilities as directed edges. The joint probability can then be factorized as

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | Pa(X_i)), \tag{2}$$

where $Pa(X_i)$ is the set of parents of the node X_i. If $Pa(X_i)$ is an empty set, then X_i is a root node and $P(X_i|Pa(X_i)) = P(X_i)$ denotes its prior probability.

A DBN [4] is defined as a pair $(B_0, 2TBN)$, where

- B_0 defines the prior distribution $P(\boldsymbol{X}_0)$ over the set $\boldsymbol{X}_0$ of random variables, and
- $2TBN$ defines a Bayesian network over two time slices with

$$P(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}) = \prod_{i=1}^{n} P(X_t^i | Pa(X_t^i)), \tag{3}$$

where X_t^i is a node at time slice t and $Pa(X_t^i)$ is the set of parent nodes, which can be in the time slice t or in the time slice $t-1$.

Note that in the definition of a $2TBN$, $Pa(X_t^i)$ is never empty, i.e., every node in time slice t has at least one parent node and therefore the left side of equation (2) differs from the left side of equation (3). An example of a $2TBN$ with 3 nodes in each time slice is shown in Figure 4. The joint probability distribution of a DBN can then be formulated as

$$P(\boldsymbol{X}_{0:T}) = P(\boldsymbol{X}_0) \cdot \prod_{t=1}^{T} \prod_{i=1}^{n} P(X_t^i | Pa(X_t^i)). \tag{4}$$

As we want to model a network of situations by a DBN, the structure of the network has to fulfill the following assumptions:

- Stationarity: the dependencies within a time slice t and the dependencies between the time slices $t-1$ and t do not depend on t.
- 1^{st} order Markov assumption: the parents of a node are in the same time slice or in the previous time slice.

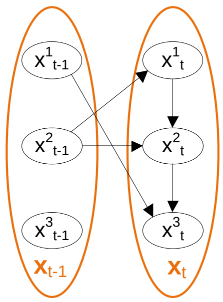

Figure 4. A example of a $2TBN$ defining dependencies between two time slices and dependencies between nodes in time slice t. Note that a $2TBN$ does not define the dependencies between nodes in time slice $t-1$.

- Temporal evolution: dependencies between two time slices are only allowed forward in time, i.e., from past to future.
- Time slice structure: The structure of one time slice is a simple Bayesian network, i.e., without cycles.

For modeling the situational network, the set of situations are divided into the set of directly inferable situations E and the set of indirectly inferable situations S, as described above. The state transition between two time slices satisfies the Markov assumption

$$P(S_t|S_{0:t-1}) = P(S_t|S_{t-1}), \tag{5}$$

and the dependencies between the directly and indirectly inferred situations is defined as

$$P(E_t|S_{0:t}, E_{0:t-1}) = P(E_t|S_t). \tag{6}$$

Due to this dependency, it is assumed that the values of the directly inferred situations are only dependent on the values of the indirectly inferred situations. The joint probability can then be calculated recursively by

$$P(S_{0:T}, E_{1:T}) = P(S_0) \cdot \prod_{t=1}^{T} P(S_t|S_{t-1})P(E_t|S_t). \tag{7}$$

By modeling the network of situations in this way, the following inference calculations are possible:

- Filtering: $P(S_t|E_{1:t})$ gives a solution to the existence probability of a set of situations S at the current time,
- Prediction: $P(S_{t+k}|E_{1:t})$ (with $k > 0$) gives a solution to the existence probability of a set of situations S in the (near) future,
- Smoothing: $P(S_k|E_{1:t})$ (with $0 < k < t$) gives a solution to the existence probability of a set of situations S in the past,
- Most likely explanation: $\mathrm{argmax}_{S_{1:t}} P(S_{1:t}|E_{1:t})$ gives a solution to the most likely sequence of situations $S_{1:t}$.

Due to this modeling, the existence probability of a set of indirectly inferable situations can be calculated in a recursive way at each point in time. A situation is represented in the world model, if the corresponding existence probability is larger than an instantiation-threshold. If the existence probability in the next time step is below a deletion-threshold, it is assumed that the situation doesn't exist any longer and its representation is removed from the world model. This way, it is tried to keep an up-to-date representation of the existing situations of the real world.

V. Application Scenario in the Maritime Domain

For a representation of the world model, the OOWM-system as described in [8] was adapted to the maritime domain. The graphical user interface of the OOWM is depicted in Figure 5. It shows observed vessels at the Mediterranean Sea between the African coast and the island of Lampedusa. Sensor observations are simulated in the system, but they are assumed to be generated by coastal radar systems or signals from the automatic identification system (AIS). In Figure 5, an observed vessel is selected and its observed attributes can be seen on the left side of the user interface. These are exactly the attributes that are stored in the world model and are used for inferring situations of interest.

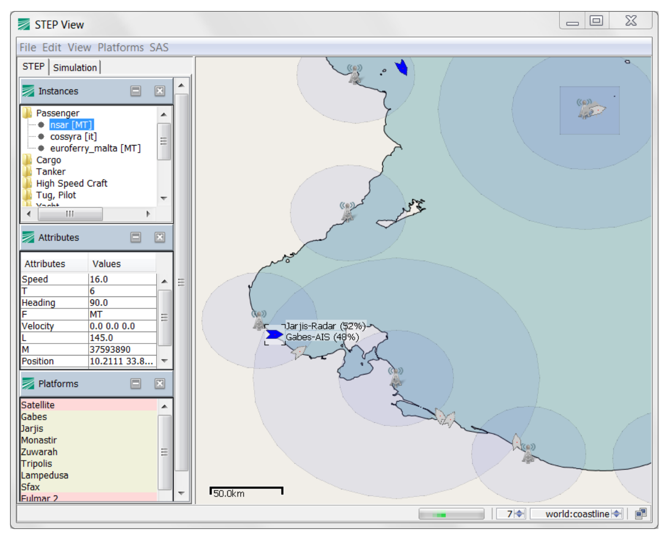

Figure 5. The OOWM system applied to the maritime domain

In the Mediterranian Sea, a situation of interest is the detection of vessels that carry refugees on board. Based on various statements by maritime experts, these vessels have the following (observable) characteristics: They start from the African coast (Tunisia or Libya), are heading towards Lampedusa, take a direct course, and don't send any AIS-Signal for identification. They are either wooden boats or motor-boats, where the wooden boats are slower and smaller than the motor-boats, and the motorboats often go the border, put the refugees into the water and make an emergency call.

An example of a dynamic Bayesian network representing the 3 situations of interest that an observed vessel is a refugee vessel, a wooden vessel, or a motor-vessel is shown

 ISBN: 978-1-61208-184-7

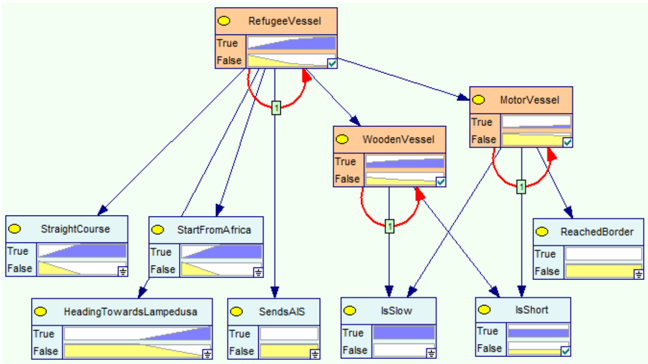

Figure 6. Dynamic Bayesian Network with 3 situations of interest (colored in orange). Temporal arcs over one time slice are marked with a "1" and colored in red.

in Figure 6. The 3 temporal arcs are pointing to the situations of interest themselves, respectively. The resulting existence probabilities (calculated by filtering) for the root node situation (refugee vessel) over 3 time steps are visualized in Figure 7. It can clearly be seen that due to the evidence that has been collected over time, the existence probability of this situation is increasing over time.

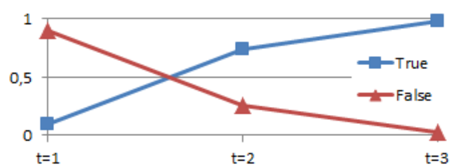

Figure 7. Resulting existence probabilities for the situation that an observed vessel is carrying refugees on board.

The challenges of designing the situational network are to model the structure and to determine the parameters, i.e., the conditional probabilities. Finally, the resulting probabilities for different configurations have to be interpreted (e.g., for the specification of the instantiation- and the deletion-threshold), which is often not straightforward.

VI. Conclusion and Future Work

In this article the information flow in an intelligent surveillance system was highlighted and it was described how situations of interest in surveillance applications can be modeled by concepts. For modeling a network of situations, the framework of dynamic Bayesian networks is suggested, in which the values of the directly inferable nodes are based on the content of the world model. This modeling fulfills the requirements resulting from the definition of situations and allows the application of efficient inference methods. An example of a situation of interest in the maritime domain was given. By extending the surveillance system with such a module for automatic interpretation of the observed environment it is able to support the situation assessment process of an operator and thus enhances his situation awareness.

Future work includes an experimental evaluation of the proposed method and an investigation on supporting the human operator in designing a situational network without having a detailed knowledge of the underlying method. Also the real-time capability of the proposed method when using a large amount of data has to be investigated.

Acknowledgment

This research is partially supported by the EU-FP7-Project WIMA^{2}S (Wide Maritime Area Airborne Surveillance), see http://www.wimaas.eu (last access: 11.12.2011).

References

[1] M. R. Endsley, "Towards a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 11, pp. 32–64, 1995.

[2] D. L. Hall and S. A. H. McMullen, *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Inc., 2004.

[3] M. Baum, I. Gheta, A. Belkin, J. Beyerer, and U. D. Hanebeck, "Data association in a world model for autonomous systems," in *Proc. of the 2010 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2010)*, 2010, pp. 187–192.

[4] A. Dore, M. Soto, and C. S. Regazzoni, "Bayesian tracking for video analytics: An overview," *IEEE Signal processing magazine*, vol. 27, no. 5, pp. 46–55, 2010.

[5] A. Bauer, T. Emter, H. Vagts, and J. Beyerer, "Object oriented world model for surveillance systems," in *Future Security: 4th Security Research Conference*. Fraunhofer Press, 2009, pp. 339–345.

[6] A. N. Steinberg, C. L. Bowman, and F. E. White, "Revisions to the JDL data fusion model," in *Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE Vol. 3719*, 1999, pp. 430–441.

[7] J. Moßgraber, F. Reinert, and H. Vagts, "An architecture for a task-oriented surveillance system: A service- and event-based approach," in *Fifth International Conference on Systems (ICONS 2010)*, 2010, pp. 146–151.

[8] Y. Fischer and A. Bauer, "Object-oriented sensor data fusion for wide maritime surveillance," in *2nd International Conference on Waterside Security,IEEE,(WSS 2010)*, 2010, pp. 1–6.

[9] Y. Fischer, A. Bauer, and J. Beyerer, "A conceptual framework for automatic situation assessment," in *IEEE First International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA 2011)*, 2011, pp. 234–239.

[10] D. Meyer-Delius, C. Plageman, and W. Burgard, "Probabilistic situation recognition for vehicular traffic scenarios," in *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*, 2009, pp. 459–464.

[11] R. Glinton, J. Giampapa, and K. Sycara, "A markov random field model of context for high-level information fusion," in *9th International Conference on Information Fusion*, 2006, pp. 1–8.

