World Modeling for Autonomous Systems

Andrey Belkin[1], Achim Kuwertz[1], Yvonne Fischer[1] and Jürgen Beyerer[1,2]

[1]*Karlsruhe Institute of Technology (KIT), Institute for Anthropomatics (IFA),*
Vision and Fusion Laboratory (IES)
[2]*Fraunhofer Institute of Optronics, System Technologies*
and Image Exploitation (IOSB)
Germany

1. Introduction

Since ancient times mankind is building tools for facilitating survival and work. With time, the tools became complex and exquisite and even obtained consciousness and limited reasoning. The recent developments in computing and robotics led to creation of intelligent systems that are autonomous. The purpose of such *autonomous systems* is to function in different environments stand-alone according to some predefined strategy, e.g. protecting some areas or rescuing people in hazards. The intelligent autonomous systems were defined within the *IBM Autonomic Computing Initiative* in IBM (2001), Kephart & Chess (2003) with the following required abilities:

- *Self-Configuration* – ability to configure and reconfigure itself under internal and external changes;
- *Self-Optimization* – ability to optimize its working;
- *Self-Healing* – ability to discover and correct faults;
- *Self-Protection* – ability to identify and protect itself;
- *Self-Awareness* – ability to know its components, current status, ultimate capacity and connections to other systems;
- *Environment-Awareness* – ability to perceive and know the surrounding environment and context.

An operation of intelligent autonomous systems requires handling of *reactive* and *proactive* activities. The reactive behavior is required for immediate reactions on changes, such as bipedal locomotion balancing, avoiding direct hazards or grasping deformable items. The proactive analysis is required to perform planning and sophisticated decision making, e.g. helping people in a household. Such activities imply so-called constant *situation awareness* that demands perception of changes in the surrounding environment, comprehension of their meaning and projection of their status in the future as defined in Endsley (1995).

In practice, reactive and proactive handling involves complex information acquisition, processing and management. An example of a complete information workflow is presented in Fig. 1. At the beginning, a (heterogeneous) sensors system acquires data from the surrounding environment, called *world of interest* (WoI). The gathered data concerns WoI *entities*: topology features, things, persons, and relations. This data is processed by sophisticated extraction
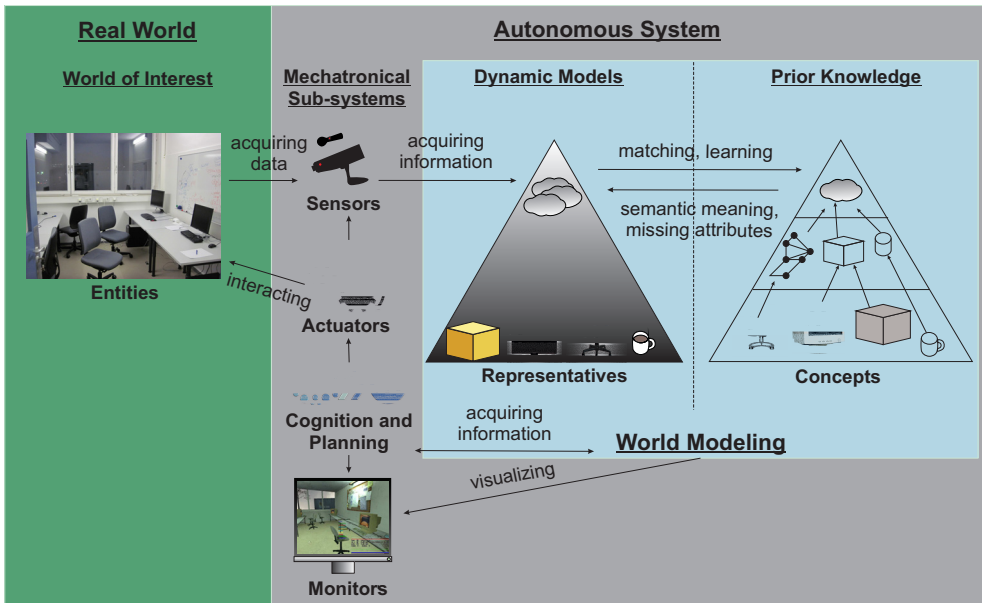
Fig. 1. Information workflow of autonomous systems.

methods in order to obtain as much information about the surrounding environment as possible and then fuse into *dynamic models* within the world modeling system. The world modeling system contains the knowledge about the environment and the autonomous system itself, necessary for the expected autonomous operation, and serves as a global information hub for all mechatronic sub-systems. The dynamic models contain up-to-date description of the WoI, expressing each real-world entity with virtual *representative*. Since such descriptions are created from sensory information, they lack semantic meaning and pre-defined characteristics expected for their type. In order to gain missing semantics and attributes, each representative can be matched and assigned to *prior knowledge concepts*. The content of the world modeling system is available to other mechatronic sub-systems, as, for example, monitors or cognition and planning modules. Based on the world modeling information cognition and planning modules assess the present situation and future states, make decisions and compound plans about future actions. These actions are then performed by actuators that interact with the world of interest. In the case of repeating properties or dependencies, the autonomous system can find out common patterns and try to extend the prior knowledge. The extension of prior knowledge is performed for automatic recognition of previously detected patterns and more efficient environment description. For this, a balance between description length of concepts and description length of dynamic models given prior concepts has to be empirically found.

The world modeling system serves as a virtual demonstrative model of the environment for the whole autonomous system. In order to perform efficiently, its following features are to be considered:

- Slim *symbolic representation* of the environment information;
- *Gathering station* for all the acquired information;

- *Interpretability* for humans and machines;
- *Object-orientedness*;
- Universal *probabilistic uncertainty handling*;
- Common *information hub* for all autonomous system components;
- Handling of *unknown objects*;
- Extensible concepts under *Open World* assumption.

More detailed operation of world modeling system is presented in the following sections: the Section 2 describes the relation between the environment and the autonomous system and defines the modeling domains. The Section 3 covers main aspects of fusing sensory information within the world modeling system. The detailed discussion of the world modeling system, including the mathematical concepts of parameter estimations is given in Section 4. Methods for autonomous prior knowledge extension are highlighted in Section 5. Some examples of practical realization of the described architecture and methods are given in Section 6 followed by conclusion and outlook.

2. Domains of the world modeling

During the operation of autonomous systems that have to make decisions, the process of acquiring and interpreting information from the environment forms the basis for the state of knowledge of such systems. As defined in Endsley (1995), this process is known as situation assessment. In today's autonomous systems, situation assessment 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. Theories and methods of multi-sensor data fusion, e.g., as described in Hall & McMullen (2004), offer a powerful technique for combining information from different sources. The object-oriented world model (OOWM) is an approach to represent relevant information extracted from the sensor signals, fused into a single comprehensive, dynamic model of the monitored area. Bauer et al. (2009) developed a data fusion architecture based on the JDL (Joint Directors of Laboratories) data fusion process model, see Steinberg et al. (1999). An application of the OOWM for wide area maritime surveillance is presented in Fischer & Bauer (2010). Another application is to provide humanoid robots with a memory structure based on a complete Bayesian network, see for example Baum et al. (2010) or Belkin (2009).

However, the challenge of intelligent systems that act in an autonomous way is not only to collect as much sensor data as possible, but also to detect and assess complex situations that evolve over time. First ideas of modeling higher level entities like situations have been presented in Fischer et al. (2011). For the situation assessment process, probabilistic methods like hidden Markov models can be used (see Meyer-Delius et al. (2009)), but they are strongly dependent on training data. In Gheţa et al. (2010), a Bayesian method for the association of observations to objects is presented, and in Glinton et al. (2006), Markov random fields are used to model contextual relationships and maximum a posteriori labeling is used to infer intentions of the observed entities. In this section, we will clarify the separation of the real world and the world model represented in the system and introduce the concepts of objects, scenes, relations, and situations, which are necessary for the internal representation of the observed entities in the environment.

2.1 Information flow in autonomous systems

In applications where autonomous systems are used, a spatio-temporal section of the real world, the so-called world of interest. The general information flow inside such systems is visualized in Fig. 2, in which information aggregates, i.e. the stores of several information, are represented by boxes and processes are represented by circles. The information flow is as follows. The world of interest consists of entities and can be observed by sensors. Sensor systems can be of extremely heterogeneous types, e.g., video or infrared cameras, radar equipment, or RFID-chips. Even human beings can act as a sensor by observing entities of the real world. Observing the real world with sensors results in *sensor data*, for example a radar image or a video stream. Sensor data is then analyzed by means of existing knowledge and the resulting information is passed to the world model. The existing knowledge contains all information that is necessary for analyzing sensor data, for example specific methods and algorithms used for the detection and localization of people in video streams.
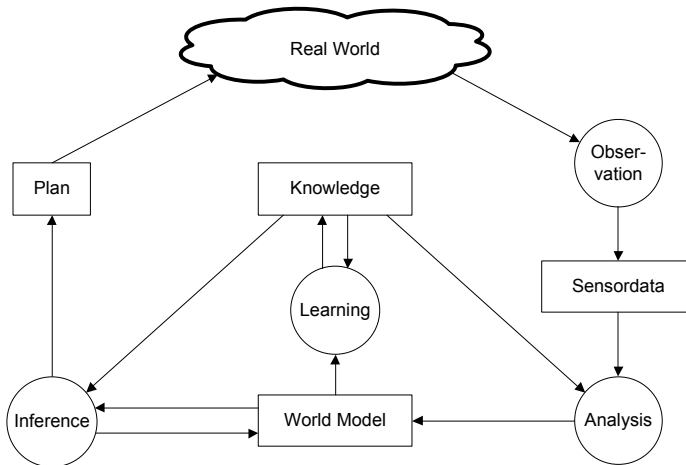

Fig. 2. General information flow in autonomous systems represented by information aggregates *(boxes)* and processes *(circles)*.

The *world model* is a representation of entities from the world of interest and contains 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 termed as concepts and are part of the knowledge. Concepts are defined at several levels of abstraction, e.g., there are concepts for objects, scenes, relations, and situations. Concepts are used for example in the analyzing process by defining how an observed person 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, entities are not always directly observable. For example the intention of a person could also be an entity and be represented in the world model. As some entities cannot be observed directly, an inference process is reasoning about unobservable (and also unobserved) entities by means

of knowledge. Therefore, the world model is always being updated and supplemented with new information by the inference process.

Summing up, knowledge contains all the information necessary for analyzing sensor data, updating the world model and supplementing it with new information. Concepts are used to describe the possible real-world entities in the world model. Characteristics of 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 can also include a *plan*, of how to act further in the real world. This could be an action plan for an agent, e.g., to call the police, or a sensor management plan, e.g., a request for more detailed information from a specific sensor.

2.2 Concepts for world modeling

We will now introduce the concepts for objects, scenes, relations, and situations. The concept of an *object* is defined as a physical entity of the real world. An object can be (spatially) movable (e.g., a person) or stationary (e.g., a room). An object has several attributes, which can be divided into properties and states. Properties are time-invariant attributes, for example the height or the name of a person. State values can change over time and are therefore time-variant, for example the position or the velocity of a person. The representation of an object in the world model includes not only observed attributes, but also inferred ones. For example, the velocity can be inferred based on observed positions of a person. Furthermore, attribute values can be quantitative or qualitative. For example, the absolute position and the velocity of a person are quantitative attributes, and the attribute value that a person is smiling is a symbolic one.

By the concept of a *scene*, we define 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, we also speak of a sequence of scenes, when the scenes are considered at several discrete points in time. However, a scene does not include any type of relations. The concepts of a scene and a sequence of scenes are visualized in Fig. 3. By the concept of *relations*, we mean a statement about dependences between at least two different attribute values of one or more objects. Relational values can also 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 distance measured 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 *situation* is defined only on the symbolic level. Thus, they have a higher level of abstraction, but the level of detail of the quantitative attribute values of objects and relations is getting lost. Situations are therefore characterized by symbolic attribute values. The simplest situation is a symbolic attribute value of an object, e.g., a person is smiling. There are also situations, which can only be inferred by observing the real world over a period of time, e.g., a person is dancing or two persons have a conversation. And, of course, situations can be inferred from other situations, e.g., if some persons are dancing and some persons are drinking beer, the inferred situation could be that there is a party going on. Although situations are also characterized by the information collected over a period of time
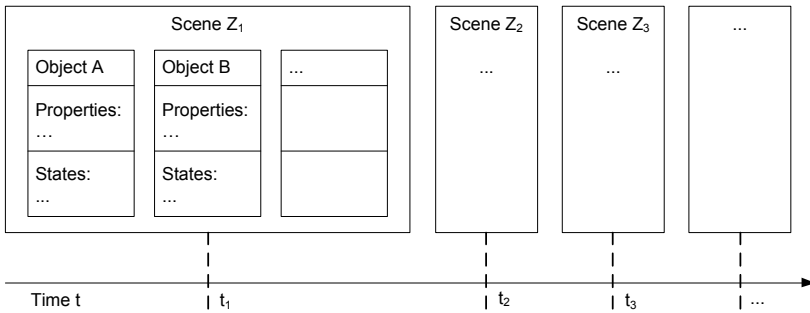
Fig. 3. The concepts of a scene and a sequence of scenes.

they only exist at a point in time. Their existence in the next time-point has to be verified again. However, the earlier existence and also the collection of further evidence over time can support the existence of a situation in the future. The concepts of a relation and a situation are visualized in Fig. 4.
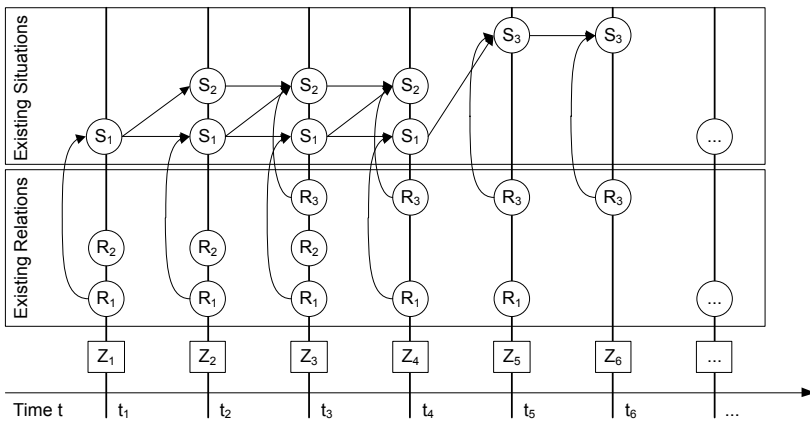

Fig. 4. The concepts of a relation and a situation.

3. Information fusion

An autonomous system perceives the surrounding environment with sensors. The acquired data are processed, analyzed and passed to a world modeling system, which contains a description of the current state of the environment. The new sensory information is fused into the existing environment description by matching and updating relevant representatives in the model. Since information from sensors contains uncertainty, this information matching and description update is not trivial. Before constructing such mechanisms, a suitable information and uncertainty representation has to be defined.

3.1 Information representation

There are many ways to represent information in the context of world modeling (see Belkin (2009)). The simplest way is to describe each parameter with a single numerical value (e.g. "*36.65*") or a fact value (e.g. "*Dog*"). However, this method does not reflect uncertainties, emerging from measurements always limited by sensor acceptance and efficiency. Accounting for uncertainty (e.g. keeping error values) allows for a more faithful representation of measured values. The next level of generalization of the information representation accounts for relations between parameters. The uncertainty, however, can be represented more accurately with probability distributions over possible values, e.g. by means of *Degree-of-Belief* (DoB). In the Fig. 5, there is a typical temperature DoB distribution of an object. The parameter relations can be incorporated into distributions by joint DoBs for groups of correlated attributes (e.g. joint DoB of type and temperature). In this case, the DoB combines discrete and continuous distributions into one common frame. The most general level of the information representation is then the joint DoB distribution for all modeling attributes that typically has combinatorially growing complexity and computational costs. The generalization hierarchy is shown in Fig. 6.
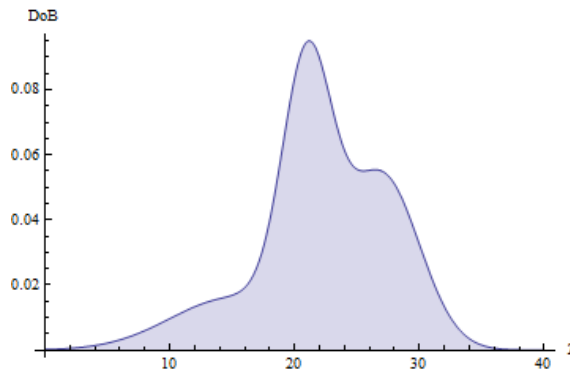

Fig. 5. A Degree-of-Belief distribution for the attribute "temperature".

3.2 Information life cycle

The sensory information has to be fused into the world model in order to update the current environment description. When a new entity is observed in the environment, a corresponding representative has to be created in the dynamic model. This can be achieved by the calculation of the posterior probability that from one hand a new entity has been observed at some time step and the other hand the observed entity exists, exceeding some *instantiation threshold*. If the entity is already present in the model with a representative, the new information is merged into the existing description by means of Bayesian fusion, raising the existence probability. In the absence of new measurements related to an entity, the aging mechanism changes the DoB distributions according to the *maximum entropy principle*, i.e. the uncertainty of the information increases. In particular, the uncertainty increase leads to the existence probability reduction. If the existence probability falls below some *reconfirmation threshold*, the world modeling system signals the need to re-observe the entity. The representative is deleted from the model, if the maximum of the DoB distribution for the existence goes below the *deletion threshold*, in order
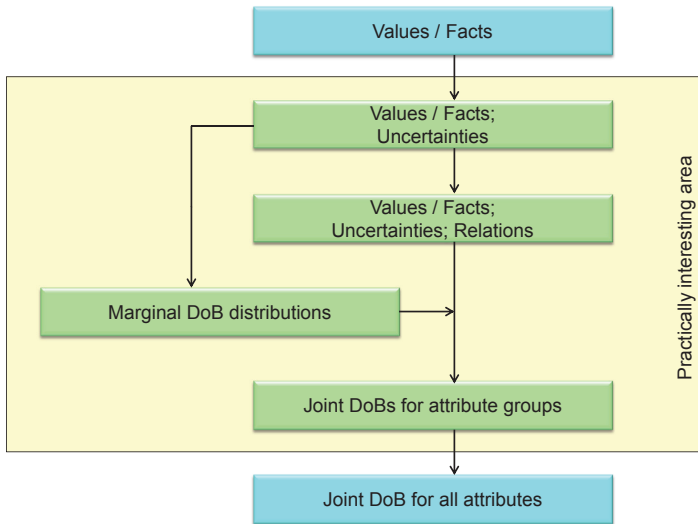
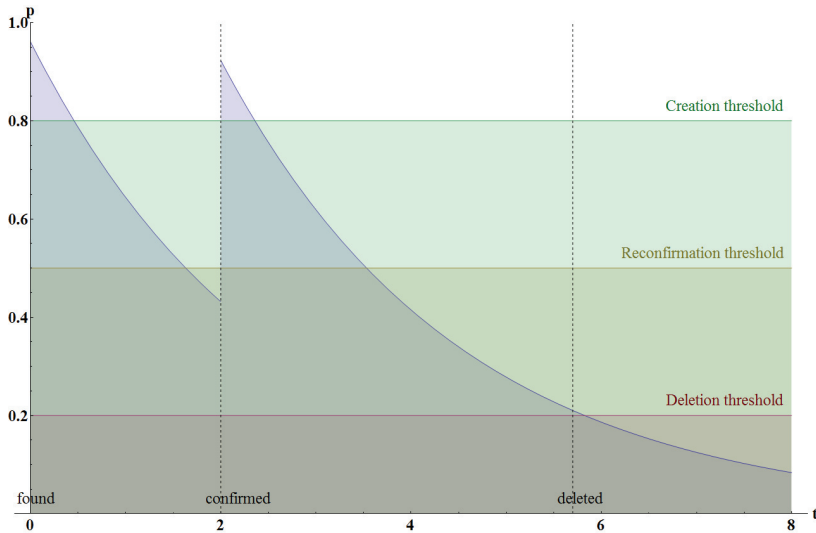Fig. 6. The hierarchy of generalizing the representation of values.

Fig. 7. An example of the existence probability life cycle.

to reduce the computational complexity of the whole model. The aging mechanism can be applied *on demand*: the existence probability of a representative is calculated as soon as the representative is requested by the system. The existence probability life cycle is presented in Fig. 7.

3.3 Bayesian framework

The management of the information life cycle requires the formal definition of observations, a mechanism to associate observations to representatives, and procedures to information update. Such a formal description is developed within the *Bayesian framework* presented briefly below and is given in details in Belkin (2010), Baum et al. (2010).

A modeled entity i at time step k is represented as a DoB distribution $p(e_k^i, \underline{a}_k^i)$, where $e_k^i \in \{0,1\}$ denotes whether the entity i exists and $\underline{a}_k^i := \left[a_k^{i,1} \ \ldots \ a_k^{i,n_a} \right]^{\mathrm{T}}$ denotes the vector of n_a descriptive attributes.

The fusion of an observation into the world model requires some *observation model*, which connects observations to representatives on some probabilistic basis. This connection can be described with the connection probability $p(d_k|\underline{e}_k)$, where $d_k \in \{0, 1 \ldots, N_k\}$ denotes the representative, N_k is the number of representatives, and $\underline{e}_k := \left[e_k^1, \ldots, e_k^{N_k} \right]^{\mathrm{T}}$.

The observation vector can be written as $\underline{y}_k^i := \left[\underline{y}_k^{i,1}, \ldots, \underline{y}_k^{i,n_y} \right]^{\mathrm{T}}$ composed of observations for each given attribute (totally n_y).

After the association of observations, the existing description is updated in *the prediction and the update steps*. The prediction step propagates representatives from the time point $k-1$ to k, resulting in the predicted existence $p(e_k^i = 1|\underline{\hat{y}}_{1:k-1})$ and the predicted attributes $p(\underline{a}_k^i|e_k^i = 1, \underline{\hat{y}}_{1:k-1})$, where $\underline{\hat{y}}_k$ is an observation on the time step k and $\underline{\hat{y}}_{1:k-1}$ is a set of all observations till the time step $k-1$. This propagation increases the entropy of DoB distributions and, thus, the uncertainty. The predicted existence probability can be defined by a *Hidden Markov Model*

$$p(e_k^i = 1|e_{k-1}^i = 1) = \beta^i, \tag{1}$$

with $\beta^i \in (0; 1]$.

The existence probabilities can then be predicted as

$$p(e_k^i = 1|\underline{\hat{y}}_{1:k-1}) = \beta^i \cdot p(e_{k-1}^i = 1|\underline{\hat{y}}_{1:k-1}). \tag{2}$$

The attribute probabilities can be predicted based on the *Chapman-Kolmogorov equation* as

$$p(\underline{a}_k^i|e_k^i = 1, \underline{\hat{y}}_{1:k-1}) = \int p(\underline{a}_k^i|\underline{a}_{k-1}^i) \cdot p(\underline{a}_{k-1}^i|e_k^i = 1, \underline{\hat{y}}_{1:k-1}) d\underline{a}_{k-1}^i, \tag{3}$$

with a propagation model $p(\underline{a}_{k+1}^i|\underline{a}_k^i)$ for entity attributes.

The update step merges the predicted state with the associated observation. The update attributes are given by

$$p(\underline{a}_k^i|e_k^i = 1, \underline{\hat{y}}_{1:k}) = \sum_{d_k} p(\underline{a}_k^i|e_k^i = 1, \underline{\hat{y}}_{1:k}, d_k) \cdot p(d_k|e_k^i = 1, \underline{\hat{y}}_{1:k}), \tag{4}$$

where the first factor is the posterior distribution

$$p(\underline{a}_k^i|e_k^i = 1, \underline{\hat{y}}_{1:k}, d_k) \propto p(\hat{\underline{y}}_k^{d_k}|e_k^i = 1, \underline{a}_k^i) \cdot p(\underline{a}_k^i|e_k^i = 1, d_k, \underline{\hat{y}}_{1:k-1}). \tag{5}$$

The updated existence is given by

$$p(e_k^i = 1|\underline{\hat{y}}_{1:k}) = \sum_{d_k} p(e_k^i = 1, d_k|\underline{\hat{y}}_{1:k}). \tag{6}$$

3.4 Kalman filter and its extensions

In the case of linear dependence of attributes on observed parameters and Gaussian measurement noise, modeling attributes can be updated by the *Kalman filter* (KF) (cf. Bar-Shalom & Fortmann (1988)) and its extensions. The KF takes multiple sequential measurements of some parameters and combines them in order to evaluate the entity's state at that time instant, resulting in a better estimate than obtained by using one observation alone. During the fusion of multiple measurements, a system's dynamics model (e.g. physical laws of motion), known control inputs, and covariances (a measure of the estimated uncertainty) are considered in the prediction step. In the update step, the filter averages a prediction of an entity's state with a new observation based on a weighted average. The weights are calculated from the covariance, estimated on the prediction step, and assure that the more exact values (i.e. with lower uncertainty) have a greater influence on the state estimate.

More specifically, the Kalman Filter model assumes the following evolution of the (unobservable) true state of the system $\underline{\tilde{a}}_k$ (index i is omitted for clarity) from time step $k-1$ to step k and the description of the observed values $\underline{\hat{y}}_k$ at time step k:

$$\underline{\tilde{a}}_k = F_k\underline{\tilde{a}}_{k-1} + B_k\underline{u}_k + \underline{w}_k, \tag{7}$$

$$\underline{\hat{y}}_k = H_k\underline{\tilde{a}}_k + \underline{v}_k, \tag{8}$$

where F_k is a propagation matrix, B_k is a control-input model matrix, $\underline{u}_k$ is a control vector, $\underline{w}_k \propto N(0, Q_k)$ is a process noise with a covariance Q_k, H_k is an observation model, and $\underline{v}_k \propto N(0, R_k)$ is an observation noise with a covariance R_k.

The prediction step gives an estimate $\underline{\hat{a}}_{k|k-1}$ of the attributes and its covariance matrix $\hat{P}_{k|k-1}$ according to:

$$\underline{\hat{a}}_{k|k-1} = F_k\underline{\hat{a}}_{k-1|k-1} + B_k\underline{u}_{k-1}, \tag{9}$$

$$\hat{P}_{k|k-1} = F_k\hat{P}_{k-1|k-1}F_k^T + Q_{k-1}. \tag{10}$$

The update step corrects the prediction using the information observed at step k:

$$\underline{\hat{a}}_{k|k} = \underline{\hat{a}}_{k|k-1} + K_k\left(\underline{\hat{y}}_k - H_k\underline{\hat{a}}_{k|k-1}\right), \tag{11}$$

$$\hat{P}_{k|k} = (I - K_kH_k)\hat{P}_{k|k-1}, \tag{12}$$

where $K_k = \hat{P}_{k|k-1}H_k^T S_k^{-1}$ is the Kalman matrix with pre-calculated residual matrix $S_k = H_k\hat{P}_{k|k-1}H_k^T + R_k$ and I is the appropriate unit matrix.

In practical application, it is useful to simplify the attributes vector. For example, in entity tracking one can describe the movements with the constant velocity model (see Das (2008)), assuming accelerations as random evolution of the process.

In the case of multiple targets being tracked in clutter by multiple sensors, an extension to Kalman filter is required. For example, the Probabilistic Data Association (PDA) filter weights the association hypotheses according to their probabilities (Bar-Shalom & Tse (1975), Bar-Shalom & Fortmann (1988)). The further development, called Joint PDA (JPDA), analyzes joint events in order to track closely located entities (Fortmann et al. (1983)). The Integrated PDA (IPDA), proposed in Mušicki et al. (1994), introduces creation and deletion of tracked entities on a probabilistic basis. One of the most advanced methods, called Joint Integrated Probabilistic Data Association Filter (JIPDAF), combines all above mentioned ideas and considers all combinations of the observation to entity assignments and the creation and deletion mechanisms of modeling entities (Mušicki & Evans (2002)).

4. World modeling architecture

The world modeling represents the complete knowledge that is required for the operation of an autonomous system. The dynamic models contain up-to-date description of the world of interest, expressing each real-world entity with a virtual representative, and are similar to short-term memory. The prior knowledge contains pre-defined information about known particular entities or generalizations of entity classes, thus representing a long-term memory. The modeling system serves as a global information hub for all other mechatronic sub-systems. Hence, the structure of the world modeling is organized in a consistent, robust and efficient way, enabling fast search over representatives and attributes on a probabilistic basis. Moreover, there are mechanisms for consistency checks, correcting improbable states, e.g. flying cups or spatially overlapping representatives. Before introducing these advanced mechanisms, we demonstrate here the practical realization of attributes.

4.1 Modeling attributes

Since modeling information, in particular attributes, can be updated by different mechatronic sub-systems (e.g. perception and cognition modules) there has to be a global scheme or protocol for the information interchange. This global scheme defines attribute and relation concepts: which attributes are allowed, which possible values do they have, their semantic meaning, units, coordinate system, prior probability distribution (e.g. prior size distribution for kitchen tables), and so on. After defining attributes, the specific Degree-of-Belief description has to be introduced.

In practice, a continuous Degree-of-Belief distribution can be conveniently given via a Gaussian mixture model (GMM) approximation. The GMM represents a weighted sum of M component Gaussian densities as given in Eq. 13. The parameter M is empirically selected to find a balance between the computational complexity and the approximation precision. As new sensor information is fused into the dynamic model, the number of components in the mixture usually increases. In order to sustain the predefined number of M densities in the mixture, reduction techniques can be applied as discussed in Williams (2003).

$$p(\underline{x}|\omega, \underline{\mu}, \Sigma) = \sum_{i=1}^{M} \omega_i g\left(\underline{x}|\underline{\mu}_i, \Sigma_i\right), \tag{13}$$

where $\underline{x}$ denotes a D-dimensional continuous data vector, ω_i are the normalized mixture weights with $\sum_{i=1}^{M} \omega_i = 1$, and $g\left(\underline{x}|\underline{\mu}_i, \Sigma_i\right)$ are the component Gaussian densities. Each

component density is a D-variate Gaussian function as given in Eq. 14:

$$g\left(\underline{x}|\underline{\mu}_i, \Sigma_i\right) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}\left(\underline{x} - \underline{\mu}_i\right)^T \Sigma_i^{-1}\left(\underline{x} - \underline{\mu}_i\right)\right\}, \qquad (14)$$

where $\underline{\mu}_i$ denotes mean vectors and Σ_i covariance matrices.

4.2 Progressive mapping

After modeling attributes are clearly defined, it is important to introduce a common mechanism for attribute groups management. It is convenient to group attributes, which correspond to the same entity in the real world. It is natural then to describe representatives within the object-oriented (OO) methodology. It is important though to distinguish the OO approach in the most common sense and OO programming paradigm, which limits many modern world modeling systems to one application field. The OO languages have a clear inheritance structure of classes that is inapplicable for modern autonomous systems due to many arising problems. One of the problems is the arbitrariness of measured attribute sequences, e.g. color before size or vice versa, which prevents creation of one clear hierarchy of concepts. Another problem is the *open world* assumption, vital for intelligent autonomous systems. In the absence of clear inheritance hierarchy of classes, modeling entity is represented by a growing conglomerate of attributes and relations, created from incoming sensory information. Each observation is associated with the sets of attributes and relations and fused on a probabilistic basis. The update of such growing unclassified information containers is called *progressive mapping* (cf. Belkin (2009), Kühn et al. (2010)).

The progressive mapping can be visually described by the *abstraction level pyramids* (cf. Gheţa et al. (2008)). The dynamic model can be viewed as a hierarchy of abstraction levels, which contain *blank representatives* at the top and more specific representatives downside the hierarchy. The blank representative is an empty *information container* that marks the existence of "something", e.g. the existence of an entity that has not yet been measured. Each observation associated with the entity is iteratively fused into the information container, lowering the representative in the abstraction hierarchy. The idea of dynamic extension of information containers allows for an arbitrary information flow and an abstract information management. In order to understand these kind of attribute sets, they can be at any time matched to known concepts in the prior knowledge.

4.3 Prior knowledge

In order to classify each attribute container, e.g. assign a type, it can be at any time matched to the prior knowledge as shown in Fig 8. The prior knowledge contains general concepts as well as information about specific entities. This match, made on a probabilistic basis, provides the missing semantics and deduces the deficient attributes. The contents of the prior knowledge database can be predefined and then extended during the operation by concepts and entities via a learning process, as described in Section 5. Since prior knowledge contains a global information compendium about the environment, it is reasonable to store it outside the autonomous system (e.g. on a computer cluster) and provide a remote connection. In a case of remote connection failure, it is possible to operate the autonomous system without prior knowledge, though, without complex cognition abilities.
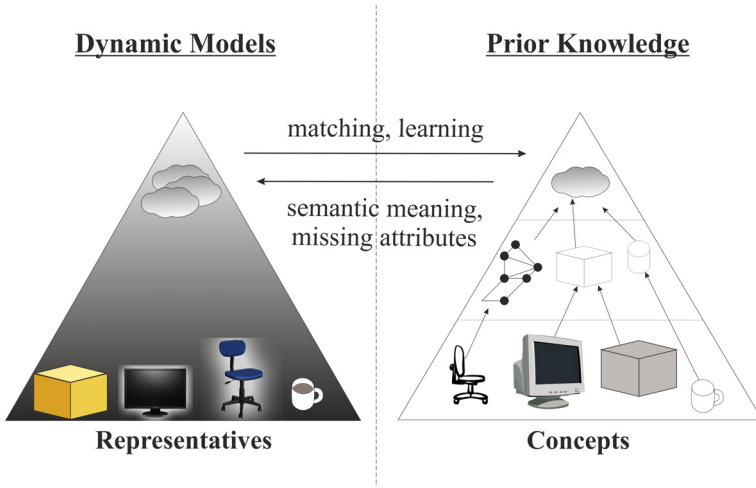
Fig. 8. World modeling architecture.

The concrete realization of matching between a dynamic representative and a class concept or specific entity is not trivial. It is possible to perform a deep search through one given tree-organized concepts hierarchy. Each node can be compared to the representative by two criteria: the structural similarity of attributes and the DoB distributions similarity. The structural similarity implies counting attributes that were observed and added to dynamic description to prior attributes described in ontology. Mathematically, such comparison can be performed by sets metrics, e.g. Tanimoto distance (Rogers & Tanimoto (1960)), Sørensen index (Sørensen (1957)) or other discussed in the overview article Jackson et al. (1989). The DoB distribution similarity between a representative attribute and the prior DoB can be calculated based on Gaussian mixture model metrics, such as Kullback-Leibler distance (Kullback & Leibler (1951), Cover & Thomas (1991)) or Wasserstein L^2 metric (Hazewinkel (2001), Jensen et al. (2007)).

4.4 Quantitative evaluation of the model

The information coming from sensors or stored in the world model can be evaluated numerically. For example, each observation or attribute vector can be characterized by some *information entropy*, that is a measure of uncertainty of a random variable (see Cover & Thomas (1991)). The difference between entropies before and after the observation and dynamic model fusion can give estimation on the information gain.

The numerical estimation of a given attribute can be quantified through the information entropy formulas:

$$H(A) = - \sum P(a) \log_2 P(a), \tag{15}$$

where A denotes some attribute and a is its discrete value and

$$h(A) = - \int p(a) \log_2 p(a) da, \tag{16}$$
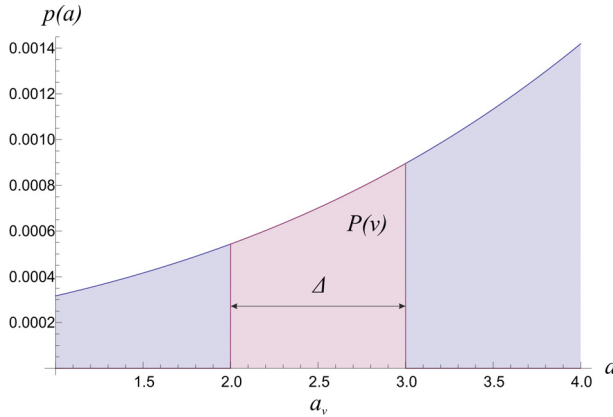
in case a is continuous.

Fig. 9. Discretization of a Degree-of-Belief distribution to Least Discernible Quantum.

In order to numerically estimate a representative as a whole, one has to analyze the entropies of all its attributes. Thus, one needs a unified framework for handling discrete and continuous distributions uniformly. However, by discretizing attribute A to A^* with smaller and smaller steps, so that $P(a^*) \rightarrow p(a)$ the corresponding entropy diverges from the continuous as $\lim_{a^* \rightarrow a} H(A^*) \neq h(A)$. One of the possible ways to overcome this problem is the introduction of the *least discernible quantum* (LDQ) (cf. Oswald (1991)), which specifies the utmost precision of any operation over the given attribute.

The least discernible quantum Δ defines the precision of each considered attribute. It is natural to set Δ equal to the accuracy of sensors. The discretization to LDQ can be performed by dissecting the distribution into sections as shown in Fig. 9 and integrating each section as in Eq. 17:

$$p(a) \rightarrow P(v) := \int_{a_v - \frac{\Delta}{2}}^{a_v + \frac{\Delta}{2}} p(a)da, \tag{17}$$

where $a \in [a_v - \frac{\Delta}{2}, a_v + \frac{\Delta}{2}]$ and $v \subseteq \mathbb{N}$.

The discretized entropy is defined then as

$$H_\Delta(A) := H(v) = -\sum_v P(v) \log_2 P(v). \tag{18}$$

5. Concept learning

For autonomous systems, concepts contained in its *knowledge* characterize the mapping from real world *entities* to *representatives* in the world model in a homomorphic way (Sydenham (1982)). The main purpose of the mapping is to share a meaningful set of symbols with its environment: on the one hand this mapping allows classification and even identification of the observed entities based on measurements of their attribute values. On the other hand, concepts allow to interoperably exchange meaningful information, e.g., when receiving requests from humans. Based on the mapping of representatives to concepts, human requests and sensor data can be interrelated and furthermore represented entities can be enhanced with rich semantics not derivable from pure observations (Heizmann et al. (2010)).

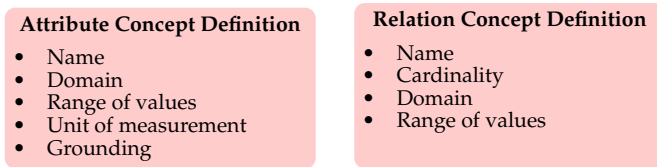| Attribute Concept Definition	Relation Concept Definition
• Name • Domain • Range of values • Unit of measurement • Grounding	• Name • Cardinality • Domain • Range of values

Fig. 10. Definition of attribute concepts and relation concepts.

Within a world model for autonomous systems, *semantics*, i.e. the meaning of symbols received from sensors or humans, have to be introduced in one or another way (Lim et al. (2011)). When preparing a system for operation, it is possible to equip the system with definitions of all the relevant concepts that are supposed to be encountered within its operational tasks. Since these concepts are defined prior to operation, they constitute *a priori* knowledge. By relying on a priori knowledge, an autonomous system is able to interrelate, classify and semantically handle only the entities occurring in a closed world. When operating in a real-world environment, however, modeling tasks often go beyond closed worlds towards an open-world modeling task: representing and handling a dynamic and changing environment which consists of entities of possibly unforeseen types (Kennedy (1998)). To cope with open-world modeling, an autonomous system must be able to extend its knowledge beyond a priori information by *acquiring new concept definitions*. Another purpose of knowledge is thus to store information *learned from experience* during operation.

The task of dynamically expanding the knowledge of an autonomous system constitutes an extension to the classical concept learning problem, i.e. learning Boolean valued functions (over attributes) from a given set of (training) data (Mitchell (1997)). In order to extend the knowledge of an autonomous system, new entity concepts must be learned. In concept learning, the process of learning can be regarded as a search within concept space (Mitchell (1997)), i.e. the space of all representable concepts. The structure of this space constitutes the bias to the concept learning problem and thus determines how concepts are learned and which concepts can be learned at all (Mitchell (1980)).

Within our presented approach to world modeling for autonomous systems, the structure of concept space is loosely defined by the object-oriented world model. In the following, this structure shall be further formalized to enable structured concept learning. Besides structuring concept learning, this formalization extends the general applicability of object-oriented world modeling by formally abstracting entity representation from the sensor data, thus making sensor components exchangeable and acquired knowledge interoperable.

Entity concepts we are interested in generally comprise a set of attributes. An attribute thereby is associated with an *attribute concept*. Attribute concepts describe the domain of an attribute (e.g., a height value, a position, a color, the number of wheels, the kind of fuel, etc.), its allowed range of values, its unit of measurement and the information on how instances of this attribute concept are grounded in the real-world (e.g., modeled a priori, measured by a sensor system, related to an effector interaction or to communication with humans). When instantiating attribute concepts within an entity concept, the expected attribute value(s) have to be specified in the form of a probability distribution interpreted in a DoB sense. Fig. 10 shows the definition of an attribute concept. Besides their attribute values, real-world entities can be characterized by their relations to other entities. Relations are associated with *relation concepts*. As shown in Fig. 10, a relation concept includes a name, the cardinality of the

Object Concept Definition
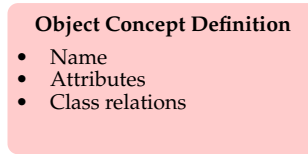- Name
- Attributes
- Class relations

Fig. 11. Object concept definition

relation, its domain and its range of values. Relation concepts define the relations occurring between entity concepts as well as relations existing between single concept instances, i.e. representatives, or even attributes. The domain in attribute concepts and domain and range of values in relation concepts are specified by appropriate entity concepts, e.g., a concept for color. The unit of measurement in attribute concepts can take values from a predefined hierarchy of unit concepts, e.g., specifying length units like meter or centimeter and their conversions. By abstracting the range of values and the unit of measurement from the domain to be represented within attribute concepts, sensors of different type can define different attributes (e.g., differing in unit and range of measured quantities) associated with the same attribute concept. Thus, these sensors are either interchangeable or can consistently interoperate.

The most basic entity concepts for world modeling are *object concepts*, which map physical real-world entities to representatives. Object concepts are defined based on attribute concepts and relation concepts. They consist of a name, a set of attributes and a set of class relations to other entity concepts (which apply to all representatives of a given concept). Fig. 11 shows the definition of object concepts. More abstract entities like roles (e.g., a customer), groups of objects, situations, etc. can be modeled by extending the given definition for object concepts. Following the definitions for entity concepts given above, *concept learning* within object-oriented world modeling allows for learning new object concepts as well as new attribute and relation concepts. By learning new attribute concepts, for example, the entity concept space is changed, thus extending beyond classical concept learning.

As with the syntactical aspects of concept representation, the *semantics of entities* and the semantics of their attributes also play an important role in concept learning. When an autonomous system for example faces the task of representing an encountered real-world entity, its learning component has to choose the relevant attributes from all available attributes (e.g., through sensor measurements or human input). To make an elaborate choice, the system has to base its decision on the semantics of the considered attributes and relate them to its general purpose. When defining a priori knowledge for an autonomous system, an ontology-like structure organizes conceptions as interrelated symbols. Within our object-oriented world model, attributes are used to further describe these conceptions, like a car having four wheels or a cup being cylindric and 9 cm tall. Nevertheless, for an autonomous system none of these conceptions has a meaning by itself. Furthermore, they do not reference any real-world entities. Thus, all concepts in the knowledge have to be grounded in either the real world or the human thought (for human-machine-interaction) (Harnad (1990)).

For the grounding of entity concepts, their attributes are considered. *Attribute concepts* in world modeling can be divided into three groups:
- perceptual,
- effector,

- and semantic attributes.

Perceptual attributes (like the height, color or speed of an entity) are grounded within the sensor components providing value information for these attributes. *Effector* attributes are grounded within the actuators used to manipulate real-world entities based on values represented by these attributes (Kennedy (1998)). An effector attribute for example can be the weight and the shape of a cup supposed to be grasped, lifted and moved. As can be seen from this example, there is no sharp boundary between perceptual and effector attributes (the weight could also be measured, e.g., by a scale). Which attribute is grounded by which sensomotoric component of an autonomous system thus in general depends upon the task currently at hand. Perceptual attributes are used for classifying and identifying of the observed real-world entities, while effector attributes assist in manipulating these entities.

Furthermore, there are attributes in the knowledge that can be directly associated with neither a sensor component nor an actuator component. Examples of such attributes may include information about if and how an entity can be moved (e.g., something is fixed to a wall and cannot be moved at all, or a door can only be rotated on its hinges), where an object is usually located (e.g., a cup in the cup section of a cupboard), etc. These attributes constitute *semantic* attributes as they are not directly stemming from either acquired knowledge or long-term experience (like long-term observation of a specific situation in conjunction with reasoning about it). Semantic attributes include the physical properties, behavior and capabilities of real-world entities. These attributes are useful in modeling the human grasp of the meaning of entities for autonomous systems, e.g., a cup as a drinking vessel, and thus enable human-machine understanding. Furthermore, they allow the semantical enrichment of the represented entities after mapping them to a known concept. By deriving information not obtainable by the pure use of sensomotoric components, they serve thus as a basis for higher-level information processing, like reasoning or planning.

After having defined a structure for entity concepts and the examined attribute semantics, we now move on to learning new entity concepts. *Concept learning* for autonomous world modeling can be structured as a task consisting of two interconnected subtasks:
- a deductive inference task on concepts in the knowledge, and
- an inductive learning task on sensor data.

The first subtask performs deductive inference on concepts already present in knowledge in order to either add new semantic properties or even learn whole new concepts (e.g., generalizations) in order to gain a better understanding of the world of interest (by a more appropriate or efficiently organized concept hierarchy). For example, a new super class drinking vessels could be created by deductive inference based on the common properties of the classes of cups and glasses. For extracting the semantic properties and hidden relations, the deductive subtask compares the concepts and their attributes for similarities, employing a variety of conceivable measures ranging from probability distribution metrics (like Kullback-Leibler divergence (Kullback & Leibler (1951))) to methods based on vector space models for entity concepts. After having found sufficiently similar concepts, one possible inference is to extract the matching parts and transfer them into a new concept, which then subsumes the deduction result within its set of inferred semantic properties, and the necessary conditions for these properties to hold in its set of attributes.

The second subtask in concept learning constitutes an inductive inference problem: learning new concepts from observations, which in case of the world modeling are given in the

form of a set of representatives. A world model enhances the observed real-world entities with semantics by mapping them to the known concepts. If a given representative cannot be properly assigned to the concepts stored in the knowledge, no useful semantics can be derived. Thus, the quality of the concept assignment is crucial for the overall quality of the world model. To ensure high-quality modeling, measures based on Shannon Entropy (Shannon (1948)) or Minimum Description Length (Rissanen (1978)) methods can be employed to evaluate the assignment of observed representatives to concepts. If a poor concept assignment is detected for some representatives, changes to the current set of concepts stored in the knowledge can be performed, including the creation of additional concepts and adaption or even deletion of the existing concepts. Created new concepts can subsequently be enriched with semantics by the deductive subtask.

6. Experimental set-up

The described theoretical analysis, as well as the complete system architecture for the world modeling has been tested in the three experimental set-ups:

- The German Science Foundation (DFG) project Collaborative Research Center (SFB) 588 "Humanoid Robots – Learning and Cooperating Multimodal Robots", see DFG SFB 588 (2001-2012);
- The Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) project "Network Enabled Surveillance and Tracking" (NEST), see Moßgraber et al. (2010);
- The Fraunhofer IOSB project "Wide Maritime Area Airborne Surveillance", see WiMAAS (2008-2011).

The main goal of the SFB 588 project is to create household assisting anthropomorphic robots (the three developed generations of such robots are presented in Fig. 12) with the following features:
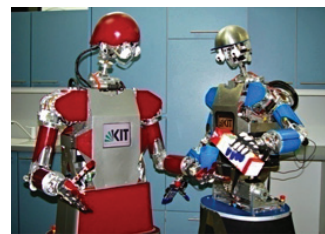
- Autonomous operation;
- Humanoid shape;
- Multimodality;
- Cooperation;
- Learning capabilities;

(a) ARMAR-I   (b) ARMAR-II   (c) ARMAR-III

Fig. 12. The three generations of the humanoid robot ARMAR (DFG SFB 588 (2001-2012)).

In fact, self-sufficient humanoid robots represent androids that are one of the ultimate goals of engineering and natural sciences. The multimodality implies that the communication

(a) Object detection in video streams

(b) World model of observed persons in indoor surveillance

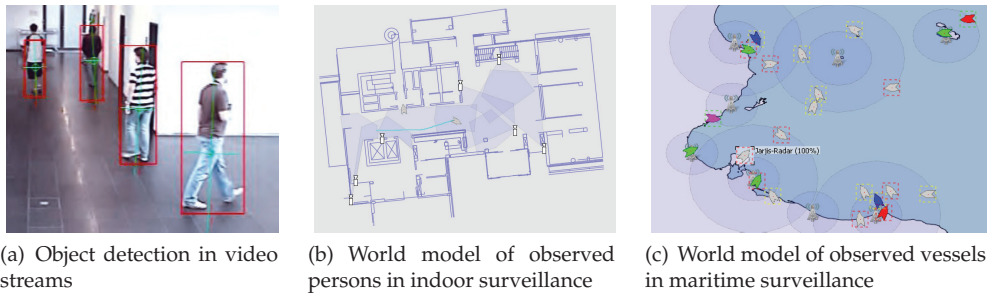(c) World model of observed vessels in maritime surveillance

Fig. 13. Applications of world modeling surveillance systems.

is intuitive for humans and machines, using speech, gestures, mimics, and haptics. The cooperation between humanoid robots and humans with support of intentions and context recognition techniques and safety regulations allows for collaborative activities. At the top of the robot's cognition abilities is the capability to learn. During the interactive explorations the new entities, concepts and even motions can be learned with an aid from a human. Due to complexity of the given tasks, such humanoid robots need a comprehensive state of knowledge on their environment. The required mechanisms and principles for the knowledge processing and management are implemented within the object-oriented world modeling system developed within the SFB 588 project (see Belkin (2009), Kühn et al. (2010), Belkin (2010), Baum et al. (2010), Gheţa et al. (2010)).

In the NEST research project, the proposed world modeling system has been employed as a virtual information hub between the evaluation and analysis services of an intelligent video-assisted security and monitoring system. In the context of indoor surveillance, detection of moving persons was performed in the video streams, see Fig. 13(a). Observations from several cameras are fused into a consistent representation and the resulting world model can be visualized, see Fig. 13(b). In the WiMAAS-project, the focus was on observing the maritime vessel behavior. Detections of vessels are coming from several different sources, e.g., navigational radar or signals from the automatic identification system (AIS). The resulting world model is visualized in Fig. 13(c). Based on the content of the world model, several methods for behavior analysis of the moving vessels can be applied.

7. Conclusion

This chapter outlines the state of the art of world modeling for autonomous systems and introduces several novel methods. First, we overview the world modeling problem, introduce the nomenclature and discuss the common concepts and methods used in the problem area. Then the global structure of the information workflow is specified along with the definition of the modeling domains. Next, each step of the information workflow is discussed in detail: the fusion of the incoming information from the sensors, the real-time autonomous modeling of the surrounding environment, the prior knowledge employment and the learning of new concepts. We also describe modern methods of the information fusion, including the information representation techniques, the information life cycle within the autonomous system, the Bayesian framework for information association and fusion and the Kalman filter. The dynamic modeling is discussed by presenting methods for modeling attributes

and entities, followed by the definition of the numerical estimate of the information content in the model. A separate section is dedicated to the learning of new concepts and the prior knowledge extension. Finally, the three experimental set-ups are described.

There are still many open questions in the area of world modeling. For example, there is no appropriate probabilistic framework for joint Degree-of-Belief distributions, many issues regarding the consistency checks or concepts learning are unsolved. In spite of many stated problems, first steps for creation of a universal world modeling for autonomous systems are carried out and the modeling system will evolve with the time, becoming more complex and intelligent.

8. References

Bar-Shalom, Y. & Fortmann, T. E. (1988). *Tracking and data association*, Academic Press.

Bar-Shalom, Y. & Tse, E. (1975). Tracking in a cluttered environment with probabilistic data association, *Automatica* 11.

Bauer, A., Emter, T., Vagts, H. & Beyerer, J. (2009). Object oriented world model for surveillance systems, *Future Security: 4th Security Research Conference*, Fraunhofer Press, pp. 339–345.

Baum, M., Gheţa, I., Belkin, A., Beyerer, J. & Hanebeck, U. D. (2010). Data association in a world model for autonomous systems, *Proceedings of IEEE 2010 International Conference on Multisensor Fusion and Integration for Intelligent Systems*.

Belkin, A. (2009). Object-oriented world modelling for autonomous systems, *Technical report*, Institute for Anthropomatics, Vision and Fusion Laboratory, Karlsruhe Institute for Technology.

Belkin, A. (2010). Information management in world modeling, *Technical report*, Vision and Fusion Laboratory, Institute for Anthropomatics, Karlsruhe Institute of Technology (KIT).

Cover, T. M. & Thomas, J. A. (1991). *Elements of Information Theory*, Wiley-Interscience.

Das, S. (2008). *High-Level Data Fusion*, Artech House Publishers.

DFG SFB 588 (2001-2012). Humanoid robots – learning and cooperating multimodal robots. URL: *http://www.sfb588.uni-karlsruhe.de*

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems, *Human Factors* 37(11): 32–64.

Fischer, Y. & Bauer, A. (2010). Object-oriented sensor data fusion for wide maritime surveillance, *2nd International Conference on Waterside Security*.

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

Fortmann, T. E., Bar-Shalom, Y. & Scheffe, M. (1983). Sonar tracking of multiple targets using joint probabilistic data association, *IEEE Journal of Oceanic Engineering* OE-8: 173–183.

Gheţa, I., Heizmann, M., Belkin, A. & Beyerer, J. (2010). World modeling for autonomous systems, *Proceedings of the KI 2010: Advances in Artificial Intelligence*, Springer Lecture Notes in Computer Science, pp. 176–183.

Gheţa, I., Heizmann, M. & Beyerer, J. (2008). Object oriented environment model for autonomous systems, *in* H. Boström, R. Johansson & J. van Laere (eds), *Proceedings of the second Skövde Workshop on Information Fusion Topics*, Skövde Studies in Informatics, pp. 9–12.

Glinton, R., Giampapa, J. & Sycara, K. (2006). A markov random field model of context for high-level information fusion, *Proceedings of the 9th International Conference on Information Fusion*.

Hall, D. L. & McMullen, S. A. H. (2004). *Mathematical Techniques in Multisensor Data Fusion*, Artech House, Inc.

Harnad, S. (1990). The symbol grounding problem, http://cogprints.org/3106/.
    URL: *http://cogprints.org/3106/*

Hazewinkel, M. (ed.) (2001). *Encyclopaedia of Mathematics*, Springer, chapter Wasserstein metric.

Heizmann, M., Gheta, I., Puente León, F. & Beyerer, J. (2010). Information fusion for environment exploration, *in* F. Puente León & K. Dostert (eds), *Reports on Industrial Information Technology*, Vol. 12, KIT Scientific Publishing, pp. 147–166.

IBM (2001). Autonomic computing: Ibm's perspective on the state of information technology, Manifesto.
    URL: *http://www.research.ibm.com/autonomic/manifesto/*

Jackson, D. A., Somers, K. M. & Harvey, H. H. (1989). Similarity coefficients: Measures of co-occurrence and assosication or simply measures of occurrence?, *The American Naturalist* 133: 436–453.

Jensen, J. H., Ellis, D. P., Christensen, M. G. & Jensen, S. H. (2007). Evaluation of distance measures between gaussian mixture models of mfccs, *Proceedings of the International Conference on Music Information Retrieval*, Vol. 2, Citeseer, pp. 107–108.

Kennedy, C. M. (1998). Anomaly driven concept acquisition.
    URL: *http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.7360*

Kephart, J. O. & Chess, D. M. (2003). The vision of autonomic computing, *Computer* 36(1): 41–50.

Kühn, B., Belkin, A., Swerdlow, A., Machmer, T., Beyerer, J. B. & Kroschel, K. (2010). Knowledge-driven opto-acoustic scene analysis based on an object-oriented world modelling approach for humanoid robots, *Proceedings of the 41st International Symposium on Robotics and the 6th German Conference on Robotics*.

Kullback, S. & Leibler, R. A. (1951). On information and sufficiency, *Annals of Mathematical Statistics* 22: 49–86.

Lim, G. H., Suh, I. H. & Suh, H. (2011). Ontology-Based unified robot knowledge for service robots in indoor environments, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 41(3): 492–509.

Meyer-Delius, D., Plagemann, C. & Burgard, W. (2009). Probabilistic situation recognition for vehicular traffic scenarios, *Proceedings of the 2009 IEEE international conference on Robotics and Automation*, pp. 4161–4166.

Mitchell, T. M. (1980). The need for biases in learning generalizations, *Technical report*.

Mitchell, T. M. (1997). *Machine Learning*, 1st edn, McGraw-Hill Education (ISE Editions).

Moßgraber, J., Reinert, F. & Vagts, H. (2010). An architecture for a task-oriented surveillance system: A service- and event-based approach, *Fifth International Conference on Systems (ICONS 2010)*.

Mušicki, D. & Evans, R. (2002). Joint integrated probabilistic data association - jipda, *Proceedings of the International Conference on Information Fusion*.

Mušicki, D., Evans, R. & Stankovic, S. (1994). Integrated probabilistic data association, *IEEE Transactions on Automatic Control* 39.

Oswald, J. R. (1991). *Diacritical Analysis of Systems: a Treatise on Information Theory*, Ellis Horwood Limited.

Rissanen, J. (1978). Modeling By Shortest Data Description, *Automatica* 14: 465–471.

Rogers, D. J. & Tanimoto, T. T. (1960). A computer program for classifying plants, *Science* 132: 1115–1118.

Shannon, C. E. (1948). A mathematical theory of communication, *Bell system technical journal* 27.

Sørensen, T. (1957). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons., *Biologiske Skrifter, Kongelige Danske Videnskabernes Selskab* 5: 1–34.

Steinberg, A. N., Bowman, C. L. & White, F. E. (1999). Revisions to the jdl data fusion model, *Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE Vol. 3719*.

Sydenham, P. H. (1982). *Handbook of Measurement Science: Theoretical fundamentals*, Vol. 1, Wiley.

Williams, J. L. (2003). *Gaussian mixture reduction for tracking multiple maneuvering targets in clutter*, Master's thesis, Graduate School of Engineering and Management, Air Force Institute of Technology, Air University, USA.

WiMAAS (2008-2011). Wide maritime area airborne surveillance.
URL: *http://www.wimaas.eu*