

Bayesian Fusion: Modeling and Application

Jennifer Sander and Jürgen Beyerer

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB)

Fraunhoferstraße 1, 76131 Karlsruhe

GERMANY

Email: [jennifer.sander|jürgen.beyerer]@iosb.fraunhofer.de

Abstract—Bayesian statistics leads to a powerful fusion methodology, especially for the fusion of heterogeneous information sources. If fusion problems are handled under consideration of the full expressiveness and the full range of methods provided by Bayesian statistics, the Bayesian fusion methodology possesses an impressive wide range of applications. We discuss this by having a closer look at selected aspects of Bayesian modeling. Thereby, also parallels to other methods used for information fusion will be drawn. With regard to the practical tractability of Bayesian fusion problems, selected approaches to deal with its potentially high complexity are discussed.

I. INTRODUCTION

By using the synergy between different information sources, information fusion can deliver information possessing a higher quality than the information that would be available if the sources were used without the synergy exploitation [1], [2]. Often, the fusion of heterogeneous information sources is particularly promising. Broadly speaking, by possessing different strengths and weaknesses, heterogeneous information sources can complement each other in an optimal way. E.g., by gathering different physical features, heterogeneous physical sensors can deliver complementary information and contribute significantly to the robustness of an overall system.

Having their beginnings in military applications, automatic methods for information fusion are used in multiple civil areas like robotics, visual inspection, and medicine, today. Information fusion is an interdisciplinary research field that uses concepts and methods from multiple areas like signal processing, applied statistics, artificial intelligence, and information theory. Due to its heterogeneity, not all concepts and approaches have been applied with optimal efficiency throughout the whole field of information fusion. E.g., the same concept has been studied repeatedly in different research works [3].

By adopting a Bayesian perspective, it is often possible achieve a unique view among different methods for information fusion [4]. Especially, by serving as an integration framework in this sense, the Bayesian methodology can counteract a fragmentation of the field of information fusion and relations among different approaches can be made transparent and used profitably. If fusion problems are handled under consideration of the full expressiveness and the full range of methods provided by Bayesian statistics, it is hard to find non-pathological real world examples in that the Bayesian methodology fails.

In this publication, we discuss this by having a closer look at selected aspects of Bayesian modeling. Thereby, also parallels to other methods used for information fusion will be drawn. With regard to the practical tractability of Bayesian fusion problems, selected already established and new approaches

to deal with its potentially high costs are discussed. Such high cost especially may arise if the space of the properties of interest is high dimensional and of heterogeneous nature. Before going in the details of the Bayesian approach, in the next section, some requirements that a mathematical fusion methodology should fulfill are summarized.

II. REQUIREMENTS ON A MATHEMATICAL FUSION METHODOLOGY

If heterogeneous information sources are to be fused, e.g., due to their different physical origin, their contributions are usually not compatible at least from a semantic point of view. In order to fuse such different sources, a mathematical fusion methodology should satisfy some essential requirements [5]:

- **Transformation:** transformation of the contributions of the sources into a common mathematical description.
- **Fusion:** combination of the transformed contributions.
- **Focusing:** concentration of the fusion result in order to derive specific conclusions for the problem at hand.

Information being useful for information fusion is completely specified if the facts, the corresponding uncertainties as well as the dependencies between the different information contributions are known and described. The ability to adequately handle uncertainties is an indispensable prerequisite for any meaningful fusion methodology. Also, the information fusion process that results from the three steps itemized above must respect dependencies between the information contributions.

While research on automated methods for information fusion has been long time focused on low level fusion, currently, there is a shift toward high level fusion [6]. High level fusion is still essentially based on human capabilities, today, and there is a strong need for automated assistance of human operators. With a view to optimize the needed man-machine interaction, automated fusion capabilities should be based on a fusion methodology being easy to understand and to communicate also to mathematical non-experts.

III. THE BAYESIAN POINT OF VIEW

A. Continuous concept to handling uncertainty

At least two approaches govern modern statistics: classical statistics and the (much older) Bayesian statistics that provides the basis for the Bayesian fusion methodology. Both approaches are grounded on the same definitions of probability, i.e., on the the definition of absolute probability as stated in the Kolmogorov axioms well as on the common definition of conditional probability [7]. However, they essentially differ with regard to the question what probability means, i.e., how

probabilistic statements must be interpreted.

In classical statistics, based on the empirical law on the stabilization of relative frequencies [8], probability is interpreted as “limit” of relative frequencies when an experiment is repeated infinitely often. In this sense, e.g., the probability therefore that a fair coin shows head is equal to 0.5 because the relative frequency of the event “head” usually¹ becomes stable around this value if the coin is thrown increasingly often. In Bayesian statistics, probability is interpreted in a wider sense as Degree of Belief, i.e., as quantification of the knowledge one has in the occurrence of the considered events.

As consequence, in the Bayesian approach, all quantities that are involved in a fusion task can be and are interpreted as being random – not only these whose value is uncertain in a sense that is consistent with the frequentist interpretation of probability used in classical statistics. By this, the Bayesian approach has the capability to adopt a symmetric treatment of all involved quantities and there is no need for indirect argumentation as in classical statistics. E.g., in classical statistics, an interval estimation regarding the value of an unknown quantity on the basis of a series of observations would be done by the calculation of a confidence interval. Such an interval has the property that, on average over repeated generations of the series of observations, it will contain the value of the unknown quantity at least in the percentage of cases being specified by the confidence level. The interpretation that the value of the unknown quantity is contained in the confidence interval with at least a certain probability is not applicable: the unknown quantity is not assumed to be random and, after the series of observations has been collected, its value is contained in the confidence interval or not. In contrast, in Bayesian statistics, a credible interval would be calculated to perform an interval estimation. Here, both, the unknown quantity as well as the observations are interpreted as being random in the sense of the Degree of Belief interpretation of probability and, given the concrete series of observations, the probability that the “true” value of the unknown quantity lies within the credible interval is at least the specified credibility level.

In high level fusion tasks, it is often not possible to interpret probabilities as “limit” of relative frequencies as done classical statistics [9]. Here, probabilistic models are often based on expert knowledge (see, e.g., [10]) and not on things like physical sensor properties which would rather admit a frequentist interpretation of probability. In contrast, the Bayesian interpretation of probability as measure of what is known is predestinated, here. In addition, prominent applications of the Bayesian methodology like target tracking methods [11] demonstrate also its success in low level fusion.

The Bayesian methodology provides a continuous concept to model uncertainty and to propagate it through the different JDL levels. Note that an objective² Bayesian analysis often has also a validation from the point of classical statistics [12]. E.g., if despite the information delivered by a series of observations no further information regarding the value of an unknown quantity of interest is available, the Bayesian maximum a posteriori estimate, i.e., the most probable value in the Bayesian sense,

¹The fact that the stabilization of relative frequencies occurs not always but only usually makes clear that the “limit” of relative frequencies is not a limit in the strict mathematical sense.

²Bayesian analyses are objective if they are based on objective distributions. A distribution is objective if two subjects equipped with the same knowledge both conclude the same distribution from that.

is identical to the frequentist Maximum-Likelihood estimate, i.e., the value that is in the sense of classical statistics mostly supported by the observations.

In contrast to other theories like the Dempster-Shafer theory, Bayesian theory needs only one single measure to describe uncertainty. This measure is conform with the founded rules of probability. Generally, the expressiveness of an analysis based on Dempster-Shafer theory and related theories can be also reached by appropriate Bayesian models [13]. We demonstrate this in Sec. IV-A by deriving a concrete Bayesian model being able to respect the unreliability of information sources.

B. Generic model and requirements satisfaction

Let the vector $z = (z_1, \dots, z_n) \in Z$, $n \in \mathbb{N}$, specify the quantities of interest, i.e., these aspects that have to be evaluated in a concrete fusion task. It is assumed that z adopts a certain “true” value that is not directly observable. Also, let d_s denote the known contribution of the information source number s , $s \in \{1, \dots, S\}$, $S \in \mathbb{N}$, such that the vector $d = (d_1, \dots, d_S)$ summarizes exactly all the information delivered by the information sources.

The Bayesian methodology is based on principles of inductive statistics. In this sense, in a given fusion task, it is assumed that the (known) value of d is a consequence of the fact that z adopts its (unknown) “true” value. Thereby, in the Bayesian methodology, uncertainties are modeled probabilistically in the sense of the Degree of Belief interpretation.

In inductive statistics, the (known) vector d is usually considered to be a realization of a random quantity. Thereby, it is assumed that it is only known that the distribution of this random quantity belongs to the distribution set $\{p(d|z)|z \in Z\}$. For d being fixed and z being variable, this assumption is communicated via the Likelihood function $l(d|z)$ that is defined to be proportional to $p(d|z)$. I.e., it holds $l(d|z) = c \cdot p(d|z)$. Thereby, the constant of proportionality c is arbitrarily but fixed for all $z \in Z$. If the components of d are conditionally independent given z , $l(d|z)$ can be substituted by the product of component specific Likelihood functions $l(d_s|z)$, $s \in \{1, \dots, S\}$. An essential advantage of the Bayesian methodology is its explicit handling of prior knowledge. Therefore, a Bayesian model complements the statistical model that has been introduced so far by the prior distribution $p(z)$.

In [5], [14], it has been shown by the authors that the Bayesian methodology fulfills the requirements on a fusion methodology that have been noted in Sec. II. Bayesian statistics provides different methods for transformation [15], [16]. E.g., transformation may be done by the application of the maximum entropy principle [17] that has a consolidated information theoretic background and is successfully applied in measuring technology [18] as well as in multiple other research disciplines [19]. The actual combination is done by the application of the well known Bayesian theorem [16] that is a direct consequence of the rules of probability. The result of this combination is the posterior distribution $p(z|d)$ which probabilistically embodies exactly all the information that is available about z . The Bayesian theorem delivers $p(z|d) \propto l(d|z)p(z)$. The posterior distribution is the basis for further calculations that may correspond to a focusing, e.g., marginalizations, estimations, or applications of decision theoretic concepts like the principle of expected utility [16].

Despite the fact that founded means are available to this

end, choosing the prior distribution in the best way may be a critical task [20]. Rather unproblematic are cases in that the information delivered by the information sources together clearly dominates the fusion task. In these cases, the posterior distribution is rather stable against different choices of the prior. If this is not the case or unclear, a promising strategy is to perform sensitivity analysis, i.e., to check the impact of different choices of the prior distribution.

Despite this difficulty, Bayesian statistics delivers a founded and practically well proven methodology that provides a wide range of methods for information fusion together with an intuitive interpretation. Most other methodologies for handling uncertainty do not possess such a lean and intuitively understandable concept. By its wide reaching interpretation of probability, in principle, the Bayesian methodology is able to adequately address any kind of uncertainty (compare also [21]). In addition, this is done in a direct manner and on the basis of a unique uncertainty measure.

IV. SELECTED EXAMPLES

A. The case of unreliable information sources

The Bayesian methodology is sometimes criticized for being not able to adequately respect the unreliability of information sources. To describe this criticism more precisely and to demonstrate that it is not necessarily warranted, we use an example that has been given in a similar form in [22].

Assume that, in a given fusion task, the type of an object is of interest. There are two possible object types, i.e., it holds $Z = \{A, B\}$. Assume further that two heterogeneous sensors being based on different physical principles both report the object type A . Sensor 1 has a reliability of $r_1 = 0.8$ and that sensor 2 has a reliability of $r_2 = 0.9$. Thereby, reliability means: in $r_s \cdot 100$ percent of the cases in that sensor number s reports a certain object type, this is true; in the remaining $(1 - r_s) \cdot 100$ percent of the cases, the sensor report is no help for differentiating the both possible object types, $s \in \{1, 2\}$.

As it is based on the principles of probability theory, the Bayesian methodology must assign the total belief (i.e., the total probability mass, here) explicitly among the atomic events of the underlying probability space. In contrast, Dempster-Shafer theory [23] allows assigning belief also explicitly to non-atomic subsets of the so called universe of discourse without the need to specify how this belief is concretely distributed among the atomic components of these subsets.

When applying Dempster-Shafer theory in the given example, one assigns a belief of r_s to A . The remaining belief of $(1 - r_s)$ is directly assigned to $A \cup B$. Then, Dempster's rule of combination delivers for A the support (i.e., the minimal total belief) 0.98 and the plausibility (i.e., the maximal total belief) 1. The length 0.02 of the uncertainty interval $[0.98, 1]$ is a quantification of the ignorance being present with regard to the total belief of A . For B , the uncertainty interval is given by $[0, 0.02]$ and $A \cup B$ has the uncertainty interval $[1, 1]$.

Critics of the Bayesian methodology argue that it is limited in cases like the described example because it is not able to assign the belief corresponding to $1 - r_s$ directly to the event $A \cup B$. It is argued that, in the Bayesian methodology, if the belief corresponding to the sensor reliability r_s shall be assigned to the event A , the remaining belief corresponding to $1 - r_s$ must be necessarily assigned to the event B (the complement of A) which is not consistent with the given

explanation of the meaning of the term reliability. As we will show, the statements pointed out in the criticism are not true if the information regarding the sensor reliabilities is explicitly incorporated in the Bayesian model via the introduction of a sufficient comprehensive probability space. To say it more concretely, if we model the information contributions of the sensors as pairs $d_s = (t_s, r_s)$ with t_s being the reported object type, the Bayesian methodology is able to adequately handle the described fusion problem.

According to the Bayesian methodology, one has to calculate the posterior distribution $p(z|d_1, d_2)$ using the Bayesian theorem. As, in the given example, there is no prior knowledge available, we obtain for $p(z)$ the uniform distribution on $Z = \{A, B\}$. As the two sensors work on the basis of different physical principles, it can be assumed that their information contributions are conditionally independent given z .

The definition of d_s and the rules of probability deliver

$$p(d_s|z) = p(t_s, r_s|z) = p(t_s|r_s, z)p(r_s|z). \quad (1)$$

The first term on the right side of (1) can be modeled as

$$p(t_s|r_s, z) = r_s \cdot \mathbf{1}_{\{z=t_s\}} + (1 - r_s) \cdot \frac{\mathbf{1}_{\{z=t_s\}} + \mathbf{1}_{\{z \neq t_s\}}}{2}. \quad (2)$$

Thereby, $\mathbf{1}_S$ is defined to be equal to 1 if statement S is true and 0 otherwise. Hence, for a completely reliable sensor with $r_s = 1$, we have

$$p(t_s|r_s, z) = \mathbf{1}_{\{z=t_s\}}. \quad (3)$$

On the other hand, for a completely unreliable sensor with $r_s = 0$, it holds that

$$p(t_s|r_s, z) = \frac{\mathbf{1}_{\{z=t_s\}} + \mathbf{1}_{\{z \neq t_s\}}}{2}. \quad (4)$$

I.e., in the latter case, there is complete ignorance if the reported object type is identical with the "true" type or not.

The sensor reliabilities (as they are defined) do not depend on the "true" object type. Hence, it can be concluded that the second term on the right side of (1) is equal to $p(r_s)$.

The derivations made so far deliver

$$p(z|d_1, d_2) \propto \prod_{s=1}^2 \left(\frac{1 + r_s}{2} \mathbf{1}_{\{z=t_s\}} + \frac{1 - r_s}{2} \mathbf{1}_{\{z \neq t_s\}} \right). \quad (5)$$

For the given example, one obtains $p(A|d_1, d_2) = 0.994$ and $p(B|d_1, d_2) = 0.006$.

This Bayesian fusion result is consistent with the result of Dempster-Shafer fusion at least in the sense that the Bayesian posterior beliefs and the Dempster-Shafer uncertainty intervals are consistent. At least for a not specially trained operator, the Bayesian result is much easier to comprehend than the result delivered by Dempster-Shafer theory is.

B. Relation to an optimization approach to information fusion

In this section, we sketch the common energy functional formalism that provides very flexible means to information fusion. We show how it is related to the Bayesian methodology. The energy functional formalism is essentially based on the formulation of the fusion problem via so called energy terms $E_t(z)$, $t \in \{1, \dots, T\}$, $T \in \mathbb{N}$. Thereby, each energy term represents a piece of the information that is given about z . Such a piece may be certain information delivered by

the information sources, prior knowledge or an additional constraint that shall be imposed on the fusion result. According to the formalism (especially, to the way in that the energy terms have to be formulated), the fusion result corresponds to the value of z that minimizes the global energy. The global energy is obtained by weighting and summing up the energy terms:

$$E(z) = \sum_{t=1}^T \omega_t E_t(z), \quad \omega_t > 0. \quad (6)$$

For a more detailed description of equation (6) and for concrete examples, we refer the reader to [5], [24], [25].

Introducing Gibbs' distribution using the global energy gives

$$\pi(z) \propto e^{-\beta E(z)} = \prod_{t=1}^T e^{-\beta \omega_t E_t(z)}, \quad \beta \rightarrow \infty. \quad (7)$$

By this transformation, the energy functional formalism is directly related to the Bayesian methodology. Especially, minimization of the global energy corresponds to Bayesian maximum a posteriori estimation. If all energy terms are quadratic, the corresponding Bayesian model is fully Gaussian.

A clear advantage of noting the relation between the energy functional formalism and the Bayesian methodology results from the range of methods and generic results of Bayesian statistics that become applicable by this. Especially, taking into account the fact that there exists no universal method for the minimization of the global energy, promising approaches for the complexity reduction of Bayesian fusion (compare Sec. V) may be profitably adopted for the energy functional formalism.

V. COMPLEXITY REDUCTION OF BAYESIAN FUSION

In Bayesian statistics, usually, the posterior distribution is not analytically calculable. As consequence, the practical tractability of Bayesian fusion may be critical because Bayesian fusion may cause unacceptable costs. If it holds $Z = Z_1 \times \dots \times Z_n$ and if each subspace Z_i possesses (possibly after discretization) the cardinality $|Z_i|$, the costs for obtaining the posterior distribution are $O(|Z|) = O(\prod_{i=1}^n |Z_i|)$.

In the next two sections, two promising state of the art approaches for dealing with the potentially enormous costs of Bayesian fusion are discussed. Subsequently, in Sec. V-C, the essentials of an alternative, namely local Bayesian fusion approaches that have been developed by the authors are introduced. These approaches circumvent high costs by avoiding the complete calculation of the posterior distribution for all possible values of Z . They can be of valuable benefit taken in isolation as well as in combination with other approaches.

A. Conjugate families

The traditional way to ensure the feasibility of the posterior distribution calculation is the use of conjugate families. A family \mathcal{P} of prior distributions is defined to be conjugate to $l(d|z)$ if it follows from $p(z) \in \mathcal{P}$ that also $p(z|d) \in \mathcal{P}$ holds. If the distribution family \mathcal{P} is parametrized by a moderate number of parameters, $p(z|d)$ will be analytically derivable in a computationally very efficient manner – simply by updating the corresponding parameters.

The most prominent example for a Bayesian model being based on the concept of conjugate families is the fully Gaussian

model. E.g., in the case $n = 1$, if the Likelihood function $l(d|z)$ corresponds to $N(z, \sigma^2)$, i.e., to a Gaussian with mean z and known variance σ^2 , the family of all (one-dimensional) Gaussian distributions constitutes a conjugate family: from

$$d|z \sim N(z, \sigma^2) \quad \text{and} \quad z \sim N(y, \sigma_0^2), \quad (8)$$

it follows that

$$z|d \sim N(\tau(\sigma^2 y + \sigma_0^2 d), \tau \sigma^2 \sigma_0^2) \quad (9)$$

with $\tau^{-1} = \sigma^2 + \sigma_0^2$. Hence, $p(z|d)$ can be obtained simply by the algebraic calculation of mean and variance in (9). Comparable results hold also for the multivariate case [15].

Even though the concept of conjugate families performs well in a lot of typical applications of information fusion, this concept is not free of problems. Firstly, one should not forget that, in general, its use is justified purely by technical arguments. Using the concept of conjugate families is beneficial with regard to the practical tractability of Bayesian fusion. However it comes along with a loss regarding the expressiveness of Bayesian statistics. Usually, the “true” prior distribution, which represents the available prior knowledge in an objective manner, is not contained in the distribution family \mathcal{P} such that the choice of an element of \mathcal{P} as prior distribution constitutes an approximation of the “true” prior distribution.

At choosing a suitable conjugate family, usually, one is restricted to conjugate families reported in literature. Even if Z possesses a homogenous structure and a moderate cardinality, the analytical derivation of a family being conjugate to a certain Likelihood function is not a trivial task³. Especially, if Z is complex, the identification of a suitable conjugate family may be extremely critical.

If no element of a conjugate family approximates the “true” prior distribution sufficiently well, mixtures of conjugate prior distributions may be used. E.g., it has been proven that mixtures of Gaussian distributions possess the capability to approximate every “true” prior distribution arbitrarily good [16], [26]. Also, it has been proven that the set of mixtures of $t \in \mathbb{N}$ elements of a conjugate family forms again a conjugate family [16]. However, as the computational costs caused by Bayesian fusion grow again with the number of mixture components, generally, there will be a trade-off between approximation accuracy and computational costs. Because of this, e.g., the dynamic reduction of mixture components is an actual research theme in computer science [26].

Another issue in the context of the concept of conjugate families is that it is known that the respective distributions may suffer from a lack of robustness [16], [27].

B. Markov Chain Monte Carlo methods

In the area of pure applied statistics, Bayesian methods have taken a major boost since modern stochastic simulation methods, especially Markov Chain Monte Carlo (MCMC) methods, have become available and applicable. In this area, MCMC methods have established themselves as almost universally applicable concept for making the computation of the posterior distribution and quantities that are further derived from it tractable.

³Usually, it is based on the identification of the Likelihood function as an element of the so called exponential family of distributions [16] (not to be confused with the exponential distribution).

The basis of MCMC methods is the generation of a Markov Chain (i.e., a stochastic process whose next state depends always only on the current state) possessing the property to converge to a stationary distribution that is identical to $p(z|d)$. If this Markov chain runs long enough, such that the convergence to the stationary distribution has been arisen, it delivers samples suitable for a posterior analysis. A sufficiently large set of drawn samples can be used to approximate theoretical quantities via its empirical counterparts with high quality. E.g., the posterior distribution itself may be approximated by a histogram calculated from the samples. To give another example, empirical moments can be used approximate the theoretical posterior moments.

To ensure that the drawn samples have the desired properties, an assessment of the speed with that the Markov chain converges to its stationary distribution is necessary. This is important because, indeed, the design principles underlying the used Markov chain ensure its convergence to $p(z|d)$. However, generally, it is not clear how fast (i.e., after what a number of steps) the stationary distribution is reached. Generally, the assessment of the convergence speed is done by applying and rating different methods for empirical convergence diagnostics [28]. Sometimes, it is also possible to ensure theoretically that the convergence has been arisen at least after a certain number of steps. However, being generally rather weak, such theoretical bounds are often of limited applicability.

In practice, the convergence speed of Markov chains varies considerably and, as consequence, the same holds also for the computational performance of MCMC methods. Especially, if Z is large and possesses a heterogeneous structure (as it is often the case in information fusion tasks), MCMC methods may prove themselves to be ineffective. This problem motivated, e.g., the research on Data Driven MCMC (DDMCMC) methods [29]. DDMCMC methods improve the effectiveness of MCMC methods in complex state spaces by the incorporation of specific knowledge regarding the underlying domain to design very specific Markov chains. Via data driven techniques, such a Markov chain is guided through the transition process. In principle, the concept of DDMCMC methods is a very promising one. However, due to their extreme adaption on the respective application and the large amount of expert knowledge necessary to realize this adaption, it is questionable if these methods have the potential to become a standard approach for Bayesian fusion.

C. Local Bayesian fusion approaches

Local Bayesian fusion approaches turn away from the principle that the probability space underlying a Bayesian model is fixed by default. They aim at performing the actual Bayesian fusion not globally with regard to the whole range Z of the quantities of interest but only locally, i.e., concentrated on these parts of Z where something being relevant for the actual fusion task takes place. Basis of the considerations that motivated the research on local Bayesian fusion approaches was an analogy to criminal investigations [30]. Here, prior to the actual detailed investigations, in a first step, clues are identified. Subsequently, clues are investigated in detail only with regard to a kind of local part of the overall context. Analogously, the primary idea of local Bayesian fusion approaches is to investigate in a first step conspicuous elements of Z . Then, on basis of the detected conspicuous elements, task

specific (local) Bayesian models are constructed and the actual Bayesian fusion is done with regard to these local Bayesian models. By this, the costs of the actual Bayesian fusion reduce themselves to $O(|U|)$ whereby $U \subset Z$ denotes the set of conspicuous elements.

To understand the principles of local Bayesian modeling, an even deeper insight into the theoretical foundations underlying the Bayesian methodology (than the one given so far) is needed. From a measure theoretic point of view, a Bayesian model corresponds to a product probability space of the form

$$(Z \times D, \sigma(\mathcal{Z}, \mathcal{D}), P_{z,d}) . \quad (10)$$

Thereby, Z and D denote the sets of possible values of the quantities of interest and of the contributions of the information sources, respectively. The sets \mathcal{Z} and \mathcal{D} are σ -fields over Z and D , respectively. They constitute these subsets of Z and D for that it is possible to formulate probability statements. I.e., they constitute the universe of knowledge states that are reachable with regard to z and d . Usually, in traditional (global) Bayesian models, the finest possible σ -fields over Z and D are chosen by default. The set $\sigma(\mathcal{Z}, \mathcal{D})$ is the smallest σ -field containing both, \mathcal{Z} and \mathcal{D} . It is the domain of the joint probability measure $P_{z,d}$. This measure is represented by the joint distribution $p(z, d)$ that factorizes into $p(z)p(d|z) = p(d)p(z|d)$. The components $p(z)$, $p(z|d)$, $p(d|z)$ of these products and their identity are the things on that the manipulations done at Bayesian fusion are essentially based.

Two possibilities to model locality lie on hand [31]. The first one is to construct a coarser Bayesian model. Here, the σ -field \mathcal{Z} of the global Bayesian model in (10) is replaced by $\sigma(\mathcal{Z} \cap U, \mathcal{Z} \setminus U)$. The resulting Bayesian model provides the possibility of having the same fine granular view on U as before. With regard to $\mathcal{Z} \setminus U$, no fine granular view is possible. The second possibility is to replace in (10) Z by U and \mathcal{Z} by $\mathcal{Z} \cap U$. As result of this, one obtains a local Bayesian model that completely “forgets” that z may adopt values in $\mathcal{Z} \setminus U$. Further research with regard to these two possibilities to model locality [31], [32] delivered that, except in special cases, the second local Bayesian model has a higher practicability than the first one. However, because of the way in that the probabilistic calculus shifts the global probability mass from $\mathcal{Z} \setminus U$ to U when “forgetting” $\mathcal{Z} \setminus U$, solely on the basis of this local Bayesian model, it is not possible to decide if U has been appropriately chosen. To solve this issue, different criteria for the appropriateness of local Bayesian models have been developed [33], [34], [35], [36].

In addition, a way to draw essentially from a local Bayesian model conclusions about the global probabilities has been worked out. More concretely, on the basis of the local posterior distribution and of a rating of the global prior probability of U , a meaningful probability interval scheme for the global posterior probabilities is constructable [37]. Of special practical interest is also the fact that the information that is incorporated into the probability interval scheme can be used as input for decision theoretic methods [33]. As consequence, it is possible to determine if decisions that are made on basis of local Bayesian fusion are surely also globally seen optimal ones. In cases in that their global optimality can not surely get validated, the expected maximum loss arising from making decisions on basis of a local Bayesian model can be determined. Hence, it is possible to ascertain if decisions on the basis of the current local Bayesian model are justifiable. If

this is not the case, a further analysis based on an enlargement of U may be done.

VI. CONCLUSION

The Bayesian methodology provides methods being mathematically fully substantiated and possesses the ability to provide a continuous concept for handling uncertainty. It distinguishes itself by a powerful expressiveness paired with an intuitively understandable interpretation of underlying models and processes. Different other methods that are applied in the field of information fusion can be seen from an integrative, uniform Bayesian view. Special examples for such methods are methods stemming from classical statistics and from optimization theory. By using the full expressive power of the Bayesian fusion methodology, it is also possible to handle fusions tasks for that it is often claimed that the probabilistic calculus is not adequate. This has been demonstrated by the derivation of a concrete Bayesian model being able to handle unreliable information sources. To support the reader at realizing complex Bayesian models in practice, selected state of the art approaches for handling the complexity of Bayesian fusion have been discussed. In this context, also a new concept, local Bayesian fusion approaches, has been introduced.

REFERENCES

- [1] B. V. Dasarathy, "Information Fusion – what, where, why, when, and how?" *Information Fusion*, vol. 2, no. 2, pp. 75–76, 2001.
- [2] L. Wald, "Some Terms of Reference in Data Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1190–1193, 1999.
- [3] B. Khaleghi, S. Razavi, A. Khamis, F. Karray, and M. Kamel, "Multi-sensor Data Fusion: Antecedents and Directions," in *3rd International Conference on Signals, Circuits and Systems (SCS)*, 2009, pp. 1–6.
- [4] H. B. Mitchell, *Multi-Sensor Data Fusion*. Springer, 2007.
- [5] J. Beyerer, M. Heizmann, J. Sander, and I. Gheta, "Bayesian methods for image fusion," in *Image Fusion*, T. Stathaki, Ed. Academic Press, 2008, pp. 157–192.
- [6] P. H. Foo and G. W. Ng, "High-level Information Fusion: An Overview," *Journal of Advances in Information Fusion (JAIF)*, vol. 8, no. 1, pp. 33–72, 2013.
- [7] A. Irle, *Wahrscheinlichkeitstheorie und Statistik*. Springer, 2005.
- [8] N. Henze, *Stochastik für Einsteiger*. Vieweg+Teubner, 2010.
- [9] G. Toth, M. M. Kokar, K. Wallenius, K. B. Laskey, M. Sudit, M. Hultner, and O. Kessler, "Higher-Level Information Fusion: Challenges to the Academic Community," in *Proceedings of the 11th International Conference on Information Fusion (FUSION)*, 2008.
- [10] Y. Fischer and J. Beyerer, "Ontologies for probabilistic situation assessment in the maritime domain," in *Proceedings of the IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2013, pp. 105–108.
- [11] W. Koch, "On Bayesian Tracking and Data Fusion: A Tutorial Introduction with Examples," *IEEE Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 29–52, 2010.
- [12] J. K. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis – Theory and Methods*. Springer, 2006.
- [13] S. Maskell, "A Bayesian approach to fusing uncertain, imprecise and conflicting information," *Information Fusion*, vol. 9, no. 2, pp. 259–277, 2008.
- [14] J. Beyerer, J. Sander, and S. Werling, "Bayes'sche Methodik zur lokalen Fusion heterogener Informationsquellen," *tm – Technisches Messen*, vol. 74, no. 3, pp. 103–111, 2007.
- [15] B. P. Carlin and T. A. Louis, *Bayesian Methods for Data Analysis*. Chapman & Hall, 2009.
- [16] C. Robert, *The Bayesian Choice*. Springer, 2007.
- [17] E. T. Jaynes, "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 3, pp. 227–241, 1968.
- [18] K.-D. Sommer, O. Kuehn, F. Puente Leon, and B. R. L. Siebert, "A Bayesian Approach to Information Fusion for Evaluating the Measurement Uncertainty," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006, pp. 507–511.
- [19] J. N. Kapur, *Maximum Entropy Models in Science and Engineering*. Wiley, 1989.
- [20] F. Papi, M. Podt, Y. Boers, G. Battistello, and M. Ulmke, "On constraints exploitation for particle filtering based target tracking," in *Proceedings of the 15th International Conference on Information Fusion (FUSION)*. IEEE, 2012, pp. 455–462.
- [21] D. V. Lindley, "The probability approach to the treatment of uncertainty in artificial intelligence and expert systems," *Statistical Science*, vol. 2, no. 1, pp. 17–24, 1987.
- [22] K. Dietmayer, "Evidenztheorie: Ein Vergleich zwischen Bayes- und Dempster-Shafer-Methoden," in *Informationsfusion in der Mess- und Sensortechnik*, J. Beyerer, F. Puente Leon, and K.-D. Sommer, Eds. Universitätsverlag Karlsruhe, 2006, pp. 39–49.
- [23] L. A. Klein, *Sensor and Data Fusion – A Tool for Information Assessment and Decision Making*. SPIE Press, 2004.
- [24] J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers, 1990.
- [25] C. Frese, *Planung kooperativer Fahrmanöver für kognitive Automobile*, ser. Karlsruher Schriften zur Anthropomatik, Bd. 10. KIT Scientific Publishing, 2012.
- [26] M. Huber, P. Krauthausen, and U. D. Hanebeck, "Superficial Gaussian Mixture Reduction," in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2011)*, 2011.
- [27] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.
- [28] I. Ntzoufras, *Bayesian modeling using WinBUGS*. Wiley, 2011.
- [29] S.-C. Zhu, Z. Tu, and R. Zhang, "Effective Bayesian inference by data-driven Markov chain Monte Carlo for object recognition and image segmentation," in *Automatic Target Recognition X, Proceedings of SPIE 4050*, F. A. Sadjadi, Ed., 2000, pp. 320–332.
- [30] J. Beyerer, M. Heizmann, and J. Sander, "Fuselets – an agent based architecture for fusion of heterogeneous information and data," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, Proceedings of SPIE 6242*, B. V. Dasarathy, Ed., 2006, pp. 235–243.
- [31] J. Sander and J. Beyerer, "Decreased Complexity and Increased Problem Specificity of Bayesian Fusion by Local Approaches," in *Proceedings of the 11th International Conference on Information Fusion (FUSION)*, 2008.
- [32] —, "A Local Approach for Bayesian Fusion: Mathematical Analysis and Agent Based Conception," *Robotics and Autonomous Systems*, vol. 57, no. 3, pp. 259–267, 2009.
- [33] —, "Decision Theoretic Approaches for Focussed Bayesian Fusion," in *Proceedings of the IEEE ISIF Workshop on Sensor Data Fusion: Trends, Solutions, Applications (SDF 2011)*, 2011.
- [34] J. Sander, M. Heizmann, I. Goussev, and J. Beyerer, "A local approach for focussed Bayesian fusion," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, Proceedings of SPIE Vol. 7345*, B. Dasarathy, Ed., 2009.
- [35] J. Sander, J. Krieger, and J. Beyerer, "The Importance of Statistical Evidence for Focussed Bayesian Fusion," in *KI 2010: Advances in Artificial Intelligence*, ser. Lecture Notes in Artificial Intelligence, R. Dillmann, J. Beyerer, U. D. Hanebeck, and T. Schultz, Eds., vol. 6359. Karlsruhe: Springer, Sep. 2010, pp. 299–308.
- [36] J. Sander, "Use of a Probability Interval Scheme for Focussed Bayesian Fusion," in *Proceedings of the 2010 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*, ser. Karlsruher Schriften zur Anthropomatik, Bd. 7, J. Beyerer and M. Huber, Eds. KIT Scientific Publishing, 2010, pp. 31–46.
- [37] J. Sander, M. Heizmann, I. Goussev, and J. Beyerer, "Global evaluation of focussed Bayesian fusion," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications, Proceedings of SPIE Vol. 7710*, J. J. Braun, Ed., 2010.