Global evaluation of focussed Bayesian fusion

Jennifer Sander\textsuperscript{a}, Michael Heizmann\textsuperscript{b}, Igor Goussev\textsuperscript{a}, and Jürgen Beyerer\textsuperscript{a,b}

\textsuperscript{a}Karlsruhe Institute of Technology (KIT), Institute for Anthropomatics (IFA), Vision and Fusion Laboratory, Adenauerring 4, 76131 Karlsruhe, Germany; \textsuperscript{b}Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany

ABSTRACT

Information fusion is essential for the retrieval of desired information in a sufficiently precise, complete, and robust manner. The Bayesian approach provides a powerful and mathematically funded framework for information fusion. By local Bayesian fusion approaches, the computational complexity of Bayesian fusion gets drastically reduced. This is done by a concentration of the actual fusion task on its probably most task relevant aspects. In this contribution, further research results on a special local Bayesian fusion technique called focussed Bayesian fusion are reported. At focussed Bayesian fusion, the actual Bayesian fusion task gets completely restricted to the probably most relevant parts of the range of values of the Properties of Interest. The practical usefulness of focussed Bayesian fusion is shown by the use of an example from the field of reconnaissance. Within this example, final decisions are based on local significance considerations and consistency arguments. As shown in previous publications, the absolute values of focussed probability statements represent upper bounds for their global values. Now, lower bounds which are obtained from the knowledge about the construction of the focussed Bayesian model are proven additionally. The usefulness of the resulting probability interval scheme is discussed.

Keywords: information fusion, Bayesian fusion, local Bayesian fusion, focussed Bayesian fusion, Maximum Entropy principle, probability interval scheme, probability bounds, restricted model, qualitative analysis, quantitative analysis

1. INTRODUCTION

In many real world tasks, desired information on a scene, a process, or a situation is to be obtained by the fusion of different information sources. A short overview on the manifold benefits of the application of information fusion is given in Sec. 2.1. The Bayesian fusion methodology is mathematically funded and intuitively understandable. It has as a wide range of application and is in principle usable for the straightforward solution of most information fusion tasks. If information from different sources is fused by the use of the Bayesian methodology, all available information is modelled probabilistically and the resulting probability distributions are combined via the application of the Bayesian theorem. By this, facts and corresponding uncertainties are processed in an adequate manner. A short introduction into the Bayesian fusion methodology is given in Sec. 2.2.

However, the main disadvantage of the Bayesian fusion methodology is its global perspective on the Properties of Interest (Pols) about which information shall be obtained. By this, the Bayesian fusion methodology causes high computational and storage costs in most real world applications. None of the common methods for complexity reduction is generally appropriate for the solution of real world fusion tasks without reducing thereby the theoretical power of the Bayesian methodology.\textsuperscript{1} As alternative solution, we investigate local Bayesian fusion approaches by which high costs caused by Bayesian fusion are circumvented, see Sec. 2.3.

The essential idea underlying local Bayesian fusion is to perform Bayesian fusion at least not in detail for the whole space which is spanned by the range of the Pols but only locally\textsuperscript{*}--at least in these regions which are

\textsuperscript{*}Within this context, the term local does not necessarily imply spatial closeness. Also, it is not used identically as in the context of Bayesian networks.
essential for the solution of the actual fusion task. Especially, the restriction of the range of possible values of
the PoIs to the corresponding local context delivers a straightforward fusion scheme. Because of the theoretical
similarity between the corresponding proceeding and the focussing mechanism of Bayesian fusion,\textsuperscript{2} we termed
local Bayesian fusion based on a restriction of the range of the PoIs focussed Bayesian fusion.\textsuperscript{1} As demonstrated
in previous publications,\textsuperscript{1,3} the validity of a focussed Bayesian analysis is verifiable for example by a statistical
error bound or by quality indicators based on information theory.

In Sec. 3.2, the validity and the usefulness of focussed Bayesian fusion are demonstrated using a practical
example. The task is the detection and classification of cars within a scene (landscape model) by the fusion
of prior knowledge from a street map, HUMINT information and various kinds of IMINT information. As
demonstrated, this task can be solved by a qualitative Bayesian analysis via significance considerations. After
starting with a general description of the corresponding focussed Bayesian fusion scheme in Sec. 3.1, we firstly
describe in Sec. 3.2.1 the necessary probabilistic modelling for the corresponding global Bayesian setup. Here,
it is also shown how the given task is solvable theoretically by global Bayesian fusion. However, to avoid high
computational and storage costs, the task is solved by focussed Bayesian fusion as described in Sec. 3.2.2. Section
3.2.3 contains an evaluation and a discussion of the high potential of focussed Bayesian fusion in particular with
respect to the given example.

If quantitative probabilistic statements are obtained via focussed Bayesian fusion, it is important to regard
the distortion of absolute values which results from the focussing: at focussed Bayesian fusion, probabilistic
statements which are based on normalized probability distributions represent upper bounds for their global
equivalents.\textsuperscript{1,3} see Sec. 4.1. In Sec. 4.2, it is shown that facts concerning the construction of focussed Bayesian
models can be used for the deviation of lower bounds for global posterior probabilities. These lower bounds
can be calculated within a focussed Bayesian model if the prior relevance of the local context is ratable. In
Sec. 4.3, the resulting probability interval scheme will be discussed. It becomes clear that this scheme also makes
conclusions concerning the validity of a focussed Bayesian model possible.

2. THE BAYESIAN METHODOLOGY FOR INFORMATION FUSION

2.1 Why Information Fusion?

The observation of a scene, a process, or a situation by an information\textsuperscript{†} source generally comes along with
information reduction. In a sensor system, information reduction can result from

- windowing: observation of a temporally and/or spatially bounded extract,
- projection: temporal, spatial and/or spectral reduction of the dimensions,
- sampling: time discretization, space discretization and/or quantization,
- addition of disturbances like thermal sensor noise.

If it is not possible to retrieve the desired information completely and robustly by the use of only one information
source, the fusion of several sources is promising.\textsuperscript{4–7}

The application of methods from information fusion may be aimed at different objectives: fusing different
information sources can deliver a result with a higher information content. Information fusion can extend sensor
covering in temporal, spatial and/or spectral dimension by the combination of appropriately overlapping observ-
ations. The resolution can get increased spatially and/or temporally by information fusion, e.g., if repeated
measurements are combined. Especially if the useful information is redundantly present within the information
sources, information fusion can increase system robustness, system reliability and the confidence in the result.
In many real world tasks, it is possible to infer the values of not directly observable PoIs from several indirect
measurements. In this case, information about the PoIs can be gained from the fusion of several sensors which
deliver respective observation quantities. E.g., by the fusion of intensity images of a surface under different illu-
mination conditions, the surface normal is determinable. This proceeding may be less expensive and faster than

\textsuperscript{†}Information denotes everything which has the potential to reduce existing uncertainty.
the direct determination of the surface normal by a coordinate measuring machine, by white light interferometry, or by a structured light 3D scanner. By means of redundant information acquisition, information fusion can reduce costs by reducing the number of system failures. Costs can also get reduced by information fusion if one expensive high-quality sensor is replaced by several considerably cheaper sensors the measurements of which are fused. E.g., the construction of a depth map by the fusion of several images from a stereo camera or a camera array may be cheaper than the direct measurement of depth values by techniques like time of flight, white light interferometry or laser triangulation.\[^8\] Heterogenous information sources often posses specific weaknesses and strengths and may complement each other in an optimal way. E.g., the fusion of sensors which are different with respect to their physical principles is promising because different physical properties of a scene are accessible.\[^9\]

In real world fusion tasks, sensory information is often complemented by non-sensory one like symbolic information, physical laws, or mathematical constraints. E.g., if the aim of the fusion task is to detect and classify objects in a scene, the fusion of (sensory) IMINT information and (symbolic) HUMINT information may be promising. As an example, on the one hand, it may be easy for a human observer to differ between similar object classes on the basis of specific object details while an automatic analysis of the IMINT information may not be able to perform this task sufficiently well. On the other hand, an automatic analysis of IMINT information will usually outperform the report of a human observer with respect to the exact determination of the object positions and other quantitative properties. Considering prior knowledge and constraints on the fusion result usually further improves the quality of the final result.

### 2.2 Why Bayesian Fusion?

A fusion methodology which is theoreticly applicable to all kinds of information fusion problems has to cope with differently scaled quantities. E.g., if objects within a scene have to be detected and classified, the PoIs contain both, ratio scaled and nominally scaled components. It also has to cope with information contributions on different abstraction levels, e.g., if IMINT and HUMINT information have to be fused.\[^10\]

Generally, to make the actual combination of information contributions stemming from different sources possible, they have to be homogenized by a transformation into a common mathematical description. A meaningful description has to represent not only the facts but also the corresponding uncertainties. Within the Bayesian methodology, all information is represented by probability distributions in the sense of the Degree-of-Belief (DoB) interpretation.\[^2\] For the mathematical description, we use the following notations: \( z := (z_1, ..., z_N) \in Z \) with \( Z := Z_1 \times ... \times Z_N, \ N \in \mathbb{N} \), denote the PoIs and \( d := (d_1, ..., d_S) \in D \) with \( D := D_1 \times ... \times D_S, \ S \in \mathbb{N} \), denote the information from the information sources. Thereby, \( d_s, s \in \{1, ..., S\} \), stands for the contribution of the
information source $s$. As usual, we assume that the PoIs adopt a certain “true” value about which it is inferred from $d$.

Prior knowledge gets incorporated explicitly and adequately by its transformation into the prior distribution $p(z)$. The information stemming from the information sources is transformed into the Likelihood function $l(d|z)$. If the contributions of the information sources are conditionally independent given the “underlying truth” $z$, which may be especially the case if the information sources are heterogenous, it is possible to transform each information contribution separately into a source specific Likelihood function $l(d_s|z)$, $s \in \{1,\ldots,S\}$.

After the transformation, the application of the Bayesian theorem delivers

$$p(z|d) \propto l(d|z)p(z).$$

(1)

If the information contributions are conditionally independent given $z$, $l(d|z)$ can be substituted by $\prod_{s=1}^{S} l(d_s|z)$, i.e., by the product of the source specific Likelihood functions. The result of a Bayesian fusion task is strictly speaking the posterior distribution $p(z|d)$ which comprehensively represents the information about the PoIs which is delivered by prior knowledge and the contributions of the information sources—provided that the transformation of information into a probabilistic representation is artefact-free, lossless and not distorting. A theoretically funded method for such a transformation of information into an objective probabilistic representation is given by the Maximum Entropy (ME) principle.\textsuperscript{12,13}

The Bayesian methodology constitutes a powerful fusion approach which distinguishes itself by its funded mathematical background, its slim calculus and its clearness. Especially, it meets the essential requirements for a reasonable fusion methodology\textsuperscript{2} and it is information theoretically lossless. By the transformation of information into a probabilistic representation in the sense of the DoB interpretation, every kind of uncertainty is represented in an adequate manner.\textsuperscript{14–16} differently scaled quantities are manageable simultaneously and the consideration on different abstraction levels is possible without difficulties. Bayesian fusion can be applied in a centralized, distributed, or hybrid manner. Theoretically, the Bayesian approach has therefore a wide range of application and it is theoretically usable for the straightforward solution of most information fusion tasks.

2.3 Why local Bayesian fusion approaches?

Bayesian methods often necessitate immense storage space and cause unacceptable computing costs: the complexity for the necessary operations to obtain the posterior distribution is $O(\prod_{i=1}^{N} |Z_i|) = O(\zeta^N)$, $\zeta = N \sqrt{\prod_{i=1}^{N} |Z_i|}$.

Local Bayesian fusion approaches have the objective to circumvent such high costs by avoiding the complete calculation of the posterior distribution for all possible values of the PoIs. By a pre-evaluation, values of the PoIs that have a higher potential to be the “true” value than others are detected. Probabilistic statements are then made with respect to modifications of the usual Bayesian model. At focussed Bayesian fusion, a restriction of $Z$ on a local context $U \subset Z$ is done. I.e., values of the PoIs are ignored completely. The set $U$ contains the part of $Z$ which is according to the pre-evaluation the most task relevant.

3. FOCUSED BAYESIAN FUSION FOR A QUALITATIVE ANALYSIS

3.1 Fusion Scheme

Focussing on the local context $U$ mathematically corresponds to a conditioning on $U$.\textsuperscript{1} The focussed posterior distribution $p_U(z|d)$ can be obtained via the application of the Bayesian theorem:

$$p_U(z|d) \propto p(d|z)p(z)I_U(z).$$

(2)

\textsuperscript{1}The term objective is used differently in the context of the Bayesian theory. We call a transformation of information into a probabilistic representation objective if different subjects derive the same probability distribution given the same information.

\textsuperscript{3}The notation $|Z_i|$ stands for the number of elements of a discrete subspace $Z_i \subseteq Z$ and for the Lebesgue measure (length, area, or volume), respectively, if the subspace $Z_i$ is continuous.
Figure 2. IMINT information used for the fusion example.

Thereby, \( 1_U(z) := 1 \) if \( z \in U \), \( 1_U(z) := 0 \) if \( z \in Z \setminus U \). If the global prior distribution \( p(z) \) is available for all \( z \in U \), an unnormalized version of the focussed posterior distribution is calculable simply by a multiplication of \( p(z) \) and the Likelihood function \( p(d|z) \) for all \( z \in U \). Obviously, also the knowledge of an unnormalized version of \( p(z) \) for all \( z \in U \) is sufficient for this. The presented focussed Bayesian fusion scheme exactly corresponds to the respective global one which is given by (1).

Generally at Bayesian fusion, a qualitative analysis based on significance considerations is possible by an examination of posterior probability ratios. Obviously, even the knowledge of an unnormalized version of \( p(z|d) \) is sufficient to realize such a comparison of posterior probability values. From (1) and (2), it becomes obvious that such kinds of conclusions for events \( E \subseteq U \) are globally completely meaningful even if they have been obtained by focussed Bayesian fusion.

### 3.2 Practical Example

#### 3.2.1 Bayesian Modelling

We continue an example which has been given in [1]. There, focussed Bayesian fusion has been used for the detection and classification of cars within a scene (landscape model, scale 1:160) on the basis of prior knowledge from a street map and IMINT information which is represented by a single gray-value image of the scene. Here, HUMINT information \( d_1 \) is used additionally, and the IMINT information corresponds to three gray-value images \( d_2, d_3, d_4 \), which have been acquired from a fixed camera position with three different interference filters (400 nm, 550 nm, 650 nm), see Fig. 2. The spectral filters have a bandwidth of 50 nm.

According to the given task, \( Z = Z_1 \times Z_2 \times Z_3 \) holds where \( Z_1 \) and \( Z_2 \) denote the horizontal and vertical position of a car in the scene, respectively, and \( Z_3 \) denotes the set of possible car types. Without loss of generality, it is assumed that \( Z_3 = \{ "Audi", "DKW", "Ford", "Mercedes", "Opel" \} \). As described below, the probabilistic representation of the IMINT information is obtained via matched filtering. Because this technique is not implicitly rotation invariant, the probabilistic modelling will be done with respect to an enlargement \( Z \times R \) of \( Z \). Thereby, \( R \) denotes the set of possible driving directions. Without loss of generality, it is assumed that \( R = \{ "south", "north", "west", "east" \} \). Mathematically, the direction \( r \) corresponds to a nuisance parameter: since \( r \) is not task relevant, this quantity has to be eliminated from the final result.\(^2\)

The given task can be performed by a qualitative analysis on the basis of Bayesian fusion. Thereby, an unnormalized version of the posterior distribution is calculated. A subsequent estimation of the posterior most significant values delivers positions and types of the cars within the scene. Significance considerations are based on posterior probability ratios with respect to neighborhoods and with respect to the highest possible value of the unnormalized posterior distribution. Thereby, the constraint that two cars in the scene do not overlap is considered.

Applying the ME principle, the prior distribution \( p(z_1, z_2, z_3, r) \) over \( Z \times R \) which represents exactly the knowledge from the street map and a measurement uncertainty \( \sigma \) is approximately a quantity which is proportional to the convolution of the uniform distribution \( U(Y) \) over the set

\[
Y := \{(z_1, z_2) | (z_1, z_2) \in Z_1 \times Z_2 \text{ and } (z_1, z_2) \text{ is a location on a street}\}
\]

with the two-dimensional Gaussian distribution \( \mathcal{N}((0, 0), \sigma^2 \cdot I_2) \), where \( I_2 = \text{diag}(1, 1) \): for the car position \( (z_1, z_2) \), the prior is given by\(^{10}\)

\[
p(z_1, z_2) \propto q_0(z_1, z_2) * u_Y(z_1, z_2), \quad q_0(z_1, z_2) \sim \mathcal{N}((0, 0), \sigma^2 \cdot I_2), \quad u_Y(z_1, z_2) \sim U(Y). \tag{3}
\]
The prior knowledge is modelled to be not informative with respect to car type and driving direction: here, 
\( p(z_1, z_2, z_3, r) \propto p(z_1, z_2) \). Fig. 3(a) illustrates the used unnormalized prior distribution for any car type \( z_3 \) and 

driving direction \( r \). We chose \( a = 10 \) pixels (px), here.

The HUMINT information \( d_1 \) is given by facts and uncertainties:

\[
d_1 = \left\{ \{ (\mu_1(k), \mu_2(k)), \sigma_\mu(k), r_h(k), q(k) \} \mid q(k) := \{ b(z_3, k) \mid z_3 \in Z_3 \} \right\},
\]

\( K \in \mathbb{N} \) denotes the number of cars which have been reported by the human observer. \((\mu_1(k), \mu_2(k))\) denotes the position of the detection \( k \in K \) and \( \sigma_\mu(k) \) specifies the corresponding spatial uncertainty. The value \( b(z_3, k) \) probabilistically specifies the knowledge of the human observer concerning the car type \( z_3 \) of detection \( k, z_3 \in Z_3 \). Additionally, the human observer delivers a statement \( r_H(k) \) with respect to the driving direction of the respective car. The statement is simplified to \( r_H(k) \in R_C := \{ \text{“west or east”}, \text{“north or south”} \} \).

\( d_1 \) is transformed into a probabilistic representation using Gaussian mixtures:

\[
l(d_1 | z_1, z_2, z_3, r) = \sum_{k=1}^{K} b(z_3, k) q_k(z_1, z_2), \quad q_k(z_1, z_2) \sim \mathcal{N}(\{ (\mu_1(k), \mu_2(k)), \sigma_\mu(k)^2 \cdot I_2 \}).
\]

Since there is no explicit information from the human observer concerning the uncertainty about the 

statement \( r_H(k) \), \( l(d_1 | z_1, z_2, z_3, r) \) is modelled to be not informative with respect to the driving direction \( r \). For \( r \) fixed, the 

used Likelihood function is illustrated in Fig. 3(b). Here, \( \sigma_\mu(k) = 0.1 \cdot |Z_1| \). Note that the position \((z_1, z_2)\) is a 

continuous quantity while the car type \( z_3 \) is a discrete one.

The probabilistic representation \( l(d_s | z_1, z_2, z_3, r) \) of the IMINT information \( d_s, s \in \{2, 3, 4\} \), is obtained by 

the application of a matched filter bank with filters designed according to \( Z_3 \). The used matching method is 

normalized correlation coefficient matching.\(^{17}\) To hold the size of the matched filter bank moderate, templates 

(corresponding to a gray-value signature which has been obtained without the use of inference filters are used for 

the evaluation of the images \( d_2, d_3, d_4 \). Thereby, the matched filtering is performed with respect to gradient 

images \( G_2, G_3, G_4 \) of \( d_2, d_3, d_4 \) using respective gradient images of the templates. Here, all gradient images have 

been obtained by the application of Sobel filtering. Although these templates are not reproducible from each of 

the gradient images \( G_2, G_3, G_4 \) solely by additive and multiplicative variations, e.g., illumination changes, 

the performance of the resulting matched filters is sufficient with respect to the given task. The information 

contributions \( d_2, d_3, d_4 \) are modelled probabilistically as being conditionally independent given \((z_1, z_2, z_3, r)\) 

because of the use of spectral filters with disjoint transmission ranges for the image acquisition.\(^{4}\) Because of the 

heterogeneity of the HUMINT and the IMINT information, conditional independence also holds between \( d_1 \) and 

\( d_s, s \in \{2, 3, 4\} \).

\(^{4}\)The conditional independence may hold only approximately: the same camera has been used for the acquisition of 

the three images and, in consequence, there may be systematic influences, and the sensor noise may not be independent.\(^{8}\)
For global Bayesian fusion, formulas (3) and (5) have to be evaluated for all \((z_1, z_2) \in Z_1 \times Z_2\) and for all \((z_1, z_2, z_3) \in Z_1 \times Z_2 \times Z_3\), respectively. The matched filtering results \(l(d_s|z_1, z_2, z_3, r)\) have to be computed for all \(s \in \{2, 3, 4\}, (z_1, z_2, z_3, r) \in Z \times R\). After that, the unnormalized posterior distribution has to be calculated and analyzed for all \((z_1, z_2, z_3, r) \in Z \times R\). In this example, we approximately have \(|Z_1| = 1280, |Z_2| = 960, |Z_3| = 5, |R| = 4\). Hence, at global Bayesian fusion, over \(2 \cdot 10^6\) values of the unnormalized posterior distribution must be calculated. This proceeding may be unpracticable in real world tasks due to unacceptable storage and computational costs—note also that the sets \(Z_3\) and \(R\) can be much larger in reality.

3.2.2 Application of Focussed Bayesian Fusion

At focussed Bayesian fusion, a theoretically optimal pre-evaluation for a certain kind of information has to base on its probabilistic representation.\(^1\)\(^,\)\(^18\) However, to avoid the costly calculation and also the storage of the global probabilistic representations of the IMINT information which are not representable in a closed parametric form like the prior distribution \(p(z_1, z_2, z_3, r)\) and the Likelihood \(l(d_l|z_1, z_2, z_3, r)\), “car detection” is used as sub-optimal pre-evaluation method.

For “car detection”, Sobel filtering is applied to each image \(d_s\), \(s \in \{2, 3, 4\}\), and the resulting gradient image \(G_s\) is searched for horizontally and vertically aligned edge pairs which may correspond to the lateral outer edges of a car\(^1\) resulting in \(J \in N\) edge pairs. The midpoints \(m_j, j \in \{1, \ldots, J\}\), of the rectangular structures which correspond to the found edge pairs \(e_j\) then indicate possible positions of cars. To account for deviations between the coordinates of the calculated midpoints \(m_j\) and the coordinates of the “true” midpoints of a car, small spatial areas \(U_1^1(j) \times U_2^1(j) \subset Z_1 \times Z_2\) around the calculated midpoints are stated to be possible positions of a car. These areas and the corresponding edge pairs \(e_j\) constitute local regions \(U^1(j) := U_1^1(j) \times U_2^1(j) \times Z_3 \times R^1(j) \subset Z \times R\) which are informative with respect to the position \((z_1, z_2)\) and the driving direction \(r\). We have \(R^1(j) = \{r_1(j)\} \subset R_C\) and \(r_1(j) = “west or east”\) if region \(U^1(j)\) corresponds to a horizontally aligned edge pair \(e_j\), and \(r_1(j) = “north or south”\) if \(e_j\) is vertically aligned.

An explicit pre-evaluation of the HUMINT information is not necessary. It is assumed that the “true” positions of the \(k\) reported cars lie with a high chance within quadratic spatial areas \(\mu_1(k) \times \mu_2(k) \subset Z_1 \times Z_2\) around the reported positions \((\mu_1(k), \mu_2(k)), k \in \{1, \ldots, K\}\). The size of the spatial areas is specified by \(\sigma_\mu(k)\). These areas also constitute local regions \(U^H(k) := U_1^H(k) \times U_2^H(k) \times Z_3 \times R^H(k) \subset Z \times R\). Assuming that the human observer is usually good at differing if a car is driving horizontally or vertically, these regions also mirror the reported information \(r_H(k)\) concerning the driving direction: it holds \(R^H(k) = \{r_H(k)\} \subset R_C\).

All in all, the local context \(U \subset Z \times R\) is then given by \(U := \left(\bigcup_{j \in \{1, \ldots, J\}} U_1^1(j)\right) \bigcup \left(\bigcup_{k \in \{1, \ldots, K\}} U^H(k)\right)\). For \(\sigma_\mu(k) = 0.1 \cdot |Z_1|, k \in \{1, \ldots, K\}, |U| \approx 3 \cdot 10^6\).

For the focussed Bayesian fusion with respect to \(U\), the calculation of \(p(z_1, z_2, z_3, r)\) and \(l(d_l|z_1, z_2, z_3, r)\), respectively, is done directly on the basis of formulas (3) and (5). It has to be stressed that it is sufficient to calculate these quantities for \((z_1, z_2, z_3, r) \in U\). To obtain the necessary extracts of the probabilistic representations of the IMINT information, the result \(l(d_s|z_1, z_2, z_3, r)\) of the matched filtering is to be computed solely for \((z_1, z_2, z_3, r) \in U, s \in \{2, 3, 4\}\).

Calculating the normalized cross correlation coefficient directly in the spatial domain may be computationally expensive. Computational costs can then be significantly reduced by the application of FFT-based methods and the use of integral images, see for example [19]. Here, we use an efficient implementation of normalized cross correlation coefficient matching which is provided by the OpenCV library.\(^20\)

In preparation for the matching, separately for each direction pair \(r_C \in R_C\), the connected parts within the respective spatial areas

\[
A_1(r_C) \times A_2(r_C) := \left(\bigcup_{j \in \{1, \ldots, J\}: r_1(j) = r_C} U_1^1(j) \times U_2^1(j)\right) \bigcup \left(\bigcup_{k \in \{1, \ldots, K\}: r_H(k) = r_C} U_1^H(k) \times U_2^H(k)\right) \subset Z_1 \times Z_2
\]  

\(^1\)The images \(d_2, d_3, d_4\) which exhibit exactly the scene have a size of 1280 px \(\times\) 960 px. Since in the template matching, image borders of a size corresponding to the half template size are not evaluated, \(|Z_1| \times |Z_2|\) is strictly speaking somewhat less than the size of the images.

---

**Proc. of SPIE Vol. 7710  77100A-7**
which are indicated by the local context \( U \) are identified and expanded to a rectangular form of minimal size. This is essentially carried out by the application of a region growing algorithm.\(^\text{17} \) After that, the local context \( U \) is enlarged to a set \( V \) according to the newly created rectangular areas \( V_1^{rc}(l_{rc}) \times V_2^{rc}(l_{rc}) \subset Z_1 \times Z_2 \), \( r_c \in R_C \), \( l_{rc} \in \{1,\ldots,L_{rc}\} \), \( L_{rc} \in \mathbb{N} \). Thereby, each of these areas specifies a local region \( V^{rc}(l_{rc}) = V_1^{rc}(l_{rc}) \times V_2^{rc}(l_{rc}) \times Z_3 \times R^{rc}(l_{rc}) \), \( R^{rc}(l_{rc}) = \{r_c\} \). By construction, \( U \subseteq V = \bigcup_{r_c \in R_c} V^{rc}(l_{rc}) \).

\( V \) is an extension of the local context \( U \) which contains additional elements of \( Z \times R \). These additional elements can be ignored at the focussed Bayesian fusion. However, if the size of \( V \) is not too large in comparison to the size of \( U \), the calculation of matched filtering results \( l(d_s|z_1, z_2, z_3, r) \) for all \( (z_1, z_2, z_3, r) \in V \) via an efficiently implemented realization of normalized cross correlation coefficient matching may be significantly less expensive than the direct calculation of \( l(d_s|z_1, z_2, z_3, r) \) in the spatial domain solely for \( (z_1, z_2, z_3, r) \in U \), \( s \in \{2,3,4\} \).

To guarantee a moderate difference \(|V| - |U|\), the following scheme is recursively applied during the construction of \( V \): starting with \( V_1^{rc}(l_{rc}) \times V_2^{rc}(l_{rc}) \), \( r_c \in R_C \), \( l_{rc} \in \{1,\ldots,L_{rc}\} \), rectangular areas are searched for the presence of a large irrelevant rectangular part. A rectangular part \( B \) satisfies this condition if it constitutes more than \( x \) percent of the area and if it is completely disjunct to the part of the area which is indicated by the local context \( U \). If such a \( B \) is present, it gets removed from the area and the rest of the area gets divided into two, three or four rectangular areas according to the location of \( B \).

### 3.2.3 Evaluation of Focussed Bayesian Fusion

Fig. 4 illustrates the result of the detection and classification which have been solved via a qualitative analysis on the basis of focussed Bayesian fusion. All cars in the scene are detected and correctly classified. By the application of focussed Bayesian fusion, it was possible to restrict the set of values for which the unnormalized posterior distribution has to be computed significantly. The shaded areas in Fig. 4 spatially indicate these values which are contained in the local context \( U \). The corresponding directions are identifiable from the kind of shading. Here, \(|U|\) constitutes less than 15 percent of \(|Z \times R|\). The rectangles in Fig. 4 spatially indicate the extension \( V \) of \( U \), i.e., for which values matched filtering results have been obtained. The corresponding directions are also identifiable from the kind of shading of the respective enclosed shaded areas. For the final constitution of the areas indicated by the rectangles, \( x = 15 \) percent has been chosen. About 89 percent of the computed matched filtering results have been used for the calculation of the unnormalized posterior distribution.

For a completely correct result, it is important that the local context \( U \) contains the “true” position, the “true” type and the “true” driving direction of each car which is present in the scene. As demonstrated, this requirement is fulfilled, here—although the pre-evaluation method “car detection” is suboptimal and although the statements of the human observer concerning the driving direction are accepted to be error free at the
determination of the local regions which correspond to the HUMINT information. This is because the used information sources complement each other with respect of the constitution of the local context \( U \).

With the chosen interference filters, \( d_2, d_3, d_4 \) complement each other with respect to the recognition of cars by the pre-evaluation via “car detection”. However, without the local regions which are obtained from the HUMINT information, the cars of type “DKW” and “Opel” would not have been detected by focussed Bayesian fusion. This is because they are badly silhouetted against the background in \( d_2, d_3, d_4 \) such that they are not found by the pre-evaluation of the IMINT information. Here, all cars which are present in the scene have been detected by the human observer. Of course, this may not always be the case, e.g., if several cars are occluded from his viewpoint. In reality, the detection performance of the human observer may be reduced additionally due to time pressure, stress or time lying between the observation and the reporting. It also has to be stressed that the statements of the human observer concerning the car positions are partially of extremely low quality, here. Hence, even if he detects all present cars, the local regions which correspond to the HUMINT information may not contain all “true” positions.

To obtain a more efficient implementation, focussed Bayesian fusion can be realized distributedly\(^1\) within \( U \). Thereby, a somehow modified region building may be appropriate, e.g., by an additional splitting of large regions into smaller ones. By a variation of the parameters for the constitution of the local context, the size of \( U \) is scalable according to the given resources. There is also the possibility to start with a smaller local context to reach first results within a predefined time and to resume the evaluation later by an extension of the local context to a larger one–of course a somehow modified region building is necessary also here. Also with a somehow modified region building, the presented focussed fusion scheme is applicable if additional information contributions become available after the actual focussed Bayesian fusion process has been started.

4. QUANTITATIVE ANALYSIS BASED ON FOCUSED BAYESIAN FUSION

4.1 Limited Quantitative Meaning of Focussed Probabilities

In a Bayesian model which is focussed on a local context \( U \), the normalized focussed posterior distribution is given by\(^*\)\(^*\) \( p_U(z|d) = \frac{p(z|d)}{P(U|d)} \mathbf{1}_U(z) \). I.e., at focussed Bayesian fusion, the posterior distribution is increased for \( z \in U \) according to a multiplication with the distortion factor \( \frac{1}{P(U|d)} \). Hence, within \( U \), the focussed posterior distribution delivers upper bounds for the values of its global equivalent:\(^3\)

\[
p(z|d) \leq p_U(z|d), \quad z \in U.
\] (**7**)

As discussed in Sec. 3, a qualitative analysis based on Bayesian fusion is done by the examination of posterior probability ratios. Because the distortion factors are equal for all \( z \in U \), probability ratios from a focussed Bayesian model are globally completely meaningful: it holds \( o(z^*, z^{**}) := \frac{p(z^*|d)}{p(z^{**}|d)} = \frac{p(z^*|U)}{p(z^{**}|U)} \) for all \( z^*, z^{**} \in U \). For a quantitative analysis based on Bayesian fusion, absolute posterior probability values must be evaluated. Therefore, the posterior distribution must be normalized. I.e., if a quantitative analysis is based on focussed Bayesian fusion, one has to bear in mind the described distortion of posterior probability values.

4.2 Additional Knowledge Usable for a Quantitative Analysis

Here, the calculation of lower bounds for the values of \( p(z|d) \) is described for the case that the information sources are conditionally independent given \( z \). If this condition does not hold, a similar proceeding is possible.\(^18\) If a theoretically optimal pre-evaluation based on the probabilistic information representations is done, the local context \( U \) can be defined to contain these values \( z \in Z \) for that \( p(z) \) or \( l(d_s|z) \) (for at least one \( s \in \{1, \ldots, S\} \)) is high enough.\(^3\) Thereby, the corresponding proceeding can get unified if the interpretation of the prior knowledge as an additional information contribution is possible.\(^3\) Generally, the introduction of thresholds \( \delta_s \) for the values of the Likelihood functions \( l(d_s|z) \) with respect to \( z \) is critical because there is no normalization requirement for

\[^*\]In the following, \( P(U) = \int_U p(z)dz \) and \( P(U|d) = \int_U p(z|d)dz \) denote the global prior and posterior probability, respectively, of \( U \). In this section, we use an integral notation for all kind of quantities. This is also mathematically correct for discrete quantities because integration with respect to the counting measure results in a summation.

---

Proc. of SPIE Vol. 7710 77100A-9
the Likelihood functions with respect to $z$. In consequence, such thresholds $\delta_s$ have to depend on parameters that give them a relative nature. E.g., the dependency $\delta_s = \delta(\max_{z \in Z} \{ l(d_s | z) \})$ makes sense.\(^3\)

Assuming the global prior probability $P(U)$ which rates the global prior relevance of the local context $U$ to be known, the inequality $l(d_s | z) \leq \delta_s$ for $z \in Z \setminus U$ delivers a lower bound for the global posterior relevance $P(U|d)$ of $U$: from the Bayesian theorem, one obtains\(^{18}\)

$$P(U|d) = \frac{\int_U l(d|z)p(z)dz}{\int_U l(d|z)p(z)dz + \int_{Z \setminus U} l(d|z)p(z)dz} \geq \frac{\int_U l(d|z)p_U(z)dz}{\int_U l(d|z)p_U(z)dz + (1 - \frac{1}{P(U)}) \prod_{s=1}^S \delta_s} =: \beta. \quad (8)$$

Obviously, $\beta$ is computable within the corresponding focussed Bayesian model–provided that $P(U)$ is ratable.

By the use of (8), the factor by that the focussed posterior distribution is distorted within $U$ relative to its global equivalent can be bounded from above: $\frac{P_U(z|d)}{P(z|d)} \leq \frac{1}{\beta}$, $z \in U$. From this, lower bounds for the values of the global posterior distribution within $U$ can be calculated:

$$p(z|d) \geq \beta p_U(z|d), \quad z \in U. \quad (9)$$

### 4.3 Probability Interval Scheme

The combination of the inequalities (7), (8), and (9) delivers a probability interval scheme:

$$p(z|d) \in [\beta p_U(z|d), p_U(z|d)], \quad z \in U, \quad \text{and} \quad p(z|d) \in [0, 1 - \beta], \quad z \in Z \setminus U. \quad (10)$$

For each $z \in Z$, the length $l(z)$ of the probability interval $[a(z), b(z)]$ for $p(z|d)$ is influenced by the posterior relevance $P(U|d)$ of the local context: a large $P(U|d)$ indicates a small $l(z)$ for all $z \in Z$. For $z \in U$, $l(z)$ also depends on the respective value of the global posterior distribution: the larger $p(z|d)$, the larger is $l(z)$. See [18] for a proof of these results. The last conclusion is further optimizable because, within $U$, the interval lengths mirror the probability ratios: for $z \in U$, it holds $l(z) = p_U(z|d)(1 - \beta)$. Hence, we have $\frac{l(z)}{l(z^*)} = o(z^*, z^{**})$ for $z^*, z^{**} \in U$.

It is not possible that the values of $p(z|d)$ arbitrarily lie within the corresponding intervals. To demonstrate this, let $z^* \in U$ denote an arbitrary but fixed element of the local context. If $p(z^*|d)$ adopts a fixed value $p^* \in [a(z^*), b(z^*)]$, it has to hold

$$p(z|d) = o(z, z^*) p^*, \quad z \in U, \quad \text{and} \quad p(z|d) \leq 1 - p^* \int_{\zeta \in U} o(\zeta, z^*) d\zeta, \quad z \in Z \setminus U. \quad (11)$$

To demonstrate the utility of the probability interval scheme practically, the following deliberately very simple example will be used: the task is the classification of a car by a focussed Bayesian fusion of prior knowledge and IMINT information $d_1$. Here, the PoIs have the form $Z = Z_1$ whereby $Z_1$ is the set of possible car types. As in Sec. 3.2, it is assumed that it holds $Z_1 = \{“Audi”, “DKW”, “Ford”, “Mercedes”, “Opel”\}$. Because no prior knowledge concerning $z$ is available, the ME principle delivers for $p(z)$ the uniform distribution over $Z$. $d_1$ is given by a gray-value image of the car and, as in Sec. 3.2, its probabilistic representation is obtained by normalized cross correlation coefficient matching with a matched filter bank which is adjusted to $Z$. The templates correspond to gray-value images of the different possible cars. They have the same size as $d_1$ here. $z \in Z$ is ignored at focussed Bayesian fusion if it holds $l(d_1|z) \leq \delta_1$ with $\delta_1 = \epsilon \cdot \max_{z \in Z} l(d_1|z)$. To illustrate the effect of this thresholding, the resulting probability intervals are illustrated in Fig. 5 for various choices of $\epsilon$. For $\epsilon = 0.0227$ and $\epsilon = 0.1$, it holds $U = \{“Audi”, “Ford”, “Opel”\}$. For $\epsilon = 0.1129$, it holds $U = \{“Audi”, “Opel”\}$, i.e., the type “Ford” is ignored additionally. For $\epsilon = 0.7$, it holds $U = \{“Opel”\}$, i.e., the type “Opel” is taken for sure within this focussed Bayesian model. In all cases, its possible to identify the global maximum a posteriori estimate “Opel” by focussed Bayesian fusion. This holds because “Opel” is contained in $U$ and $b(“Opel”) \geq b(z)$ for all $z \in U$–note that the question if the probability intervals overlap is not relevant for
Figure 5. The bars illustrate the probability intervals for various choices of $\epsilon$. The exact value of the posterior distribution are indicated by the color change of the bars. The white crosses indicate the values of the ME distribution. (A: “Audi”, O: “Opel”, D: “DKW”, F: “Ford”, M: “Mercedes”)

this conclusion. Apart from the influence of $P(U|d)$ and $p(z|d)$ on the size $l(z)$ of the probability intervals, the sharpness of the thresholding also influences $l(z)$, see especially Fig. 5(a) and Fig. 5(b). The reason for this is that it influences via $\beta$ the lower interval bounds $a(z)$. $\epsilon = 0.0227$ delivers a quite sharp threshold therefor that exactly the types “DKW” and “Mercedes” are ignored at focussed Bayesian fusion. From Fig. 5(a), it can be seen that also in this case the values of $p(z|d)$ lie distant from $a(z)$ and $b(z)$. This effect holds because, apart from the identity $l(d_1|”Mercedes”) = \epsilon \cdot \max_{z \in Z} l(d_1|z)$, it nevertheless holds $l(d_1|”DKW”) \ll \epsilon \cdot \max_{z \in Z} l(d_1|z)$. Analogously, $\epsilon = 0.1129$ delivers a quite sharp threshold therefor that the type “Ford” is ignored additionally. However, because the resulting threshold is less sharp with respect to $l(d_1|”DKW”) and $l(d_1|”Mercedes”) than a threshold which corresponds to smaller values of $\epsilon$ and because $P(U|d_1)$ is reduced, the expressiveness of the resulting probability interval scheme is much more lesser than in the both foregoing cases, see Fig. 5(c). Clearly, the probability interval which corresponds $\epsilon = 0.7$ is little expressive, see Fig. 5(d).

In [18], a strategy for the identification of globally optimal actions on the basis of the probability interval scheme has been discussed. This method is not optimal—in particular because the constraint (11) is not considered. In consequence, there may be less actions identified as being globally optimal than it was theoreticly possible. Here, we abandon the discussion how (11) can get incorporated within the described strategy. Instead, it is shown that it is possible to reduce the probability intervals adequately to a single probability distribution prior to a possible subsequent decision making.

The probability intervals are reducible to a unique probability distribution via the ME principle, see also Fig. 5. In all cases, the ME distribution corresponds to the lower bounds of the probability intervals, here. I.e., the ME principle rates the posterior relevance of the local context as low as the interval scheme allows. The respective entropy maximization under the constraints (10) and (11) corresponds to the maximization of a term which solely depends on $p^* \in [a(z^*), b(z^*)]$. Because of the grouping axiom for entropy, the ME distribution is constant for $z \in Z \setminus U$. Of course, also additional knowledge like knowledge concerning the quality of the estimation (8) may get incorporated in the ME analysis, in principle.

In a sense, also choosing the midpoints of the probability intervals for $z \in U$ is consistent with the ME principle: if for $z \in U$, the value $p(z|d)$ is treated as a random quantity, the ME principle delivers the uniform distribution $U([a(z), b(z)])$ for it. A subsequent reduction of the resulting distributions for $p(z|d)$ by expectation building delivers the midpoints $\frac{b(z) - a(z)}{2}$. Note that the resulting posterior distribution is consistent with (11).

5. CONCLUSION

The fusion of several information sources is a promising technique and the Bayesian approach to information fusion provides many benefits. To reduce the disadvantage of the Bayesian methodology which is its global view on the fusion problem, local approaches are investigated. Focussed Bayesian fusion is a local Bayesian
fusion method which is straightforward realizable. Qualitative statements on the basis of a focussed Bayesian model can be easily obtained. Provided the validity of a focussed Bayesian model, such qualitative statements are globally completely meaningful. If quantitative statements shall be obtained by focussed Bayesian fusion, the corresponding distortion of probability values must be considered. Quantitative statements get additional expressiveness if an interval scheme for the global probability distribution is calculated. Such an interval scheme is obtainable within a focussed Bayesian model if the prior relevance of the local context is ratable and if a construction rule for the focussed Bayesian model is known. The ME principle delivers funded mechanisms for the reduction of a probability interval scheme to a single distribution.

REFERENCES