

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# A Possibilistic Framework for Sensor Fusion with Monitoring of Sensor Reliability

Volker Lohweg<sup>1</sup>, Karl Voth<sup>2</sup> and Stefan Glock<sup>3</sup>

<sup>1,2</sup>*Ostwestfalen-Lippe University of Applied Sciences, inIT - Institute Industrial IT, Lemgo*

<sup>3</sup>*Friedrich-Alexander University of Erlangen-Nuremberg, Institute for Electronics  
Engineering, Erlangen  
Germany*

## 1. Introduction

Sensor and Information fusion is recently a major topic, not only in traffic management, military, avionics, robotics, image processing, and e. g. medical applications, but becomes more and more important in machine diagnosis and conditioning for complex production machines and process engineering. Several approaches for multi-sensor systems exist in the literature, cf. (Hall & Llinas, 2001); (Bossé et al., 2007).

The diagnosis and flaw detection in machines and process plants is a complex task, which is dependent on many effects. In the past decades, autonomous and automatic systems have become ubiquitous in our daily life. As famous examples, one may cite advanced driver-assistance systems in vehicles, quality monitoring in production processes, or semi-autonomous unmanned aircraft systems in military forces. All these systems have the common feature that they (partially) capture the state of the environment and generate fused information based on the data gained. More technically, the analogous environment is captured applying several sources (sensors, experts, etc.). After obtaining these analogous signals and experts' information, they are preprocessed and transformed into the digital domain, so that they can be processed on computers. In many cases the information captured from the environment may be imprecise, incomplete or inconsistent. Furthermore, signal sources may be not reliable. Therefore, it is necessary to extend known fusion concepts insofar that they are able to handle and to measure imprecision and reliability. In this chapter we will highlight, to our best knowledge, a new human-centric based fusion framework. Additionally, an application of the fusion method is shown for printing machines in the area of quality inspection and machine conditioning. The goal is the early recognition of errors in machines in order to avoid flaws by combining measuring data from several sensors with expert knowledge for the improvement of quality by

- using several information sources,
- integrating expert knowledge and perception,
- extracting reasonable features, and
- generating intuitive results.

This chapter investigates *context-based anticipatoric multi-sensor fusion* (CBAMSF) using evidence-based theories. A new algorithmic concept which considers a sensor's imprecision and process vagueness is presented and compared with other Fuzzy and Evidence based algorithms. In particular, the algorithm is inspired by the human behaviour in group-decision situations. This behaviour enables the algorithm to discount defective sensors. The performance of the novel fusion algorithm is evaluated with reference to real-life applications. In this context, the conditions of an intaglio printing machine (banknote printing machine) are monitored whilst the machine is producing security prints.

The sections are organised as follows: After a short introduction aleatoric (statistical) uncertainty versus epistemic (subjective, human-oriented) uncertainty is defined and explained. Furthermore, uncertainty in diagnosis of technical systems is highlighted. Thereafter, the concept of evidence theories is pointed-out in a short review—namely: *Probability Theory, Belief (Dempster-Shafer) Theory, Fuzzy Set Theory and Possibility Theory*. In the next sub-section we describe fusion algorithms which are based on human decision processes (Two-Layer Conflict Solving). From these concepts we derive a possibilistic framework for sensor fusion with reliability strategies. The following sub-section describes and compares a certain application. The succeeding section concludes the chapter and gives an outlook on further research topics.

## 2. Taxonomy of uncertainty

In today's production world we are able to generate a huge amount of data from analogue or digital sensors, PLCs, middleware components, control PCs, and, if necessary, from ERP systems. However, creating reliable knowledge about a machine process is a challenge because it is a known fact that  $Data \neq Information \neq Knowledge$ .

Insofar, a fusion process must create a low amount of data which creates reliable knowledge. Usually the main problems in sensor fusion can be described as follows: Too much data, poor models, bad or too many features, and improperly analysed applications. One major mis-belief is that machine diagnosis can only be handled based on the generated data—knowledge about the technical, physical, chemical or other processes are indispensable for modelling a multi-sensor system.

### 2.1 Knowledge and ignorance

Specialists in Engineering and Computer Science world accentuate the value of data, assuming that all necessary information about behaviour of a system, etc. is sufficiently contained in the data. Insofar, it is believed that information is an equivalent for knowledge. Therefore, human beings tend intentionally or unintentionally to sweep aside uncertainty in information or ignorance about a state of affairs of a (technical) system. In this sense, we overestimate the amount of information in data which is usable. Following these explanations, it is straightforward to cite Stephen Hawking: "*The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge*". The human knowledge and ignorance behaviour is described in Fig. 1. Referring to (Ayyub & Klir, 2006), the evolutionary certain knowledge (ECK) about a system or state of affairs can intrinsically not be interfered or camouflaged due to its evolutionary nature of this kind of knowledge. The current state of reliable knowledge (RK) has to be interpreted as trained knowledge under the general condition of provable know-how within well-grounded reliability levels, cf. Eq. (1).

The intersection of ECK and RK defines the human knowledge base of certain claims and propositions (CCK). It is assumed that this kind of knowledge withstands any

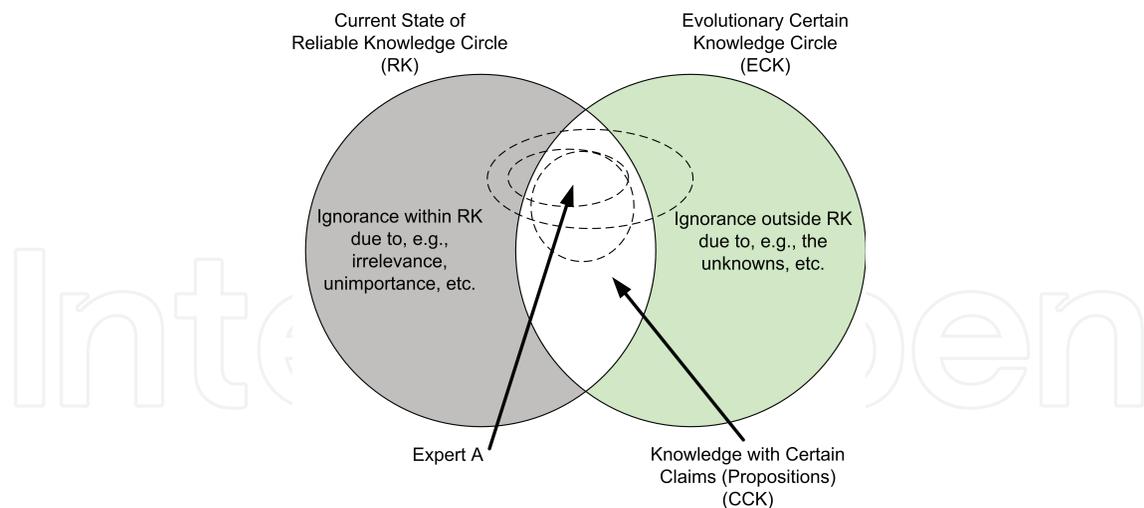


Fig. 1. Human evolutionary and trained knowledge. (Ayyub & Klir, 2006)

paradigm change or temporal opinion shifts. It can be formulated as:

$$CCK = ECK \cap RK. \quad (1)$$

Referring to Fig. 1, two types of ignorance can be distinguished: i) ignorance within the RK circle, based on irrelevance, etc.; and ii) ignorance outside RK, based on unknowns, missing know-how, etc. An expert (Expert A in Fig. 1), equipped with some know-how about a system or state of affairs, pictured with ellipses in Fig. 1, can be represented in three ways; 1) equipped with trained, created, or captured knowledge as a subset of ECK; 2) equipped with self-perceived knowledge by the expert; and 3) equipped with perception by others of the expert's knowledge. As (Ayyub & Klir, 2006) pointed out, the ECK of an expert may be smaller than the self-perceived knowledge, and therefore, the difference of the two types can be interpreted as a measure of overconfidence. Furthermore, an ellipse can also extend beyond the RK base as a result of creativity and imagination. The difference between the expert's knowledge and the ignorance space outside CCK can be interpreted as a human intuition process (creative process). It is unquestionable, that other experts will have other individual ellipses which may or may not overlap each other. If the ellipses of two or more experts do not overlap, based on a conjoint topic, a conflict situation may occur. Usually two types of ignorance can be identified: i) *blind ignorance* with its sub-categories *unknowable*, *irrelevance*, and *fallacy (camouflage)*; ii) *conscious ignorance* with its sub-categories *inconsistency (confusion, conflict, inaccuracy)* and *incompleteness (absence, unknowns, uncertainty)*.

In technical systems the latter ones are of main interest. Inconsistency occurs mainly in the case of inaccurate measurements and conflict-afflicted sensor information. Incompleteness of information, in the sense of uncertainty, causes in most instances problems in system modelling. Uncertainty is defined as *intrinsic absence of necessary knowledge*. In many cases, as, e.g. in complex systems, it is usually impossible to acquire all necessary information to create certain knowledge. Therefore, major attention should be given to this type of ignorance. As it was stated in (Klir & Wierman, 1998) uncertainty can rarely be avoided, when dealing with real-world problems. At sensory and pre-processing level, uncertainty is an inseparable companion of any measurement, resulting from measurement errors, stochastic effects, such as noise and resolution limits in terms of sampling and quantisation (Lohweg & Mönks, 2010a). At knowledge level (cognitive level), we are conditioned by vagueness and ambiguities in the description of processes or natural languages.

The concept of uncertainty was and is philosophically widely discussed in the community of Bayesian and Probability advocates.

Bayesians argue that it is invariably feasible to design a probability model for any on random variables based technical task. This argumentation is questionable in a sense that it does not fulfil a real-world scenario in all cases. Usually at least two kinds of uncertainty species have to be taken into account: i) the aleatoric uncertainty (random effects, signal noise, etc.); and ii) the epistemic uncertainty (lack of knowledge or incomplete knowledge regarding a complex system, etc.), cf. (Hacking, 1975).

The latter one is challenging in problem solving, because the unknown knowledge has to be modelled in different ways, other than probability based concepts. This argumentation leads to Shafer (Shafer, 1976), Zadeh (Zadeh, 2008) and others, who stated that the lack of knowledge is precisely reflected by the situation where the probability of events is ill-known or expert's knowledge is necessary. Furthermore, recently Salicone has presented a possibilistic approach to handle measurement uncertainty in electrical engineering applications (Salicone, 2007).

### 2.1.1 Aleatoric uncertainty

If data is complete and intrinsically non-deterministic in nature, it can be assumed as random (stochastic). The uncertainty is attributed to real-world phenomena and it can not be reduced or even eliminated by expanding an underlying knowledge base. Probabilistic approaches, such as classical Probability Theory (frequentist) and Bayesian Probability Theory are an effective way to model stochastic uncertainties, like measurement noise, etc., cf. (Ayyub & Klir, 2006). This type of uncertainty is referred to as *aleatoric uncertainty* (cf. Table 1).

### 2.1.2 Epistemic uncertainty

In many situations we lack information, that is, not all intrinsically necessary knowledge is available at state. In this case, the uncertainty range should be reduced by expanding the underlying knowledge base. When data is scarce the probabilistic approach may not be appropriate to reduce the system's uncertainty. Major types of this uncertainty are inconsistent and incomplete data as well as inconsistent information or knowledge. In many cases this uncertainty can be reduced by multi-sensory fusion and expert's knowledge, cf. (Ayyub & Klir, 2006). This type of uncertainty is referred to as *epistemic uncertainty* (cf. table 1).

	<i>Aleatoric Uncertainty</i>	<i>Epistemic Uncertainty</i>
<i>Type</i>	irreducible	reducible
<i>Data</i>	random, stochastic	scarce
<i>Origin</i>	intrinsic variations in data	inconsistent & incomplete data, lack of knowledge
<i>Model</i>	Probability Theory	Evidence and Fuzzy Theories

Table 1. Classification of aleatoric and epistemic uncertainty.

## 3. Basics on Evidence Theory

The Evidence Theory was introduced by Glenn Shafer in the 1970s, summarised and condensed in the book *A Mathematical Theory of Evidence* in 1976 (Shafer, 1976). He extended the seminal work of Arthur P. Dempster from the 1960s about upper and lower bounds of probability measures (Dempster, 1967). Today many quantitative approaches are available to

model uncertainty and knowledge processing in the light of sensor and information fusion. It is an interesting fact that many approaches have a common theoretical base.

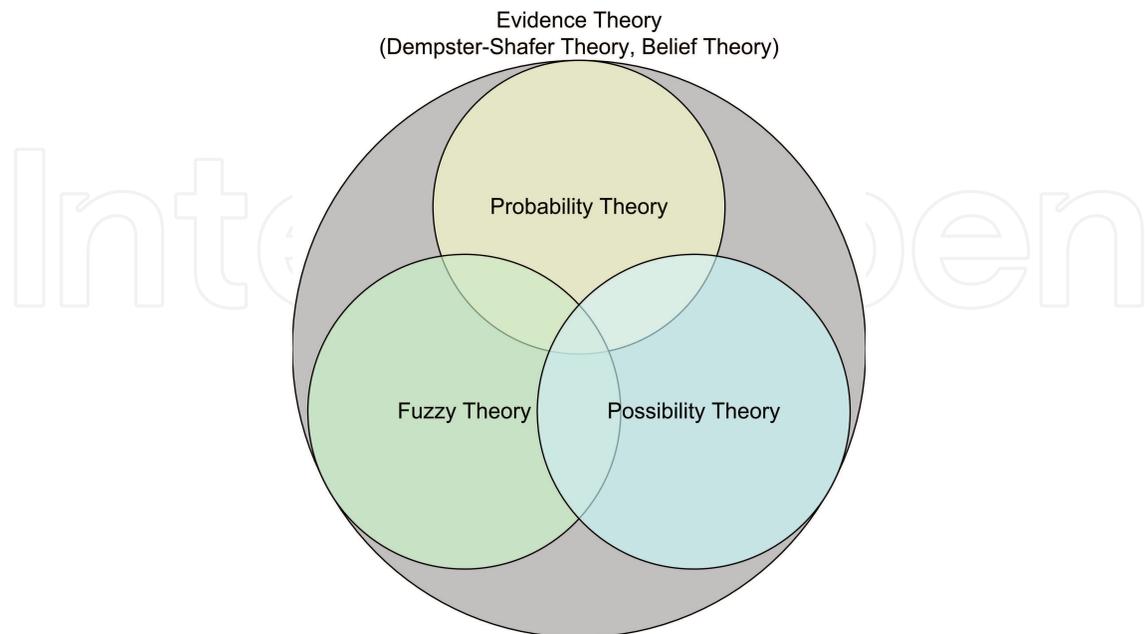


Fig. 2. Theories embedded in evidence theory.

Dempster-Shafer Theory (sometimes called *Belief Theory*) serves as an interconnection between many quantitative approaches, namely: Probability Theory, Possibility Theory, Fuzzy Theory, and others. The above mentioned theories are (partly) special branches of the Evidence Theory (Dempster-Shafer Theory). Fig. 2 illustrates symbolically the interconnection of different approaches insofar they are needed in this chapter.

### 3.1 Dempster-Shafer (Belief) Theory

Data fusion deals with data which is received from sensors, experts or human linguistic words, etc. Furthermore, much of such knowledge is cognitive and imprecise (incomplete) to some degree. To deal with uncertain knowledge, various strategies based on Dempster-Shafer Theory (DST) (Dempster, 1967), (Shafer, 1976), are researched and applied, because DST is capable of managing uncertainty due to its framework. DST acts as the pioneer in data fusion algorithms, which was proposed by Dempster and extended by Shafer subsequently. Serving as a seminal fusion approach, DST stirs up many discussions and studies in data fusion. DST is a mathematical theory of evidence, which combines independent sources of information (Dempster, 1967), (Shafer, 1976). With combination of evidence sources obtained from sensors (experts), more reliable and convincing fusion results are expected. First a finite frame of discernment, which forms a set  $\Theta$ , is defined,  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with an arbitrary event variable (proposition, hypothesis)  $\theta_i, i \in \mathbb{N}$ . A power set,  $\mathcal{P}(\Theta) = 2^\Theta$ , includes all the possible combinations of propositions  $\theta_i$ . Propositions are regarded to be *mutually exclusive* and *exhaustive*. Thus, the power set contains  $2^n$  elements. As an example, assume the following:  $\Theta = \{A, B\} \rightarrow \mathcal{P}(\Theta) = \{\emptyset, A, B, \Theta\}$ .

A function  $m : 2^\Theta \rightarrow [0, 1]$  is called a mass function, also known as Basic Probability (Belief) Assignment (BPA, BBA) of an event, e. g.  $A$ , with

$$m(\emptyset) = 0, \quad \sum_{A \subseteq \mathcal{P}(\Theta)} m(A) = 1. \quad (2)$$

If there is no element in the BBA, then its mass is zero. On the other hand, as  $\mathcal{P}(\Theta)$  is a power set composed of all the subsets, the sum of all the masses must be equal to one. Furthermore, the focal element (mass is larger than zero) is defined as:

$$\{(A, m(A)) | A \subseteq \Theta \text{ and } m(A) > 0\}. \quad (3)$$

A belief function,  $\text{Bel}(\cdot): 2^\Theta \rightarrow [0, 1]$ , satisfies the following three axioms (cf. Eq. (4) - Eq. (6)):

- Axiom 1:  $\text{Bel}(\emptyset) = 0.$  (4)

- Axiom 2:  $\text{Bel}(\Theta) = 1.$  (5)

- Axiom 3:  $\text{Bel}(A \cup B) \geq \text{Bel}(A) + \text{Bel}(B)$  if  $A \cap B = \emptyset.$  (6)

Contrary to the probability theory, the belief measure is *super-additive*. That is, the belief of the union of two propositions may be larger than the addition of the two belief measures of each proposition  $A$  and  $B$ .

Belief ( $\text{Bel}$ ) and plausibility function ( $\text{Pl}$ ),  $\text{Pl}(\cdot): 2^\Theta \rightarrow [0, 1]$ , are essential concepts in DST, which are used in decision-making. Plausibility has to be interpreted as the dual function of the belief.

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B) \text{ and } \text{Pl}(A) = 1 - \text{Bel}(A^c), \quad (7)$$

where  $A^c$  is the complement of proposition  $A$ .

In DST,  $\text{Bel}(A)$  has to be interpreted after Shafer (Shafer, 1976) as “a person’s (sensor information) degree of belief that the truth lies in  $A$ ”, whereas  $\text{Pl}(A)$  can be considered as a measure of maximum support for a proposition  $A$  under the constraint of disbelief in  $A$ . Thus,  $\text{Bel}(A)$  is the total belief mapped to  $A$ ;  $\text{Pl}(A)$  is the total mass assignment, which can move into  $A$ , and  $m(A)$  is the basic belief assignment, which is exclusively dedicated to  $A$ . Furthermore, plausibility can be derived as follows:

$$\text{Pl}(A) = 1 - \text{Bel}(A^c) = 1 - \sum_{B \subseteq A^c} m(B) = \sum_{B \subseteq \mathcal{P}(\Theta)} m(B) - \sum_{B \subseteq A^c} m(B) = \sum_{B \cap A \neq \emptyset} m(B). \quad (8)$$

From Eq. (7) the following can be directly derived:

$$\text{Bel}(A^c) + \text{Pl}(A) = 1 \text{ and } \text{Bel}(A) + \text{Pl}(A^c) = 1. \quad (9)$$

$\text{Bel}(\cdot)$  is, in an early interpretation, called lower bound probability, while  $\text{Pl}(\cdot)$  is the upper bound probability, for the reason that  $\text{Bel}(\cdot)$  is the “must-be”-probability and on the other hand  $\text{Pl}(\cdot)$  is the “might-be”-probability. Therefore,  $\text{Pl}(\cdot)$  includes more mass than  $\text{Bel}(\cdot)$ , which is illustrated with

$$\text{Bel}(A) \leq \text{Pl}(A). \quad (10)$$

The *Ignorance*  $\text{Igr}(A)$  can be interpreted as

$$\text{Igr}(A) = \text{Pl}(A) - \text{Bel}(A). \quad (11)$$

After obtaining the above-mentioned concepts, we are able to use DST to fuse independent data sources by applying *Dempster's Rule of Combination*, cf. (Shafer, 1976). It is also called normalised conjunctive combination rule:

$$\bigoplus_{i=1}^n m_i(A) = K \cdot \sum_{A_1 \cap \dots \cap A_n = A} \prod_{i=1}^n m_i(A). \quad (12)$$

The term  $\sum_{A_1 \cap \dots \cap A_n = A} \prod_{i=1}^n m_i(A)$  aggregates the consonant opinions (non-conflicting parts) from sensors which are subsequently multiplied with *conflicting factor*  $K = (1 - k_c)^{-1}$ , where

$$k_c = \sum_{A_1 \cap \dots \cap A_n = \emptyset} \prod_{i=1}^n m_i(A_i). \quad (13)$$

The variable  $m_i(A_k)$  denotes the mass of propositions from sensor  $i$ . According to the definition of  $k_c$  it defines the empty intersection of the propositions of all sensors. Therefore, it is also called *conflicting coefficient*.

The DST offers two main advantages: i) belief and possibility model two levels of probability, generating and assigning ignorance (range of uncertainty) or uncertainty to a proposition  $A$ , furthermore *total uncertainty* can be modelled by the vacuous belief, that is,  $m(\Theta) = 1$ ; ii) as the measures are defined on subsets of  $\mathcal{P}(\Theta)$ , non-specificity and conflict can be modelled.

In order to demonstrate the concept of DST, let us assume a banknote was given to an expert for forgery analysis. Is it genuine or is it counterfeit? The proposition  $A$  corresponds to the possibility that the banknote is genuine, whereas the proposition  $B$  reflects the proposition that the banknote is counterfeit; then  $\Theta = \{A, B\}$  is the frame of discernment and  $\mathcal{P}(\Theta) = \{\emptyset, A, B, \Theta\}$  is the corresponding power set. The belief functions  $\text{Bel}(A)$  and  $\text{Bel}(B)$  represent the belief measures in genuineness and counterfeit of the banknote.

As, in this case,  $B = A^c$ , Eq. (6) changes to  $\text{Bel}(A \cup A^c) \geq \text{Bel}(A) + \text{Bel}(A^c)$  and  $A \cap A^c = \emptyset, A \cup A^c = 1 \Rightarrow \text{Bel}(A) + \text{Bel}(A^c) \leq 1$ . If the expert assumes the genuineness of the banknote, one would set  $\text{Bel}(A)$  near to 1 and—independently of  $\text{Bel}(A)$ — $\text{Bel}(A^c)$  to a low value. In case one believes in the opposite, one sets  $\text{Bel}(A^c)$  to a high value and  $\text{Bel}(A)$  to a low value. If the expert is uncertain about the decision, but tends to genuineness, one might vote for  $\text{Bel}(A) = 0.5$  and  $\text{Bel}(A^c) = 0.3$ , and in the desperate case of total uncertainty (total ignorance), one will set both belief measures to zero.

It has to be pointed out that the literature on DST is manifold. For deeper insights we refer to, e. g., (Dempster, 1967), (Shafer, 1976), (Yager, 1987), (Hall & Llinas, 2001), (Ayyub & Klir, 2006), (Campos, 2006), (Yager & Liu, 2008).

### 3.2 Probability Theory

Probability Theory (PT) is the oldest concept for quantifying uncertainty. It is one of the last major mathematical domains to be formed at the end of the seventeenth century. Based on a correspondence between Fermat and Laplace about gambling, Probability Theory is also today an active field of research. PT typically models *random phenomena* and *chance*.

Although, PT is an “old” theory, it was put into question in the twentieth century, whether the Probability Theory can only be viewed as a theory of chance. In every day situations it is used in two different ways: i) when randomness is assumed; ii) in cases of lacking knowledge. Therefore, two schools of interpretation can be found: objectivists (frequentists) and subjectivists (Bayesians). Hacking (Hacking, 1975) calls probability a “Janus-faced” theory. We refer to the philosopher Karl Popper (Popper, 1934) on the one hand and the

subjectivists John Maynard Keynes (Keynes, 1921) as well as to Rudolf Carnap (Carnap, 1947) or Frank P. Ramsey (Ramsey, 1960) on the other hand. The dispute in the community still exists.

Based on random experiments, like throwing the dice or tossing a coin, classical probability is defined by the relative frequency of events in the long run. If  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  denotes the so called *sample space*, the discrete, finite, non-empty set of all possible outcomes, then  $\theta_i, i \in \mathbb{N}$ , are the elementary events (singletons). A non-negative number  $P(A)$ ,  $P : 2^\Theta \rightarrow [0, 1]$ , is assigned to each subset  $A \in \Theta$ —an event of any collection of outcomes:  $A = \{\theta \in \Theta \mid \theta \in A\}$ . This number is called *probability of the event A*. The axiomatic approach of probability was formulated by Kolmogorov in 1933 (Kolmogorov, 1933). The number  $P(A)$  must satisfy the three following axioms; cf. Eq. (14) - Eq. (16):

- Axiom 1:

$$P(\emptyset) = 0. \quad (14)$$

- Axiom 2:

$$P(\Theta) = 1. \quad (15)$$

- Axiom 3:

$$P(A \cup B) = P(A) + P(B) \text{ if } A \cap B = \emptyset. \quad (16)$$

The probability measure of two (or more) is *additive*. This property, defined by axiom 3, is one of the key features in Probability Theory. Once a probability  $P(A)$  is set, the complement  $P(A^c)$  is fixed. In other words, the event  $\theta \in \Theta$  can either be *true* or *false*. If for each possible outcome  $\theta \in \Theta$ , the probability  $p(\theta)$  is given, then a function  $p : \Theta \rightarrow [0, 1]$  is called *probability density*. The function  $p(\theta) = P(\{\theta\}), \forall \theta \in \Theta$  satisfies the following conditions:

$$0 \leq p \leq 1, \quad (17)$$

$$\sum_{\theta \in \Theta} p(\theta) = 1. \quad (18)$$

Following the additivity axiom, the probability of  $A$  can be determined by its individual probabilities of each of its elementary events:

$$P(A) = \sum_{\theta \in A} p(\theta). \quad (19)$$

Again, we follow the example of genuine or counterfeit banknotes. The proposition  $A$  corresponds to the possibility the banknote is genuine, whereas the proposition  $B$  reflects the proposition that the banknote is counterfeit; then  $\Theta = \{A, B\}$  is the set of events. The probabilities  $P(A)$  and  $P(B)$  represent the chances in genuineness and counterfeit of the banknote. As, in this case,  $B = A^c$ , Eq. (16) changes to  $P(A \cup A^c) = P(A) + P(A^c)$  and  $A \cap A^c = \emptyset, A \cup A^c = 1 \Rightarrow P(A) + P(A^c) = 1$ . If the expert assumes the genuineness of the banknote, one sets  $P(A)$  near to 1 and  $P(A^c) = 1 - P(A)$ . In case one believes in the opposite, one sets  $P(A^c)$  to a high value and  $P(A) = 1 - P(A^c)$ . If the expert is uncertain about his decision, the only thing one can do, is to set  $P(A) = 0.5$ , that is,  $P(A^c) = 0.5$  (total ignorance). The above mentioned property is essential in Probability Theory.

In the subjective view, the Bayesian view, knowledge and uncertainty is represented based on probability measures  $P(\cdot)$ , where the measure describes a *likelihood* of the occurrence of an event  $A$ . In this context,  $P(\cdot)$  is often referred to as *a-priori probability* which is determined on previous knowledge based on adequate data *and* taken for granted. The probability

of an event  $A$  is interpreted in a sense of *degree of certainty* or *degree of belief*. Knowledge about the occurrence of an event is either modelled by  $P(A) = 1$  in case of a certain event or  $P(A) = 0$  for an impossible event. Once an event  $A$  is given by a certain probability  $P(A)$ , the complement  $P(A^c)$  is naturally given. Therefore, the only possibility to model uncertainty in a set  $\Theta$  is, as there is no evidence for a certain event, the probability density  $p(\theta) = \frac{1}{|\Theta|}, \theta \in \Theta$ . The cardinality of the set is  $|\Theta|$ . The modelling of uncertainty is a weakness of Probability Theory. As wider the set  $\Theta$  is, as more improbable is a single event  $\theta_i$ , but this is counter-intuitive, because a certain event will occur to  $p(\theta_i) = 100\%$  and not to  $p(\theta_i) = \frac{1}{|\Theta|}$ .

The basis of fusion in PT is well-grounded on the *Bayes Rule* (Bayes, 1763). Two events  $A \in \Theta_1$  and  $B \in \Theta_2$  are assumed to be conjoint,  $A \cap B \neq \emptyset$ . That is, that the joint probability can be determined by the joint elements  $\theta$  in the joint set  $\Theta = \Theta_1 \times \Theta_2$ , defined by the cross product, and is denoted by  $P(A \cap B)$ .

The *conditional probability*  $P(A|B)$  is defined as the probability of  $A$  being true under the condition of the event  $B$ . It is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0, P(A) > 0. \quad (20)$$

The conditional probability is called *a-posteriori probability* of the event  $A$ . That is, that the probability is not only based on the previous knowledge, but also takes into account the evidence of the event  $B$ .

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)}. \quad (21)$$

In Eq. (21)  $P(A)$  is usually defined as the *total probability*, which is defined as

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i). \quad (22)$$

It has to be pointed out, that  $B_1, \dots, B_n$  have to be *pairwise exclusive* and *mutually exhaustive*. Furthermore,  $P(B_i) > 0 \forall i$ . Now, we show basically how information is fused in the sense of Bayesian inference. Let  $F_i$  be a measured machine readable feature on a banknote (e.g. a certain coding) which has to be detected by an optical sensor  $S_i$  as a "true"— $S_i$ —or "false"— $S_i^c$ —information. Furthermore, it can be assumed, that the accuracy of the sensors  $P(S_i|F_i)$  may be 0.9, that is, the probability of a correct measurement (the sensor delivers a correct result under the condition of a correct code) is 90%. Then, the probability of the detection of the feature  $F_i$  under the condition of the correct sensor information is:

$$P(F_i|S_i) = \left( \frac{P(S_i|F_i)}{P(S_i)} \right) \cdot P(F_i). \quad (23)$$

With Eq. (20) and Eq. (22) the sensor probability  $P(S_i) = P(S_i \cap F_i) + P(S_i \cap F_i^c)$  can readily be formulated:

$$P(F_i|S_i) = \frac{P(S_i|F_i) \cdot P(F_i)}{P(S_i \cap F_i) + P(S_i \cap F_i^c)} = \frac{P(S_i|F_i) \cdot P(F_i)}{P(S_i|F_i)P(F_i) + P(S_i|F_i^c)P(F_i^c)}. \quad (24)$$

The a-priori probability  $P(F_i)$  is given under the assumption, that it is very likely that the correct feature is used in the measurement session (e.g. 99.5%), that is,  $P(F_i^c) = 0.005$ .

Furthermore, as  $P(S_i|F_i) = 0.9$  is assumed, the probability  $P(S_i|F_i^c) = 0.1$ , that is, the detection of a feature which is not present. It has to be pointed out, that all four combinations, based on the equivalent a-priori probabilities, have to be determined:  $P(F_i|S_i)$ ,  $P(F_i|S_i^c)$  (false positive),  $P(F_i^c|S_i)$  (false negative), and  $P(F_i^c|S_i^c)$  to get the full picture.

In case of multiple measurements for one feature we get  $P(F_1|S_1, S_2)$ , which can be interpreted, e. g., as two temporarily *independent* measurements  $P(S_1, S_2) = P(S_1) \cdot P(S_2)$ . In this case we obtain:

$$P(F_1|S_1, S_2) = \frac{P(S_1, S_2|F_1) \cdot P(F_1)}{P(S_1, S_2)} = \frac{P(S_1|F_1)P(S_2|F_1) \cdot P(F_1)}{P(S_1)P(S_2)} = \frac{P(S_2|F_1)P(S_1|F_1) \cdot P(F_1)}{P(S_2)P(S_1)}. \quad (25)$$

With Eq. (23)  $P(F_1|S_1, S_2)$  can be simplified to

$$P(F_1|S_1, S_2) = \left( \frac{P(S_2|F_1)}{P(S_2)} \right) \cdot P(F_1|S_1). \quad (26)$$

The conditional probability  $P(F_1|S_1)$  from the first measurement is combined with the appropriate probabilities of the second measurement to generate a fused result. The first conditional probability serves as the new a-priori probability for the second measurement. This is how Bayesian inference combines old information with new information to update the belief states.

Probability and its frequentist representations concerns random effects. The events in Probability Theory are either assumed to be true or false. The probability of an event is connected directly to the event and to its complement. The Bayesian view leads to modelling of uncertainties. However, Probability Theory has its weakness in the area of epistemic uncertainty because of its additivity axiom. A-priori probabilities are the decisive part in PT. Furthermore, getting these a-priori probabilities is not easy to achieve. Probability Theory is very well adapted to problems based on independent random variables, but in fusion processes this independence is seldom to occur. The limitation of Boolean logic in PT leads to Fuzzy Theory.

### 3.3 Fuzzy Set Theory

As we have noticed, DST and PT are able to describe epistemic and aleatoric uncertainty. However, both theories are based on crisp sets of a frame of discernment. In many measurement or sensor related concepts we have to handle both, noise and systematic effects. The latter ones are assumed to be *systematic errors* which can possibly be compensated, as they are treated like a known bias or offset. Unfortunately, in complex systems it is nearly impossible to obtain all systematic measurement data. Therefore, in many instances we have to handle these errors as systematic uncertainties. The normal situation is as follows: The systematic input is recognised, but the exact information (data, values) is unknown. However, in most cases it is possible to locate a data value in a closed measurement interval—an *interval of confidence*. A systematic contribution to a measurement always takes the same value, even if unknown, within a certain interval. That is, not each value has the same probability to occur. A certain unknown value will occur to 100 %, others will definitely not occur. However, values in the interval have, in absence of further evidence, the possibility to occur. The above mentioned idea is in contrast to Probability Theory.

This fact leads to Fuzzy Set Theory (Fuzzy Theory, FT). The concept of fuzzy sets was first introduced by Zadeh in 1965 (Zadeh, 1965) in order to represent incomplete knowledge.

However, in 1937 the mathematician and philosopher Max Black developed the concept of *vagueness* and *fuzzy membership* (Black, 1937).

The traditional set approach is based on *crisp sets*. That is, a crisp variable can only take the value 1 or can be interpreted to be *true*, if the variable belongs to the set, or 0 (*false*), if it does not belong to the set. Fuzzy sets, in contrast, are not based on a Boolean value representation, but on a *membership function*, which describes the membership of a variable to its appending fuzzy set in a closed interval  $[0, 1]$ . It represents a multi-valued logic. Fuzzy sets can thus be described as a generalisation of crisp sets. This is not an antagonism in respect of PT, but rather an addendum to it, since both theories address a different kind of uncertainty: *randomness* and *vagueness* (fuzziness).

Let  $\Theta$  be a frame of discernment, then a fuzzy set  $A$  of  $\Theta$  is defined by a membership function  $\mu_A : \Theta \rightarrow [0, 1]$ . The membership function  $\mu_A(\theta) \in \mathbb{R}$  is the degree of membership for a fixed element  $\theta$  in the fuzzy set  $A$ , that is, the degree of belonging  $x = \theta$  and  $x \in A$ , the degree of certainty of the proposition  $\theta \in A$ . The *normalised* membership function of a fuzzy set  $A$  satisfies the following axioms; cf. Eq. (27) - Eq. (29):

- Axiom 1:

$$\exists x_0 \in A : \mu_A(x_0) = 1. \quad (27)$$

- Axiom 2:

$$0 \leq \mu_A(x) \leq 1. \quad (28)$$

- Axiom 3:

$$\exists x, x_0 \in A \forall x \neq x_0 : \mu_A(x) \leq 1. \quad (29)$$

If  $A$  is a fuzzy set defined by its membership function  $\mu_A(x)$  and  $B$  is another fuzzy set defined by  $\mu_B(x)$ , then  $\mu_A(x) \leq \mu_B(x)$ , iff  $A$  is a subset of  $B$ :

$$A \subseteq B \Rightarrow \mu_A(x) \leq \mu_B(x). \quad (30)$$

The membership function  $\mu_A(x)$  of a fuzzy variable  $x$  can also be described in terms of  $\alpha$ -cuts—different vertical levels  $\alpha \in [0, 1]$ . Each  $\alpha$ -cut is defined as

$$A_\alpha = \{x \in A \mid \mu_A(x) \geq \alpha\}. \quad (31)$$

The representation of a fuzzy variable in terms of its  $\alpha$ -cuts describes a group of nested sets (consonant sets)  $A_\alpha$  in terms of *intervals of confidence* with its associated *levels of certainty*  $\alpha$  (cf. Fig. 3). The level of certainty describes how certain a measurement value is—this is the case iff a crisp value is taken into account. The level of confidence is mapped to a set and represents the available knowledge based on an unknown value in the range of the set  $A_\alpha$ .

Under the assumption of a finite support of  $A$ , we can assume that a membership function is fully represented by its  $\alpha$ -cuts. Furthermore, the  $\alpha$ -cuts of each fuzzy variable are closed intervals of real numbers  $\mathbb{R}$ .

The more the level of certainty increases, the more the interval of confidence decreases. This fact is in line with the human knowledge representation: If the level of certainty increases, the *level of confidence* must decrease. The level of certainty indicates how sure a measurement is, while the level of confidence specifies the probability that a certain value falls in its interval. In other words, as higher  $\alpha$  is, as lower the level of confidence must be. The level of confidence is associated to  $(1 - \alpha)$ . The above mentioned fact will be a subject-matter in Possibility Theory. In the framework of fuzzy sets, knowledge is represented by membership functions, which map a measurement value in a sense of a proposition to  $x = \theta \in A$  to  $\mu_A(x)$ , that is, the belief

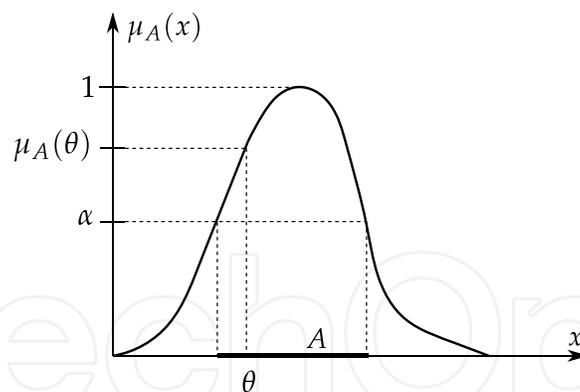


Fig. 3. Membership function and  $\alpha$ -cut.

in the proposition is expressed by  $\mu_A(x)$ . This fact corresponds to the concept of vagueness as a value  $\theta$  may belong to more than one membership function with different degrees of membership (e. g,  $\mu_B(\theta) < \mu_A(\theta)$  and  $A \cap B \neq \emptyset$ ), cf. Fig. 4).

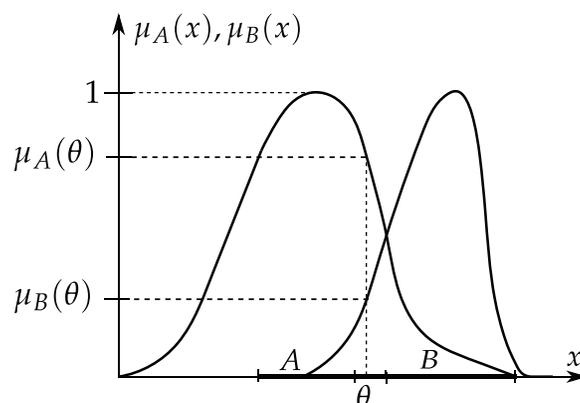


Fig. 4. Membership functions  $\mu_A(x)$  and  $\mu_B(x)$  with conjoint subset  $A \cap B \neq \emptyset$ .

Here, set  $A$  belongs to  $\mu_A(x)$  and set  $B$  belongs to  $\mu_B(x)$ . The value  $x = \theta$  is a member of both sets—a value which belongs to the subset  $A \cap B$ . As the degree of membership  $\mu_A(\theta)$  is higher than the degree of membership  $\mu_B(\theta)$ , the value  $\theta$  belongs more *certain* to set  $A$  than to set  $B$ . Fuzzy Set Theory can handle vagueness and *non-specificity* as it is an extension of classical set theory.

Total uncertainty (total ignorance) is expressed in terms of a rectangular shaped membership function with  $\mu_A(\theta_i) = 1 \forall \theta_i \in A$ .

Aggregation of information is based on elementary fuzzy set operations. The set operators aim for combining information in order to generate a single fuzzy variable. The basic operations are the equivalents to the above mentioned operations in classical set theory. Hence, we are able to define the complement, intersection and union of fuzzy variables. Following the base fuzzy set operations defined by Zadeh (Zadeh, 1965) the complement of a normalised membership value of the fuzzy set  $A$  is defined as:

$$\mu_{A^c}(x) = 1 - \mu_A(x). \quad (32)$$

The complement  $\mu_{A^c}(x)$  describes the degree to which  $x$  does not belong to the fuzzy set  $A$ . The *fuzzy union*  $\mu_{A \cup B}(x)$ , based on a s-norm,  $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ , is defined as:

$$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)). \quad (33)$$

Again, contrary to the probability theory, fuzzy union is *non-additive*. The *fuzzy intersection*  $\mu_{A \cap B}(x)$ , based on a t-norm,  $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ , is defined as:

$$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)). \quad (34)$$

The Eq. (33) and Eq. (34) are based on so-called triangular norms, *t-norms* and *t-conorms* (or *s-norms*). As many operators exist, we refer to the literature (Klir & Yuan, 1995).

The union and intersection can be interpreted as *aggregation operations*. In general, aggregation operations combine several fuzzy sets or membership functions resulting in a single membership function. The aggregation operator,  $h : [0, 1]^n \rightarrow [0, 1]$ , generates the output function by

$$\mu_{agg}(x) = h(\mu_A(x), \mu_B(x), \dots). \quad (35)$$

The aggregation operator  $h$  is a bounded, monotonic, increasing, symmetric, and idempotent operator (Klir & Yuan, 1995). Beside the fuzzy union and intersection, which can be defined in different ways, the so called *averaging operators* play an important role in aggregation. A mapping  $M : [0, 1] \times [0, 1] \rightarrow [0, 1]$  is an averaging operation iff it is continuous and satisfies the properties Symmetry:  $M(x_0, x_1) = M(x_1, x_0)$ ; Idempotency:  $M(x_0, x_0) = x_0$ ; Monotonicity:  $M(x_0, x_1) \leq M(x_2, x_3)$  if  $x_0 \leq x_2$  and  $x_1 \leq x_3$ ; Boundaries:  $M(0, 0) = 0$  and  $M(1, 1) = 1$ ;  $x_i \in [0, 1]$  (Dujmović & Larsen, 2007). Furthermore, the following inequality is valid for all averaging operators

$$\min(\mu_A(x), \mu_B(x)) \leq M(\mu_A(x), \mu_B(x)) \leq \max(\mu_A(x), \mu_B(x)). \quad (36)$$

By applying different averaging operators it is possible to construct a more “and-like” (min) or a more “or-like” (max) aggregation. Often the following operators are used: Harmonic mean, Geometric mean, and Arithmetic mean which are certain forms of the *generalised mean*. Another important basic operator covers the full range of Eq. (36). This class of operators is called *ordered weighted averaging operators* (OWA) (Dujmović, 1974), (Yager, 1988), and (Dujmović & Larsen, 2007). A *weighting vector*  $\vec{w} = (w_0, w_1, \dots, w_{N-1})$  with  $w_i \in [0, 1]$  with the constraint

$$\sum_{i=0}^{N-1} w_i = 1. \quad (37)$$

The OWA operator  $h_{OWA}$  is defined by first ordering the membership functions  $\mu_i(x)$  in *non-increasing order*, denoted with index ( $i$ ), for a given value  $x = \theta$ , based on their individual fuzzy sets  $A, B, \dots$ , and then applying

$$\mu_{agg}(\vec{w}, \theta) = \sum_{i=0}^{N-1} w_i \cdot \mu_{(i)}(\theta). \quad (38)$$

OWA aggregation can be freely positioned between the borders of Eq. (36). Dujmović (Dujmović, 1974) introduced a measure (degree of disjunction) that indicates how much an operator is near to the max-operation. Yager (Yager, 1988) called this operator *orness*. The operator models any degree of orness between 1 (pure OR) and 0 (pure AND) by adjustment of its corresponding parameters, the OWA weights after Eq. (37).

$$\text{orness}(h_{OWA}(\vec{w})) = \frac{1}{N-1} \sum_{i=0}^{N-1} (N-i) \cdot w_i. \quad (39)$$

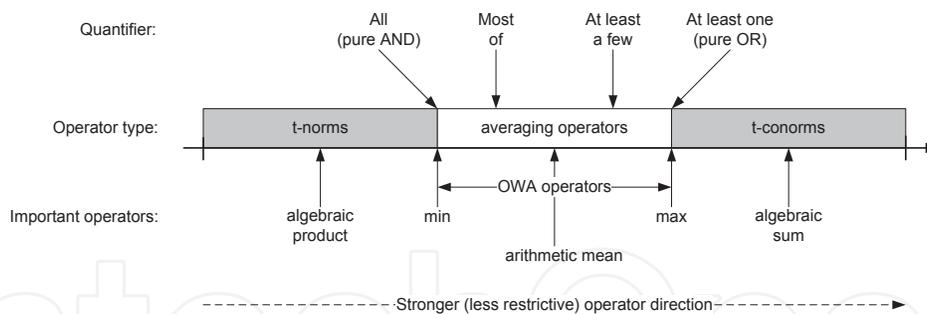


Fig. 5. Different rules of aggregation, with their associated operators (Larsen, 1999).

The *andness* degree is defined as:  $\text{andness}(h_{OWA}) = 1 - \text{orness}(h_{OWA})$ .

Fuzzy Set Theory operates with vague and non-specific information, based on a human-centred way of thinking. Usually, knowledge is represented by membership functions and their aggregation. In contrary to PT the propositions are multi-valued. The construction of membership function can be established by different concepts; e. g. cf. (Bocklisch & Priber, 1986), (Klir & Yuan, 1995), (Bossé et al., 2007), (Lohweg & Mönks, 2010a).

### 3.4 Possibility Theory

Possibility Theory (PosT) was introduced in 1978 by Zadeh (Zadeh, 1978). As imprecision and vagueness are rather possibilistic than probabilistic (frequentist's view) in its description (... *it is possible that we will have snow in summer, however, it is highly improbable...*), it is reasonable to describe the meaning of information, especially the meaning of *incomplete* information within a possibilistic framework. Possibility Theory reflects Probability Theory insofar, as it is also based on set-functions. Instead of one function, PosT is based on two functions, which are called *possibility measure* and *necessity measure*. We will concentrate on numerical possibility theory with its special branches based on consonant (nested) sets (Shafer, 1976) and on membership functions (Zadeh, 1978).

It has to be pointed out that PosT has been exhaustively studied by Dubois and Prade; e. g. (Dubois & Prade, 1978; 1983; 2000; 2003). Possibility Theory can be interpreted as a framework for handling incomplete information and aggregate information coming from multiple sources (sensors, experts, databases, etc.). If  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  denotes the so called *finite frame of discernment*,  $\theta_i, i \in \mathbb{N}$ , then a *possibility measure*  $\Pi$  is a mapping  $\Pi : 2^\Theta \rightarrow [0, 1]$ , which follows the axioms (cf. Eq. (40) - Eq. (42)):

- Axiom 1: 
$$\Pi(\emptyset) = 0. \quad (40)$$

- Axiom 2: 
$$\Pi(\Theta) = 1. \quad (41)$$

- Axiom 3: 
$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \quad (42)$$

The measure  $\Pi(A)$  denotes the possibility that an element  $\theta$  belongs to the *crisp* set  $A$ . As it can be noted from Eq. (42) the union of two sets is *non-additive*. The *necessity measure*  $N(A)$  is the dual set-function of  $\Pi(A)$ :

$$N(A) = 1 - \Pi(A^c). \quad (43)$$

The intersection of two sets (Eq. (44)) is determined as follows:

$$N(A \cap B) = \min(N(A), N(B)). \quad (44)$$

Interpreting Eq. (43), the necessity measure describes the fact that  $N(A)$  is entirely true, iff the “complement of  $A$ ” is impossible. The following implications hold for possibility and necessity measures:

$$N(A) > 0 \Rightarrow \Pi(A) = 1,$$

$$\Pi(A) < 1 \Rightarrow N(A) = 0.$$

A fuzzy variable can be interpreted as a representation of an element of a fuzzy set with a graded membership within its set. The membership function can also be interpreted differently: Let  $\mu(\theta)$  be a *possibility level* for which  $\theta \in A$  represents an existing, but unknown element (incomplete information) within in the crisp set  $A$ . We describe an unknown element (number, measurement value, etc.) with a certain level of certainty ( $\mu(x) \mid x = \theta$ ). In this case the normalised fuzzy set is called *possibility distribution*  $\pi_x(\theta)$ , where  $x$  is a fuzzy variable of  $\Theta$ , and  $\theta$  or  $\theta_i$  are appointed values of  $\Theta$  that  $x$  can take. A possibility distribution is a mapping  $\pi : \Theta \rightarrow [0, 1]$  with at least one  $\pi(\theta_i) = 1$ ,  $\theta_i \in \Theta$ .

Possibility and necessity measures are represented by the associated possibility distribution ( $\theta \in A$ ), that is,

$$\Pi(A) = \max(\pi_x(\theta)), \quad (45)$$

$$N(A) = \min(1 - \pi_x(\theta)). \quad (46)$$

If  $\Theta$  is infinite, then  $\Pi(A) = \sup_{\theta \in A} (\pi_x(A))$  and  $N(A) = \inf_{\theta \in A} (1 - \pi_x(A))$ .

Now, we will give a short insight in the interconnection between PosT and DST in the case that the focal elements within particular sets are nested, that is, the sets are ordered in a way that each is contained in the next one. Belief and Plausibility functions are termed to be *consonant* (Shafer, 1976). Let  $\Theta$  be the frame of discernment with its subsets  $A_i = \{\theta_0, \theta_1, \dots, \theta_{i-1}\}$  for  $i = 0, 1, \dots, N - 1$ . The subsets  $A_i$  match the relationship:

$$\Theta \equiv A_0 \subset A_1 \subset A_2 \subset \dots \subset A_{N-1} \quad (47)$$

In the case of nested sets Belief and Plausibility satisfy the following relationship with  $A, B \in \mathcal{P}(\Theta)$ , cf. Eq. (42), Eq. (44):

$$\text{Pl}(A \cup B) = \max(\text{Pl}(A), \text{Pl}(B)), \text{Bel}(A \cap B) = \min(\text{Bel}(A), \text{Bel}(B)). \quad (48)$$

When Plausibility and Belief functions satisfy Eq. (48), they are called Possibility and Necessity functions. Possibility Theory is defined axiomatically via Eq. (40) - Eq. (42) as an independent theory, though, as a second branch it can be defined by the possibility distribution as a special form of the membership function (Zadeh, 1978).

$$\pi_x(\theta) = \mu_x(\theta). \quad (49)$$

However, it has to be pointed out that the membership function may not be readily interpreted as a possibility distribution function, because the meaning of both functions is completely different. Whereas the membership function describes a fuzzy variable  $\theta \in A$ , the possibility distribution function characterises a representation of what is known (expert, sensor measurement, etc.) about the value of some quantity  $\theta$  ranging on  $A \subseteq \Theta$  (not necessarily a random quantity). The function  $\pi_x(\theta)$  reflects the more or less plausible values of the unknown quantity roughly. We tend to another notation proposed by Dubois and Prade to circumvent any ambiguity:

$$\pi_x(\theta \mid A) = \mu_x(A \mid \theta). \quad (50)$$

The possibility that  $x = \theta$ , knowing that  $\theta \in A$  is  $\pi_x(\theta | A)$ , whereas  $\mu_x(A | \theta)$  is the level of possibility for  $\theta \in A$ . An interesting interrelationship exists between  $\alpha$ -cuts (cf. Eq. (31)), basic belief assignments  $m$ , possibility distributions  $\pi_x(\theta | A)$ , and membership functions  $\mu_x(A | \theta)$  iff the sets are nested (cf. Eq. (47)). Let  $\alpha_i$  be a certain value of  $\mu_x(A_{\alpha_i} | \theta)$ , then  $A_{\alpha_i}$  denotes its corresponding  $\alpha$ -cut. The associated possibility function is then defined by (Klir & Yuan, 1995):

$$\pi_x(\theta) = \sum_{\substack{i=0, \\ \theta \in A_{\alpha_i}}}^{N-1} m(A_{\alpha_i}), \quad (51)$$

under the condition, that

$$m(A_{\alpha_i}) = \alpha_i - \alpha_{i+1}$$

and  $i = 0, 1, \dots, N - 1$ , and  $\alpha_N = 0$ . Furthermore, keeping in mind, the higher  $\alpha_i$ , the lower the possibility is that an element  $\theta$  will fall in its associated  $\alpha$ -cut, that is,  $\Pi(A_{\alpha_i}^c) = \alpha_i$ . Therefore, with Eq. (43),

$$N(A_{\alpha_i}) = 1 - \alpha_i$$

— *the level of confidence*. For details refer to Salicone (Salicone, 2007).

Possibility Theory is well suited for sensor and information fusion. Let  $\pi_A(x)$  and  $\pi_B(x)$  be two possibility distributions from two different signal sources, then, proposed by Dubois and Prade (Dubois & Prade, 1994), the sources are fused by:

$$\pi_{AB}(x) = \max \left[ \frac{t(\pi_A(x), \pi_B(x))}{1 - k_c(\pi_A(x), \pi_B(x))}, \min[1 - k_c(\pi_A(x), \pi_B(x)), s(\pi_A(x), \pi_B(x))] \right]. \quad (52)$$

The functions  $t(\cdot)$  and  $s(\cdot)$  are t-norms and s-norms, where  $K(\cdot)$  is the conflicting factor known from Dempster's rule of combination (cf. Eq. (13)). The fusion concept takes into account the reliability of different sources. If both sources are reliable an "and-like" fusion is applied; otherwise, an "or-like" fusion is executed.

Possibility Theory provides a promising framework for sensor and information fusion based on incomplete or partly unknown data. It is able to handle uncertainty, ignorance, and vagueness based on crisp sets instead of fuzzy sets. However, being a distinct theory, its branches interconnect Dempster-Shafer Theory, Fuzzy Set Theory, and Probability Theory. Insofar PosT can be interpreted as an unifying Theory.

#### 4. Human behaviour based sensor fusion

Human beings have remarkable potential for reasoning and rational decision-making, even when confronted with such uncertain data. Therefore, it is highly desirable to design technical systems in a way which is inspired by human behaviour. Zadeh called this human level machine intelligence (HLMI), cf. (Zadeh, 2008). Although HLMI has been intensively investigated under the umbrella of artificial intelligence (AI) and its subdivision multi-source data fusion for the past fifty years, it is still a hot research topic (Zadeh, 2008).

#### 4.1 Two-layer Conflict Solving Fusion

Dempster-Shafer based fusion creates counter-intuitive results in high conflict situations. If two sensors, that are measuring the same effect, deliver different information, then a conflict situation occurs. This inherent defect pointed out by Zadeh (Zadeh, 1986), brings criticism to the DST. Therefore, many other alternatives were researched. For instance, we refer to Campos' rule (Campos, 2006) and the work of Dezert and Smarandache (DSmT) (Smarandache & Dezert, 2006). No matter whether the original DST rule or other ad-hoc rules are applied, none of them have been regarded as a superior method compared to others. Therefore, a Two-Layer Conflict Solving (TLCS) data fusion approach is suggested, which includes two layers to combine pieces of evidence. The conflict is solved in some degree during combination—therefore, it is named as conflict solving. The first layer resolves the conflict in some extent, and the second continues to solve it and achieves more stable results. Psychologically, as clearly stated in (Lipshitz et al., 2001), *“Decision making has been traditionally studied at three levels: individual, group and organizational.”*, also cf. (Sunita, 1999). This shows that decision is made at three layers, in which conflict is unavoidable to be considered and solved: The individual level is the basic element that holds conflict; group level has a larger range which includes conflict while organisational level is the largest. The latter can be instantiated by connecting several layer 2 outputs via TLCS to generate a higher hierarchical level—the organisational one. In such a way, humans believe that conflict can be solved in an straightforward way, although it is impossible to totally eliminate its negative impacts. The approach is also applicable if several groups of sensors are considered in a larger system. Fig. 6 depicts the scheme of TLCS. In Fig. 6, layer 1 is regarded as working at the individual

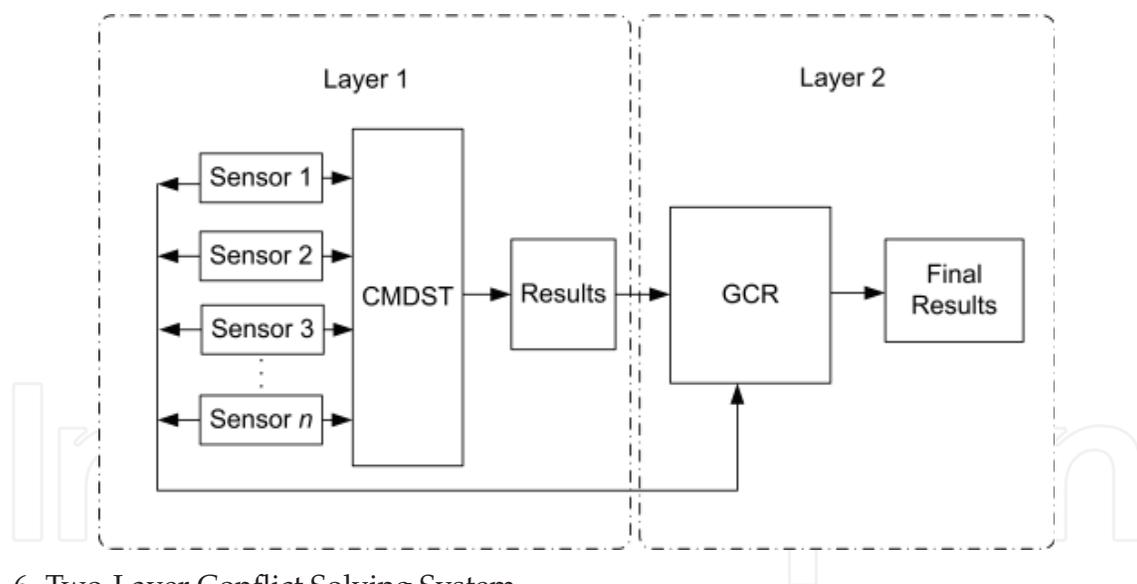


Fig. 6. Two-Layer Conflict Solving System.

level because the Conflict-Modified DST (CMDST) is an approach which combines every two sensors' data so that a conflict is considered and solved between two individuals. After receiving the results from the previous layer, layer 2 collects all sensors' original information and fuses it with the combined results from the CMDST. The conflict is further resolved at a group level.

##### 4.1.1 Aggregation rules

Especially due to the inherent defect residing in the DST (Zadeh, 1986), (Campos, 2006), many other authors have proposed their own data fusion approaches which serve as ad-hoc

alternatives. Just to name only a few, e. g., Murphy's rule (Murphy, 2000)

$$m_M = \frac{m_1(A) + m_2(A)}{2}, \quad (53)$$

is a trade-off rule, which takes the arithmetic average value of two masses  $m_1(A)$  and  $m_2(A)$ . Yager's rule (Murphy, 2000) regards that the universal set  $\Theta$  should include the mass from the conflicting parts, so that the universal set (set with all propositions) is always introduced in Yager's rule:

$$m_Y(C) = \sum_{A,B \in 2^\Theta, A \cap B = C} m_1(A)m_2(B), \quad (54)$$

$$m_Y(\Theta) = m_1(\Theta)m_2(\Theta) + \sum_{A,B \in 2^\Theta, A \cap B = \emptyset} m_1(A)m_2(B). \quad (55)$$

Campos' rule

$$m_C(A_i) = \frac{DST}{1 + \log \left( \frac{1}{1 - \sum_{A_1 \cap \dots \cap A_n = \emptyset} \prod_{i=1}^n m_i(A_i)} \right)}, \quad (56)$$

is explained in (Campos, 2006). Dezert-Smarandache Theory (DSmT) is rather a comprehensive theory; we refer to (Smarandache & Dezert, 2006).

#### 4.1.2 Conflict-modified DST

Based on the idea of DST, CMDST calculates the conflict in a different manner as shown in the formula:

$$k_{cm} = \sum_{\substack{A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, \dots, \\ A_i \cap A_k = \emptyset, \dots, A_{n-1} \cap A_n = \emptyset}} \prod_{i=1}^n m_i(A_k). \quad (57)$$

Within this definition, conflicts are calculated between every two sensors instead of all the sensors together (which is used in DST). This difference can be seen from the condition of summation in Eq. (13),  $A_1 \cap \dots \cap A_n = \emptyset$  and  $A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, \dots, A_1 \cap A_n = \emptyset, \dots, A_{n-1} \cap A_n = \emptyset$ . Due to the specified way of determining conflicts in Eq. (57), this kind of conflict will very likely be larger than one, whereas in DST the denominator is  $1 - k_c$ . Therefore, the denominator in DST must also be modified. First of all, the denominator is modified as:

$$\binom{n}{2} - k_{cm} = \binom{n}{2} - \sum_{\substack{A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, \dots, \\ A_1 \cap A_n = \emptyset, \dots, A_{n-1} \cap A_n = \emptyset}} \prod_{i=1}^n m_i(A_k). \quad (58)$$

The reason for choosing  $\binom{n}{2} = B_c(n)$  (binomial coefficient) is that there are  $B_c(n)$  possible combinations for calculating conflicts ( $n$  is the number of sensors). Thus,  $K_{cm}$  ( $K$  in DST) is:

$$K_{cm} = \frac{1}{B_c(n) - \sum_{\substack{A_1 \cap A_2 = \emptyset, A_1 \cap A_3 = \emptyset, \dots, \\ A_i \cap A_k = \emptyset, \dots, A_{n-1} \cap A_n = \emptyset}} \prod_{i=1}^n m_i(A_k)} \quad (59)$$

with  $\frac{1}{B_c} \leq K_{cm} \leq \infty$ . The CMDST is formed as:

$$\text{CMDST}(A) = \bigoplus_{i=1}^n m_i(A) = K_{cm} \sum_{\substack{A_1 \cap A_2 = A, \dots \\ A_1 \cap A_3 = A, \dots \\ A_{n-1} \cap A_n = A}} \prod_{i=1}^n m_i(A). \tag{60}$$

**4.1.3 Group Conflict Redistribution**

As pointed out in (Sunita, 1999), decision making is also studied in a group level. Hence, Group Conflict Redistribution (GCR, layer 2 in Fig. 6) acts as group conflict solving strategy, solving a conflict in a larger extent compared to individual level (CMDST). Distinguishing from layer 1 (CMDST), GCR combines sensors’ propositions in a group manner, which means all sensors will participate in this procedure (Li & Lohweg, 2008):

$$m(a) = \frac{\sum_{A_1 \cap \dots \cap A_n = A} m_i(A) + (B_c(n) + |\log(B_c(n) - k_{cm})|) \cdot \text{CMDST}(A)}{n + B_c(n) + |\log(B_c(n) - k_{cm})|}. \tag{61}$$

The variable  $n$  defines the number of sensors and  $\log(x)$  is logarithm to the base 10. The denominator includes the number of sensors and how many combinations  $B_c(n)$  among sensors are possible as well as conflict evaluation term  $|\log(B_c(n) - k_{cm})|$ . In the numerator part, sensors’ original propositions  $\sum_{A_1 \cap \dots \cap A_n = A} m_i(A)$  are calculated with corresponding CMDST results, which are obtained from layer 1. Finally, the sum of final fused results remains ‘1’. Numerical examples on TLCS which describe the aggregation effect can be found in (Li & Lohweg, 2008).

**4.1.4 Balanced TLCS**

The intention to optimise the TLCS concept is threefold. Balanced TLCS (*BalTLCS*) intends to cover three topics: First, the general TLCS approach should be utilised again. Second, the result for two information sources (sensors, cognitive knowledge) should generate intuitive results. Third, the conflict redistribution scheme for data fusion should be adapted and simplified in a way that the fusion process i) should be an additive function of all inputs and ii) could be extended to multiple sources and focal elements (masses). The approach is based on the normalised version of CMDST. The normalisation is based on the Eq. (57) - Eq. (60). The normalised conflicting factor  ${}^N K_{cm}$  is defined as follows:

$${}^N K_{cm} = \frac{B_c(n)}{B_c(n) - k_{cm} + \epsilon} < \infty \tag{62}$$

where  $\forall \epsilon > 0, \forall n > 1, n \in \mathbb{N}, k_{cm} = 0 \exists L \in \mathbb{R} : L = (1 - {}^N K_{cm}) < \epsilon$ . Therefore, as  $0 \leq k_{cm} \leq B_c(n)$ , the normalised conflict factor ranges between

$$\frac{B_c(n)}{B_c(n) + \epsilon} \leq {}^N K_{cm} \leq B_c(n) \cdot \epsilon^{-1} < \infty. \tag{63}$$

The non-conflicting part is determined and coupled with the conflicting part according to Eq. (60) as

$${}^N \text{CMDST}(A) = \bigoplus_{i=1}^n m_i(A) = \frac{{}^N K_{cm}}{B_c(n)} \sum_{\substack{A_1 \cap A_2 = A, \dots \\ A_1 \cap A_3 = A, \dots \\ A_{n-1} \cap A_n = A}} \prod_{i=1}^n m_i(A). \tag{64}$$

The *BalGCR* (Balanced Group Conflict Redistribution) approach combines the sensors' propositions in a group manner which implies that in this case all sensors will participate additively in this procedure.

The intention is to utilise the inverse of the normalised conflict factor as a control parameter. If no conflict has occurred, mainly the <sup>N</sup>CMDST fusion result should contribute to the overall result. If the conflict is high, then all different information sources have to be taken into account. None of the information sources is allowed to dominate the other ones, and none of them is allowed to play a part in the overall result with more than  $\frac{1}{n}$ .

Furthermore, in a strong conflict case it is not intended to shift the information content (hypotheses) to the universal set  $\Theta$ , defining ambiguity or ignorance, because a conflict has to be solved in any case. As the set must be complete regarding the sources, all information sources (hypotheses) which can appear must be covered in a set. The only situation where hypotheses can be transferred to the universal set  $\Theta$  occurs when a source delivers no reliable data. Formally *BalGCR* fusion consists of two parts. The first one describes the non-conflicting part:

$$m_{nc} = {}^N K_{cm}^{-1} \cdot \text{CMDST}(A) = \frac{1}{B_c(n)} \sum_{\substack{A_1 \cap A_2 = A, \dots \\ A_1 \cap A_3 = A, \dots \\ A_{n-1} \cap A_n = A}} \prod_{i=1}^n m_i(A). \quad (65)$$

The second part characterises the conflicting part. Eq. (65) tends to zero in heavy conflicts ( $k_{cm} \rightarrow B_c(n)$ ). Therefore, it is proposed here that all masses supporting a certain hypothesis have to be averaged to fulfil the above mentioned statements. Furthermore, the average value has to be controlled, e. g., by the normalised conflicting factor or coefficient. It is proposed to use

$$m_c(A) = \frac{k_{cm}}{B_c(n)} \cdot \frac{1}{n} \sum_{i=1}^n m_i(A) \quad (66)$$

for the conflicting part. It can be recognised that, in the case of maximum conflict, the average value of all sensory hypotheses is determined. In the case of minimum conflict  $m_c(A) \rightarrow 0$  both parts (Eq. (65) and Eq. (66)) are additively connected.

$$m(A) = m_c(A) + m_{nc}(A) = \frac{k_{cm}}{B_c(n)} \cdot \frac{1}{n} \sum_{i=1}^n m_i(A) + \frac{1}{B_c(n)} \sum_{\substack{A_1 \cap A_2 = A, \dots \\ A_1 \cap A_3 = A, \dots \\ A_{n-1} \cap A_n = A}} \prod_{i=1}^n m_i(A). \quad (67)$$

This equation describes the data fusion approach by the balanced TLCS concept. All masses are fused regarding their pro-and-con-hypotheses. Finally, the sum of final fused results remains always '1', if all sensors deliver acceptable results in the sense of an adequately functioning sensor. As the fusion process is based on averaging, the final result is relatively insensitive against noise effects. The fused mass of the universal set is determined for  $n$  sensors as follows:

$$m(\Theta) = 1 - [m(A) + m(B) + m(C) + \dots]. \quad (68)$$

The masses take into account all information from the hypotheses. For further details and results we refer to (Li & Lohweg, 2008) and (Lohweg & Mönks, 2010b).

## 4.2 Majority-Observation-Guided Possibilistic Fusion

One promising mathematical theory that appears to be capable of (partially) modelling a human-like decision process is *Possibility Theory*. Although Possibility Theory was introduced in 1978 by Zadeh (Zadeh, 1978) and has been primarily advanced by Dubois and Yager over the past few decades (Dubois & Prade, 2000; Yager, 2004), it is still in its infancy. This section is focused on possibilistic multi-sensor systems for industrial automation. With reference to the Sect. 4.1, *Dempster-Shafer Theory*-based fusion approaches were recently applied to this problem statement and reached satisfactory results (Lohweg & Mönks, 2010b). On the contrary, possibilistic approaches are rarely applied to those industrial-automation processes that can be classified under the problem theme of context-based anticipatoric machine conditioning. As Possibility Theory appears to be crucial for modelling the human behaviour when confronted with vague data (Zadeh, 2008), this section presents a novel possibilistic multi-source data fusion approach, the so-called *Majority-Observation-Guided Possibilistic Fusion Rule (MOGPFR)*. The MOGPFR is capable of handling the imprecision of sensors. Moreover, it has the ability of merging imprecise, incomplete or inconsistent data in a human-like group-decision process.

The rest of this section is organised as follows: Section 4.2.1 provides the basics needed for the understanding of the framework. Sections 4.2.2– 4.2.6 introduce the novel possibilistic fusion rule.

### 4.2.1 Fusion operators

As pointed out in Sect. 3.4, conjunctive aggregation of possibility distributions is debatable when not all sources are completely reliable. In this context, when all sources are completely unreliable, the disjunctive rule is most appropriate. However, when confronted with a group of reliable and unreliable sources the class of averaging operators, which lies between the conjunctive and the disjunctive rule (cf. Fig. 5), might be beneficial. OWA operators consider all fuzzy sets equally important. However, in multi-criteria decision problems, such as data fusion, different criteria, represented by different fuzzy sets, might have different importance. Hence, the effect on the overall outcome of satisfying a criterion should decrease as the importance of satisfying the criterion decreases. To meet this desire, the concept of OWA aggregation is expanded by attaching a so-called *importance weight*,  $w_i$ , to each criteria,  $a_i$  to the importance weighting vector  $\vec{v}$  of a  $n$ -dimensional attribute  $A$ . This leads to the class of *Importance Weighted Ordered Weighted Averaging (IWOWA)* operators as defined below:

$$\lambda_{IWOWA}(\vec{v}, \vec{w}, \vec{a}) = \sum_{i=1}^n w_i b_{(i)}, \quad (69)$$

where  $i$  is a permutation on  $\{1, \dots, i, \dots, n\}$ , such that  $b_{(1)} \geq \dots \geq b_{(i)} \geq \dots \geq b_{(n)}$  and  $b_i = h_{OWA}^\alpha(v_i, a_i)$  (Larsen, 1999). Larsen has shown (Larsen, 1999) that the class of IWOWA operators is order equivalent to the *Weighted Arithmetic Mean (WAM)* (Yager, 1994). Order-equivalence is sufficient when the operator is applied to provide preference ordering (Larsen, 2002). However, in situations where the aggregated value is used for other purposes, such as multi-source data fusion, a full equivalence to the WAM is needed. This property can be obtained by normalizing Eq. (69) in the interval of  $\lambda_{IWOWA}(\vec{v}, \vec{w}, 0)$  and  $\lambda_{IWOWA}(\vec{v}, \vec{w}, 1)$ , which leads to the class of *Implicative Importance Weighted Ordered Weighted Averaging (IIWOWA)* operators as defined as follows (Larsen, 2002):

$$\lambda_{IIWOWA}(\vec{v}, \vec{w}, \vec{a}) = \frac{\lambda_{IWOWA}(\vec{v}, \vec{w}, \vec{a}) - \lambda_{IWOWA}(\vec{v}, \vec{w}, 0)}{\lambda_{IWOWA}(\vec{v}, \vec{w}, 1) - \lambda_{IWOWA}(\vec{v}, \vec{w}, 0)}. \quad (70)$$

#### 4.2.2 Possibilistic fusion frameworks

Various multi-source data-fusion frameworks based on the Possibility Theory have been proposed in the past thirty years (Yager, 1979; 1994). These frameworks were mainly influenced and developed by Dubois and Yager (Dubois et al., 2001). Therefore, this section introduces the two possibilistic fusion frameworks previously mentioned.

Dubois' adaptive fusion is grounded on the discussions on the usefulness of the *maximum specificity principle*. It is based on the supposition that  $j \leq n$  sources of a  $n$ -dimensional fusion problem are reliable. First, the  $j$  reliable sources are combined conjunctively. Secondly, the intermediate result of the combination of the  $j$  reliable sources is combined disjunctively with the  $(n - j)$  unreliable sources. However, in real-life, a precise number of reliable sources is sometimes unknown or is complex to evaluate. For this reason, a good alternative is an *optimistic* estimation,  $j_o$ , and a *pessimistic* estimation,  $j_p$ , of the number of reliable sources using the following derivation (Dubois et al., 2004): Let  $T \subseteq S$  be a subset of sources,  $S = \{S_1, \dots, S_i, \dots, S_n\}$  be the set of sources to be combined, and  $h$  be the consistency index

$$h(\pi_v^1(x), \pi_v^2(x)) = \sup_{x \in X} [\min[\pi_v^1(x), \pi_v^2(x)]] \quad (71)$$

of the subset  $T$ . Then,  $j_o$  and  $j_p$  are determined as stated below:

$$T_p = \sup[T|h(T) = 1], \quad (72)$$

$$T_o = \sup[T|h(T) > 0], \quad (73)$$

$$j_p = |T_p|, \quad (74)$$

$$j_o = |T_o|, \quad (75)$$

where Eq. (72) describes the biggest subset in  $S$  that satisfies  $h(T) = 1$ , that is, it is completely consistent; and Eq. (73) describes the biggest subset in  $S$  that satisfies  $h(T) > 0$ , that is, it is not completely inconsistent. With reference to Eq. (72) and Eq. (74), it is feasible that at least  $j_p$  sources are reliable, because  $j_p$  represents the number of sources with full agreement. With this in mind, the numerical adaptive combination rule can be stated as follows (cf. Eq. (52)):

$$\pi_v^{1, \dots, n}(x) = \max \left[ \frac{\pi_v^{j_o}}{h(T_o)}, \min \left[ \pi_v^{j_p}, 1 - h(T_o) \right] \right]. \quad (76)$$

Yager's fusion approach (Yager, 1979) is based on Zadeh's extension principle (Zadeh, 1978) and is defined as follows:

$$\pi_v^{1, \dots, n}(x) = \left\{ \frac{\min [\pi_v^i(x_i)]}{F(x_1, x_2, \dots, x_n)} \right\}, \quad (77)$$

where  $x_i \in X$  and  $F(x)$  is an averaging operator. One problem that arises in blindly using the extension principle is that the aggregated possibility distribution may not be satisfactory because it does not reflect any of the source's opinion. For example, if there are two estimates of a person's age,  $p_1 = 30$  and  $p_2 = 50$ , then the average is 40. However, this is a value that is not very compatible with either of the estimates. The reason for this lack of compatibility is the attempt to aggregate disparate values. To solve this problem, Yager introduced an "intelligent" component, the so-called *compatibility* relationship,  $R$ , in the merging process to address

conflicts in the data to be fused. In this context, the compatibility relationship is defined by expert knowledge. Fusion by considering compatibility is defined as follows (Yager, 1994):

$$\pi_v^{1,\dots,n}(x) = \left\{ \frac{\min [R(x_1, x_2, \dots, x_n), \min [\pi_v^i(x_i)]]}{F(x_1, x_2, \dots, x_n)} \right\}, \quad (78)$$

where  $x_i \in X$ ,  $R(x_1, x_2, \dots, x_n) = \min_{i,j} [R(x_i, x_j)]$ , and  $F$  is an averaging operator. As observed in Eq. (78), the merging of incompatible data is indicated by a low possibility degree.

#### 4.2.3 Fusion rule

The possibilistic frameworks introduced in Sect. 4.2.2 offer a promising setting for merging observations from different sensors. However, both approaches do not consider sensor's reliability. In case of highly unreliable or defective sensors, this might affect the fusion result negatively, which makes both approaches less suited to real-life machine conditioning applications. To overcome the drawbacks mentioned before, this section presents a novel possibilistic data-fusion approach, which is based on Yager's and Dubois' frameworks for data fusion (Dubois et al., 2001; Yager, 2004) and the *theory of collective group decisions* (Eisenführ, 2003). It is an extended close-to-practice approach, which has the ability of considering sensor reliability. Before defining the fusion rule in detail, the hypotheses based on Yager's and Dubois's concepts have been widened are discussed in the following paragraphs.

- H1. In real life, group decision-making by humans is mainly driven by the concept of majority opinion. This is derived from the fact that unanimous consensus rarely occurs during group decisions. Hence, it is natural to provide a fusion approach for the ability of majority-guided aggregation.
- H2. During group decision-making processes, decision-makers might have different priorities or reliabilities. It is natural that group decisions are mainly based on the voting of prioritised and reliable decision-makers.

Undoubtedly, the desired characteristics are intrinsically vague. Accordingly, different fuzzy approaches assessing majority-guided decisions are ideally suited. In this context, the class of IIWOWA operators, introduced in Sect. 4.2.1, is relevant. However, the IIWOWA operator cannot be applied directly to sensors' observations, because it is explicitly defined for inputs in the range of  $[0, 1]$ . Therefore, the sensor observations are fuzzified to make the data measured applicable to IIWOWA operators. The fuzzification of the sensor observations is accomplished using the *Modified-Fuzzy-Pattern-Classifier* (Lohweg & Mönks, 2010a). In addition to the necessity of fuzzification because of application of an IIWOWA operator, the fuzzification of the measurement scale has an important practical benefit: the fuzzification process facilitates the usage of different sensors that measure different physical quantities on the same attribute. Considering the preceding discussion, the so-called *Majority-Opinion-Guided Possibilistic Fusion Rule (MOGPFR)* is defined as follows:

$$\pi_v^{1,\dots,n}(x) = \max_i [\rho^i] \cdot \hat{\pi}_v^{1,\dots,n}(x) + 1 - \max_i [\rho^i], \quad (79)$$

where

$$\hat{\pi}_v^{1,\dots,n}(x) = \left\{ \frac{\lambda_{\text{IIWOWA}}(\vec{v}, \vec{w}_p, \vec{\pi}_v)}{\lambda_{\text{IIWOWA}}(\vec{v}, \vec{w}_m, \vec{\mu})} \right\}, \quad (80)$$

$$\vec{\pi}_v = \left( \pi_v^1(x_1), \dots, \pi_v^i(x_i), \dots, \pi_v^n(x_n) \right),$$

$$\vec{\mu} = (\mu_1(x_1), \dots, \mu_i(x_i), \dots, \mu_n(x_n)).$$

The variable  $\mu_i(x_i)$  represents the fuzzified sensor observation  $x_i$  of the  $i$ th sensor  $S_i$ . Furthermore,  $\pi_v^i(x_i)$  represents the corresponding possibility distribution. The characteristics of IIWOWA operators can be flexibly adjusted by modifying the importance weighting vector  $\vec{v}$  and the weighting vectors  $\vec{w}_p$  and  $\vec{w}_m$  (Larsen, 1999).

The following paragraphs describe the methods for a problem-dependent evaluation of these parameters, so that the stated hypotheses are satisfied:

Consider a  $n$ -dimensional fusion problem of a set of  $n$  sensors,  $S = \{S_1, \dots, S_i, \dots, S_n\}$ , with their associated reliabilities,  $\{\rho^1, \dots, \rho^i, \dots, \rho^n\}$ . Then, the following properties, which have to be fulfilled by the MOGPFR, can be derived from H1 and H2:

**Lemma 1.** *If at least one sensor is completely reliable, the MOGPFR combines the possibility distributions conjunctively, whereas when all sources are unreliable, the MOGPFR combines the possibility distributions disjunctively (cf. Sect. 4.2.2).*

**Lemma 2.** *The fusion approach of the MOGPFR proceeds progressively from the conjunctive to the disjunctive mode, when the highest reliability among all sources decreases from one to zero (cf. Sect. 4.2.2).*

The function defined by Eq. (79) satisfies Lemma 1 and 2 iff the andness-degree of the IIWOWA operator of Eq. (80) used for fusing the possibility distributions is determined as follows:

$$\alpha_p = \min_{i \in \{1, \dots, n\}} [\rho^i]. \quad (81)$$

*Proof.* The satisfaction of Lemma 1 and 2 follows directly from the definition of the OWA and IIWOWA operators (Larsen, 1999; Yager, 1994). □

**Lemma 3.** *If only one source is fully reliable and all other sources are totally unreliable, the fusion result of the MOGPFR models the possibility distribution provided by the reliable source (cf. H2).*

The function defined by Eq. (79) satisfies Lemma 3 iff the importance weighting vector  $\vec{v}$  of Eq. (80) is defined as stated below:

$$\vec{v} = (\rho^1, \dots, \rho^i, \dots, \rho^n). \quad (82)$$

*Proof.* Consider that  $S_1$  is totally reliable and all other sources,  $\{S_2, \dots, S_i, \dots, S_n\}$ , are completely unreliable. Hence, with reference to Eq. (82), the importance weighting vector accounts to  $\vec{v} = (1, 0, \dots, 0)$  and Eq. (79) results in

$$\pi_v^{1, \dots, n}(x) = \hat{\pi}_v^{1, \dots, n}(x), \forall x \in X.$$

Consider the definition of IIWOWA operators (Larsen, 1999), the importance weighting function mutates to

$$h(0, \pi_v^i) = I \quad \forall i \neq 1,$$

$$h(1, \pi_v^1) = \pi_v^1.$$

Therefore, the importance weighting function models for all sources except  $S_1$  the identity element  $I$  and hence,  $\pi_v^{1, \dots, n}(x) = \hat{\pi}_v^{1, \dots, n}(x) = \pi_v^1(x) \quad \forall x \in X$ . □

**Lemma 4.** *When all sources are completely unreliable, then the MOGPFR models total ignorance (cf. Sect. 4.2.2).*

The function defined by Eq. (79) satisfies Lemma 4.

*Proof.* Consider that all sources are totally unreliable. Hence, with reference to Eq. (82), the importance weighting vector accounts to  $\vec{\sigma} = (0, \dots, 0)$  and Eq. (79) can be written as follows:

$$\pi_v^{1, \dots, n}(x) = 1, \forall x \in X.$$

Thus, Eq. (79) models total ignorance.  $\square$

**Lemma 5.** *If there is a conflict between sources, it is solved by the MOGPFR so that the majority observation is prioritised (cf. H1).*

To meet the requirement postulated in Lemma 5, the majority observation of the sensors has to be evaluated. The evaluation of the majority observation is discussed in Sect. 4.2.4.

#### 4.2.4 Data consistency

The evaluation of the majority observation, necessity to satisfy Lemma 5, arises the need of a *consistency* measurement amongst sensors' observations, which is introduced in the following: Consider a set of sensors  $S = \{S_1, \dots, S_i, \dots, S_n\}$  with their corresponding reliabilities  $\{\rho^1, \dots, \rho^i, \dots, \rho^n\}$ . Let each sensor provide an observation through a possibility distribution  $\pi_v^i, i \in \{1, \dots, n\}$ . Then, the largest set of sensors  $T_c$ , which forms a consensus observation is defined as follows:

$$T_c = \sup [T | h(T) > 0], \quad (83)$$

where  $T \subseteq S$ ,  $h(T)$  is the consistency index as defined in Eq. (71).

Naturally, observations of sensors which are elements of  $T_c$  are completely consistent with the consensus observation and therefore, their consistency should be equal to one. In the same way, the consistency of an observation should decrease when the distance to the consensus observation increases. Hence, the measurement of consistency rises the need for evaluating the range of the consensus observation and an appropriate definition for measuring the distance amongst observations.

To simplify the measure of consistency, the possibility distribution (observation) of each sensor is defuzzified using its *centre of gravity*, denoted as  $C$  (Klir & Yuan, 1995). The range of the consensus observation,  $[c_{\min}, c_{\max}]$ , can then be calculated as follows:

$$c_{\min} = \min_{S_i \in T_c} [C(\pi_v^i)], \quad (84)$$

$$c_{\max} = \max_{S_i \in T_c} [C(\pi_v^i)]. \quad (85)$$

Knowing the extent of the consensus observation, the measure of consistency can be defined by the following:

$$Co_c(\pi_v^i, T_c) = \begin{cases} |c_{\min} - C(\pi_v^i)|, & C(\pi_v^i) < c_{\min}, \\ |c_{\max} - C(\pi_v^i)|, & C(\pi_v^i) > c_{\max}, \\ 1, & \text{otherwise.} \end{cases} \quad (86)$$

Considering Eq. (86) sensors that have their *centres of gravity* within the range of the consensus observation are 100% consistent and, thus, are modelled by  $Co_c(\pi_v^i, T_c) = 1$ . Furthermore, when the distance of a sensor's observation to the boundaries of the consensus observation increases, the consistency of the observation with reference to the consensus observation decreases, and hence,  $Co_c(\pi_v^i, T_c) = [0, 1)$ .

However, in practical applications, no consensus observation may be found amongst all the sensors, e. g.  $|T_c| = 1$ . Then, the measurement of consistency, as defined by Eq. (86), becomes inadequate. Nevertheless, a consistency measure that evaluates the amount of conflict among the sensors is both obvious and adequate. It should be noted that reliabilities amongst sensors can vary. It is debatable whether a reliable sensor, that is in conflict with an unreliable sensor should be considered inconsistent. Accordingly, the consistency measurement should take the reliabilities of sensors into account. WAM meets the desired features stated above and thus, is used as follows:

$$Co_a(\pi_v^i, \pi_v^*) = \max \left[ 1 - \max_{j \neq i} [r_j], \lambda_{WAM}(\vec{\rho}^i, \vec{d}_i) \right], \quad (87)$$

where

$$\pi_v^* = \{\pi_v^j\}_{j \neq i}, \quad \vec{\rho}^i = (\rho^j)_{j \neq i}, \quad \vec{d}_i = (d_{i,j})_{j \neq i},$$

and

$$d_{i,j} = 1 - \left| C(\pi_v^i) - C(\pi_v^j) \right|. \quad (88)$$

With reference to Eq. (87), the consistency of a sensor's observation is the weighted average of all distances from the observation of the  $i$ th sensor  $\pi_v^i$  to the observations of all other sensors  $\pi_v^*$ . The weighting vector of the weighted arithmetic mean corresponds to the reliabilities of the sensor's observations  $\pi_v^*$ . Moreover, due to the max-operator in Eq. (87), the consistency of  $\pi_v^i$  will be equal to one, if all observations of  $\pi_v^*$  are completely unreliable. Both versions of evaluating the consistency of a sensor's observation are brought together in the following equation:

$$Co(\pi_v^i) = \begin{cases} Co_c(\pi_v^i, T_c), & |T_c| > 1, \\ Co_a(\pi_v^i, \pi_v^*), & \text{otherwise.} \end{cases} \quad (89)$$

After determining the consistency of each observation it is natural to consider this information during the fusion process. Thus, an observation with a low consistency should contribute little to the fusion result and vice versa. Therefore, the MOGPFR is modified, so that the consistency index of each observation is explicitly taken into account.

From Eq. (80), it follows that the determination of the weighting vector  $\vec{v}$  as defined in Eq. (82) has to be modified to meet the desired features mentioned above and stated in Lemma 5. A functional description, that meets the needs postulated previously is called *credibility*, denoted as  $Cr(\rho^i, Co(\pi_v^i)) = \min[\rho^i, Co(\pi_v^i)]$  (Yager & Kelman, 1996). Thus, Eq. (82) is modified as follows:

$$\vec{v} = \left( Cr(\rho^1, Co(\pi_v^1)), \dots, Cr(\rho^n, Co(\pi_v^n)) \right). \quad (90)$$

Next, whether Eq. (90) is in accordance with Lemma 2 is checked.

*Proof.* The modification of Eq. (82) does not violate Lemma 2, since if all sources  $\{S_2, \dots, S_n\}$  except  $S_1$  are unreliable, e. g.  $\vec{v} = (1, 0, \dots, 0)$  (cf. Eq. (82)) then

$$\vec{v} = \left( Cr(1, Co(\pi_v^1)), \dots, Cr(0, Co(\pi_v^n)) \right) = (1, 0, \dots, 0).$$

With respect to the proof of Lemma 2, only  $\pi_v^1$  is considered. □

Using Eq. (90) for determining the importance weighting vector  $\vec{v}$  of Eq. (79) satisfies Lemma 5.

*Proof.* Consider a  $n$ -dimensional fusion problem of a set of  $n$  sources  $S = \{S_1, \dots, S_i, \dots, S_n\}$  and let all sources be reliable, e. .  $\vec{v} = (1, \dots, 1)$ . Moreover, let  $T_c = \{S_1, \dots, S_k\}$ . With reference to Eq. (89), the importance weighting vector  $\vec{v}$  becomes  $\vec{v} = (1, \dots, 1, a_{k+1}, \dots, a_n)$ , where  $\{a_{k+1}, \dots, a_n\} < 1$  and therefore, the group of sources building the consensus observation, that is the majority observation, is prioritised.  $\square$

#### 4.2.5 Sensor reliability

As information provided by physical sensors can vary in quality, reliability is important for the success of multi-source data fusion. Two general concepts of obtaining source's reliability are established in literature (Rogova & Nimier, 2004): *static* and *dynamic reliability*. Static reliability indicates the probability that a source is working correctly. However, it does not consider unpredictable disturbances, such as environmental influences, mis-applications, or failures caused by accidents. Paying no heed to those ascendancies effectively lowers the performance of a fusion system. That is why here both static reliability, based on expert know-how, and consensus-driven dynamic reliability estimation are used.

Consensus-driven approaches presume that unreliable sources result in conflicting situations. In this context, it is assumed that a higher conflict results in a larger distance from the consensus observation. Accordingly, the measurement of consistency (cf. Eq. (89)) is ideally suited for the determination of the reliability of a source. If there are unreliable sources that do not cause conflicts, consensus-driven reliability estimation methods become inadequate.

The consistency of a source can vary widely between fusion steps, because of noise in real-life fusion problems. Therefore, it makes sense to smooth the reliability of a source by considering its former reliabilities. The family of *exponential moving average filters*, introduced below, allows modelling the needs postulated herein.

The exponential moving average filter can be classified under *infinite impulse response* filters and is defined as follows:

$$y[n] = \omega \cdot x[n] + (1 - \omega) \cdot y[n - 1], \quad (91)$$

where  $\omega \in [0, 1]$ . The filter output  $y[n]$  and the filter input  $x[n]$  are defined for the discrete time step  $n$ . Moreover,  $y[n - 1]$  is the filter output at the previous discrete time step  $[n - 1]$ .

Exponential filtering of the consistency of a source allows modelling of its dynamic reliability,  $\rho_d^i[n]$ , intuitively. Additionally, the static reliability,  $\rho_s^i[n]$ , can be used as a convex envelope for the smoothed consistency of a source. This leads to the following definition of dynamic reliability:

$$\rho_d^i[n] = \min [\rho_s^i[n], \hat{\rho}_d^i[n]], \quad (92)$$

where

$$\hat{\rho}_d^i[n] = \omega \cdot Co(\pi_v^i) + (1 - \omega) \cdot \hat{\rho}_d^i[n - 1], \quad (93)$$

and

$$\hat{\rho}_d^i[n] = 1, \quad \forall n < 0. \quad (94)$$

The inertia of a process monitored by a source can be taken into account by aligning  $\omega$ . Thus, a high inertia should be accounted by a small  $\omega$  and vice versa. To illustrate the effect of exponential filtering, Fig. 7 depicts the filtered and non-filtered reliability of an arbitrary source. It is clearly visible that the influence of process noise is considerably reduced due to exponential averaging.

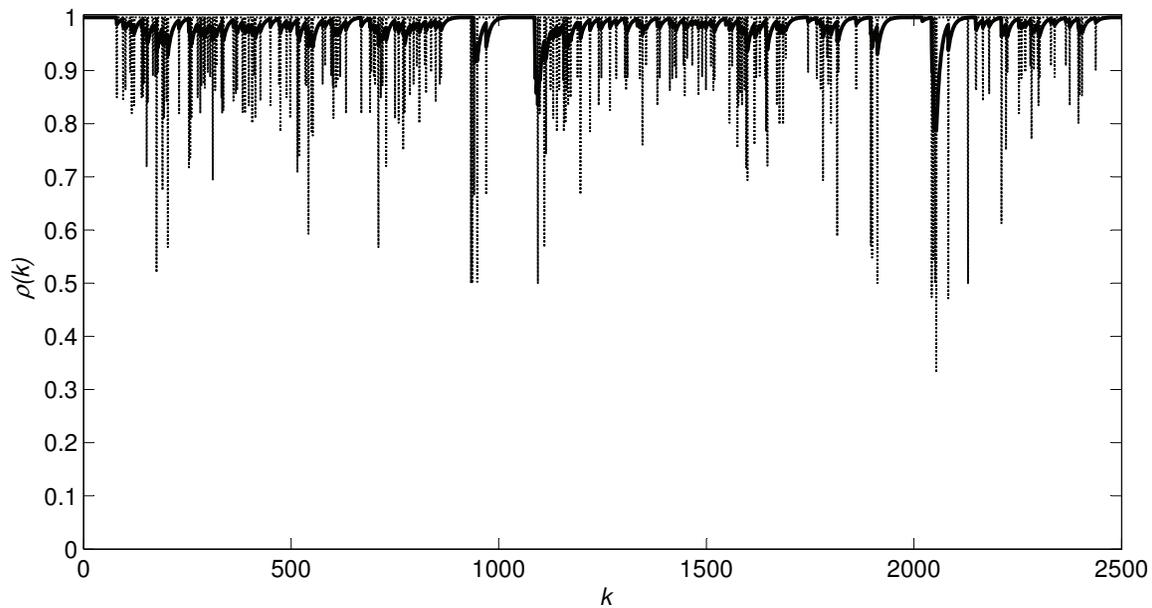


Fig. 7. The smoothed and non-smoothed reliability of an arbitrary source. The dashed line visualises the non-smoothed reliability. The thick line illustrates the smoothed reliability when an exponential moving average filter with  $\omega = 0.2$  is applied.

#### 4.2.6 Trustworthiness of the fusion result

An important problem after obtaining a fusion result is the question of its trustworthiness. Delmotte et al. (Delmotte, 2007) proposed an approach that measures the trustworthiness of a fusion result by considering internal contradictions and reliabilities of sources. Here, Delmotte's idea of determining the trustworthiness of fusion results is captured and modified so that it qualifies for the MOGPFR.

In Sect. 4.2.4, the measurement of credibility was introduced. Here, the average credibility of all sensors is accounted for the trustworthiness of the fusion result. In this context, a high overall credibility should result in a high trustworthiness and vice versa. As reliabilities among sources can vary, it is natural that the trustworthiness should mainly be influenced by highly reliable sources, whereas sources with low reliability should influence the trustworthiness of the fusion result only to a small amount. Consequently, the weighted arithmetic mean is appropriate for calculating the trustworthiness. In this context, the importance weighting vector  $\vec{v}$  equals the reliabilities of the sources to be merged.

With reference to the MOGPFR, the consistency index  $h(T_c)$ , (cf. Eq. (71)), contains information about the trustworthiness of the fusion result too. That is, a high consistency index indicates a high agreement amongst the sources that build the consensus observation. On the contrary, a low consistency index indicates a low harmony in the consensus observation. Accordingly, a high consistency index should result in a high trust in the fusion result and vice versa.

As both the overall credibility of the sources and the consistency index are equally important to the trustworthiness of a fusion result, denoted by  $Tr$ , it is defined as the average of both the indexes:

$$Tr \left( \pi_{\vec{v}}^{1, \dots, n}(x) \right) = 0.5 \cdot h(T_c) + 0.5 \cdot \lambda_{\text{WAM}}(\vec{\rho}, \vec{c}), \quad (95)$$

where

$$\vec{c} = \left( Cr(\rho^1, Co(\pi_{\vec{v}}^1)), \dots, Cr(\rho^n, Co(\pi_{\vec{v}}^n)) \right),$$

and

$$\vec{\rho} = (\rho^1, \dots, \rho^n).$$

### 4.3 Two-Layer Possibilistic Fusion

Complex fusion tasks, such as condition monitoring of printing machines, require the fusion of different attributes, because sophisticated processes cannot be described sufficiently by one attribute. In general, an attribute describes the state of a variable that influences the condition of a process to be monitored. Furthermore, an attribute consists of several criteria that determine the attribute's condition. These criteria are monitored by a set of sensors. Thus, the fusion of highly sophisticated processes is a twofold process: i) the criteria corresponding to one particular attribute are fused; ii) the attributes are aggregated into one evidence about the condition of the machine. The fusion approach introduced in this section is called the *Two-Layer Possibilistic Fusion Rule (TLPFR)*. The TLPFR is presented in an application-oriented manner for the purpose of machine conditioning. However, other applications of the TLPFR are also conceivable.

Consider a  $n$ -dimensional fusion problem of a set of  $n$  attributes  $A = \{A_1, \dots, A_i, \dots, A_n\}$ . Furthermore, let each attribute  $A_i$  consist of  $m_i$  criteria observed by  $m_i$  sources  $S_{A_i} = \{S_{A_{i,1}}, \dots, S_{A_{i,j}}, \dots, S_{A_{i,m_i}}\}$ . Moreover, let each source  $S_{A_{i,j}}$  provide its observation using a possibility distribution  $\pi_{A_i}^j(x)$ . Thus, the fusion of  $m_i$  criteria into one result about the condition of one particular attribute  $A_i$  can be considered as a  $m_i$ -dimensional possibilistic fusion problem. The fusion on the first layer, the so-called *criteria-fusion layer*, is accomplished using the MOGPFR. Furthermore, the compound consequence about the state of a particular attribute  $A_i$  is defuzzified using the *mean of maxima* (Klir & Yuan, 1995) and additionally, the attribute's trustworthiness (cf. Eq. (95)) is considered. Thus, for each attribute, one obtains numerical values indicating its current condition  $x_{A_i}$  and its corresponding trustworthiness  $Tr(\pi_{A_i}^{1, \dots, m_i}(x))$ . Figure 8 illustrates the criteria-fusion layer in its lower part.

After obtaining the condition of each attribute and its related trustworthiness, the attributes' conditions are merged on the second layer. In this context, it is obvious that when one attribute of the machine indicates that the machine is in a "bad" condition, then the holistic condition of the machine should accommodate this condition up to a certain degree. This degree depends on the trustworthiness of an attribute's observation. That is to say, if the trustworthiness of an attribute is high, the holistic condition accounts for this condition to a high degree and vice versa. Therefore, the fusion method on the second layer, denoted as *attribute-fusion layer*, should pay attention to the trustworthiness of an attribute's observation and should model a high degree of pessimism. Once again, the class of IIWOWA operators qualifies for the fusion on the second layer. The attribute-fusion layer method is defined as follows:

$$x_{A_1, \dots, A_n} = \lambda_{\text{IIWOWA}}(\vec{t}, \vec{w}, \vec{\pi}), \quad (96)$$

where

$$\vec{\pi} = (\pi_{A_1}^{1, 2, \dots, m_1}(x), \dots, \pi_{A_n}^{1, 2, \dots, m_n}(x)),$$

$$\vec{t} = (Tr(\pi_{A_1}^{1, 2, \dots, m_1}(x)), \dots, Tr(\pi_{A_n}^{1, 2, \dots, m_n}(x))),$$

where  $\vec{w}$  is determined with reference to the desired andness degree, using O'Hagens approach of maximum entropy (O'Hagen, 1987). The second layer of the TLPFR is depicted in the upper part of Fig. 8.

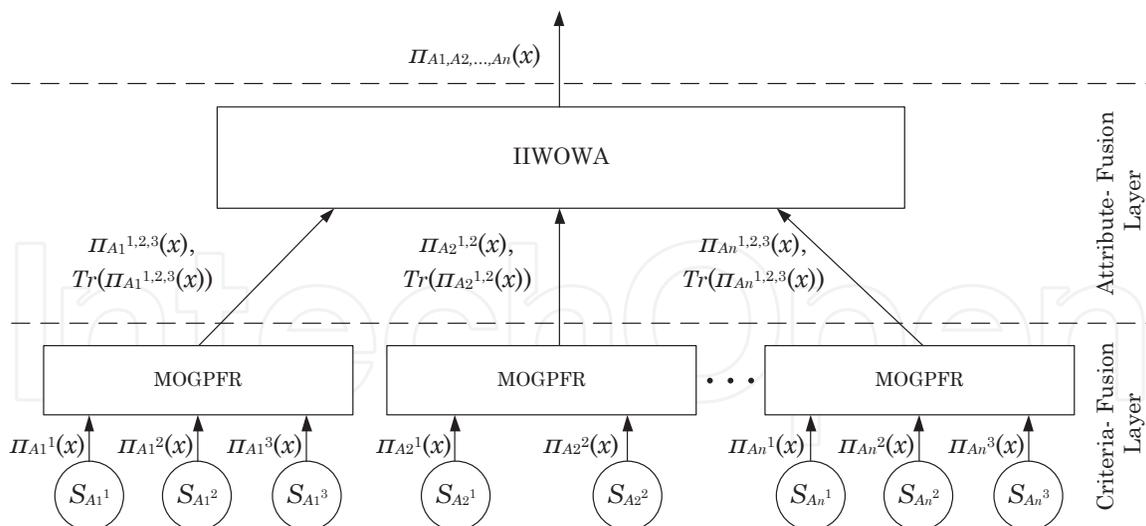


Fig. 8. The TLPFR for machine conditioning.

## 5. Application example

Nowadays, security prints, such as banknotes are highly sophisticated and exhibit many security features. To reach this high security level, several printing methods like line offset, letter-press printing, foil printing, and intaglio printing are applied. Especially the latter one named printing technique—intaglio printing—is of major importance for the security of banknotes. The term “intaglio” is of Italian origin and means “to engrave”. The printing method of the same name uses a metal plate with engraved characters and structures. Large areas and differences in brightness are produced by different hatchings. During the printing process the engraved structures are filled with ink and pressed under huge pressure directly onto the paper. As a result, a tactile relief and fine lines are formed, unique to the intaglio printing process and almost impossible to reproduce through commercial printing methods.

### 5.1 Experiments

In the steel engraving machine the wiping unit is the most observed part. It is responsible for removing surplus ink around the engravings. Even small parameter manipulations cause errors on products (cf. Fig. 9). Experienced machine operators are able to recognise errors before they occur and stabilise the production by changing mainly the wiping unit parameters. In



Fig. 9. Wiping error (right) in an intaglio printing process (Glock et al., 2011).

total, six sensors are mounted on the intaglio printing machine. The main focus of attention is on the wiping process of the printing machine, because here a large number of printing flaws can be caused (Lohweg & Mönks, 2010a). The following functional components are monitored:

- The motor current (MC) of the machine's main drive. It is mainly influenced by the friction between the wiping and the plate cylinder.
- The printing pressure of side 1<sup>1</sup> (PPS1) and the printing pressure of side 2 (PPS2).
- The pressure of the drying blade onto the wiping cylinder of side 1 (DBPS1) and side 2 (DBPS2).
- The flow rate of the wiping solution (WSFR), that is, the amount of wiping solution that is sprayed onto the wiping cylinder.

An experiment was performed by measuring all six sensor signals when the intaglio printing machine was running (Dyck et al., 2007). The test was performed in two steps: i) the machine was operated in "good" production mode; ii) the wiping pressure was decreased little-by-little until the machine printed only flawed sheets. For both experiments, a sample frequency of 7 kHz was used and the production rate kept constant at 6500 sheets per hour. During the first and the second experiment, 479 sheets (within approximately 4.4 minutes) and 796 sheets (within approximately 7.4 minutes) were printed (Dyck et al., 2007).

Sensors with equal causal relationships were arranged to the same attribute. The causality relationships were determined using expertise of the printing machine manufacturer. In this context, a specific machine attribute is related to one particular malfunction, which may cause printing errors and may give rise to machine faults. In this experiment, three machine attributes were monitored. They are stated in the following:

#### A1. *Drying Blade Pressure*

A misalignment of the drying blade pressure, that is, the pressure between the wiping cylinder and the drying blade, may cause so-called *wiping errors*. In particular, an excessive pressure may ruin the drying blade, whereas an undersized pressure may directly cause flaw print errors. The drying blade pressure is proportional to the motor current (MC) and of course, can be measured by the DBPS1 and DBPS2 sensors.

#### A2. *Cleanness of the Wiping Cylinder*

A dirty wiping cylinder may provoke another type of wiping error. Naturally, the cleaning power of the wiping unit is influenced by the amount of detergent (wiping solution) that is sprayed onto the wiping cylinder. This flow of detergent is monitored by the WSFR sensor. Moreover, dried ink that adheres onto the wiping cylinder rubs against the drying blade and increases the friction between the wiping cylinder and the drying blade. This leads to an increase of the drying blade pressure that can be detected using the DBPS1 and DBPS2 sensors.

#### A3. *Printing Pressure*

A misalignment of the printing pressure may bring out imperfect printed structures. Naturally, the printing pressure is monitored by the PPS1 and PPS2 sensors.

### 5.2 Performance evaluation of the Two-Layer Possibilistic Fusion Rule

With reference to Fig. 10, it is clearly observable that the holistic machine condition decreases gradually because the wiping pressure is decreased little-by-little. With reference to the optical inspection system (the green line in Fig. 10) from sheet number 560 onward, the printing machine produced only low-quality products, that is, wiping errors occurred. Consider a threshold of 50% (0.5) of  $\max(\pi(x)) = 1$ ; the TLPFR diagnosed this "bad" machine condition

<sup>1</sup> In the field of printing machines, the side where the machine's main drive is located is called "side 1" and the side where the machine's control centre is located is called "side 2".

approximately 160 sheets, or 1.47 minutes, prior to the occurrence of the wiping errors. The "bad" machine condition was detected by the machine's attributes  $A_1$  and  $A_2$  (wiping pressure and cleanliness of the wiping cylinder), whilst attribute  $A_3$  indicated a "good" machine condition. This behaviour of the machine's attributes is intuitive, because the wiping pressure was decreased little-by-little, whilst the printing pressure was kept constant.

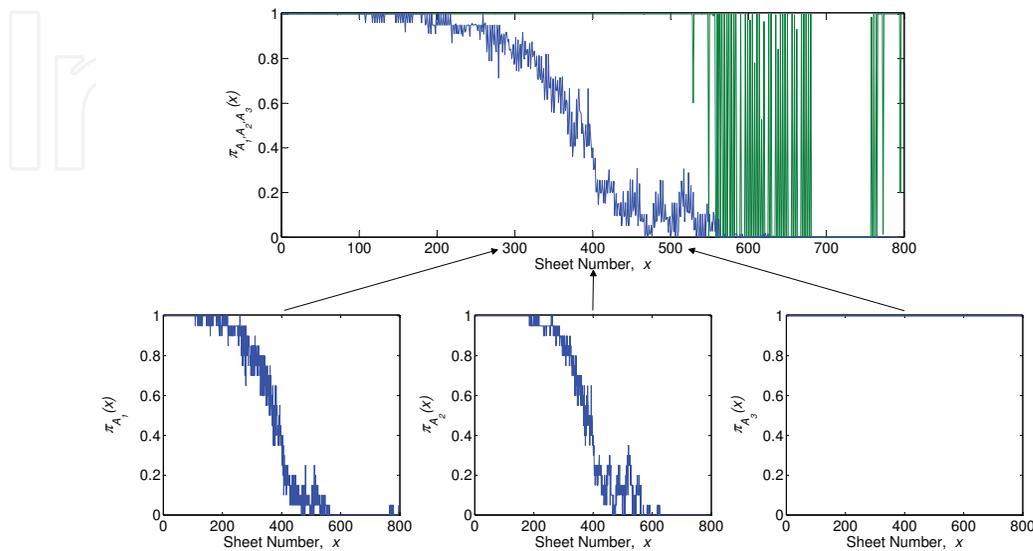


Fig. 10. The fusion results of the wiping error experiment. In the plot above, the blue line illustrates the holistic machine condition determined by the TLPFR; and the green line illustrates the membership grades of the optical inspection system. The plots underneath depict the conditions of the attributes.

The fuzzified values of all the sensors obtained during the wiping error experiment are depicted in Fig. 11. With reference to Fig. 11 (a), all sensors that describe attribute  $A_1$

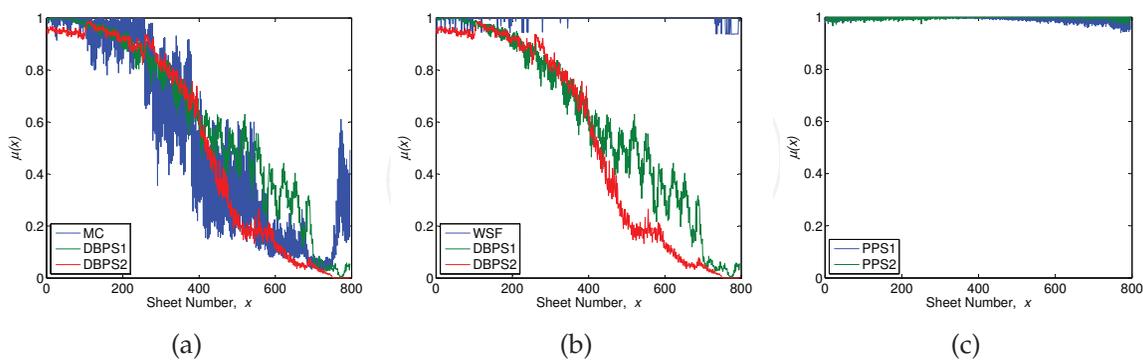


Fig. 11. The membership grades of the sensors during the wiping error experiment. Panel (a) illustrates the fuzzified sensor outputs of attribute  $A_1$ , (b) – of attribute  $A_2$ , and (c) – of attribute  $A_3$ .

detected the wiping error. Furthermore, the motor current was highly afflicted with additive noise, which caused slight conflicts. This arose with a decrease of the trustworthiness of  $A_1$  (cf. Fig. 12 (a)) and in a discounted reliability of the motor current sensor (cf. Fig. 12 (d)).

The conflicts were mainly solved at sheet number 560, that is, when the printing machine was producing only erroneous sheets. Therefore, the reliability of the MC sensor was upgraded. Considering Fig. 11 (b), the WSFR sensor did not detect the decrease of the wiping pressure, whereas the DBPS1 and the DBPS2 sensors did detect. As a result, the reliability of the WSFR sensor was decreased, until it was considered totally unreliable from sheet number 600 onwards. The trustworthiness of  $A_2$  was lowered because of the conflict between the WSFR sensor and both DBPS sensors. From sheet number 600 onwards, that is, when the reliability of the WSFR sensor accounted to zero, the trustworthiness of  $A_2$  increased steadily until it accounted to one, because the observation of the WSFR sensor was no longer considered during the fusion. As mentioned above, the third attribute did not detect the loss of the wiping pressure. Thus, the membership grades of both sensors remained constant and accounted to one (cf. Fig. 11 (c)). Therefore, no conflicts occurred, and the reliability of both sensors, as well as the trustworthiness of the attribute accounted constantly to one (cf. Fig. 12 (c) and (f)). The decrease of the wiping pressure was clearly indicated in the holistic machine condition, because of the high degree of pessimism ( $\alpha = 0.8$ ) for the fusion of the attributes, although it was not detected by  $A_3$ .

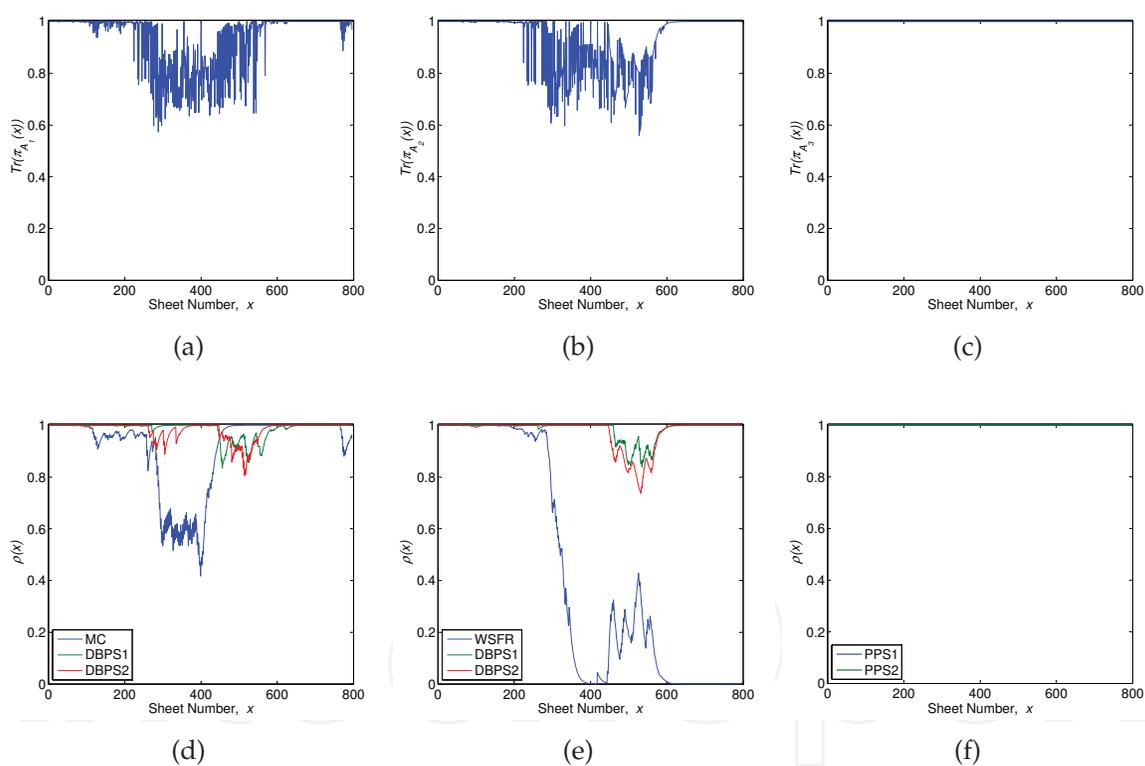


Fig. 12. The trustworthiness of the attributes conditions and the reliabilities of the sensors during the wiping error experiment. Panel (a) illustrates the trustworthiness of attribute  $A_1$ , (b) – of  $A_2$ , and (c) – of  $A_3$ . Furthermore, panel (d) illustrates the reliabilities of the sensors forming attribute  $A_1$ , (e) – of  $A_2$ , and (f) – of  $A_3$ .

Considering the discussions in this section, one can conclude that the TLPFR is insofar superior to an optical inspection system in respect of print flaws generated by machine errors. This is mainly because faulty machine conditions can be detected prior to the occurrence of printing errors. This makes it possible to readjust the printing machine to avoid print flaws.

## 6. Conclusion and outlook

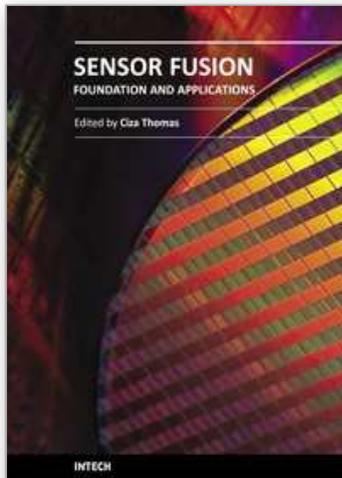
In this chapter we have reviewed the nature of knowledge perceived by experts and the representation of incomplete, inconsistent, and vague information in technical systems. Based on various theories, namely: Probability Theory, Evidence Theory, Fuzzy Set Theory, and Possibility Theory, different concept of information fusion in the appearance of incomplete and unreliable data was presented. Furthermore, two human-centric fusion approaches were characterised. In particular, a new possibilistic framework for sensor fusion was presented. In the future we will concentrate on implementation aspects of the above mentioned framework.

## 7. References

- Ayyub, B. M. & Klir, G. J. (2006). *Uncertainty Modeling and Analysis in Engineering and the Sciences*, Chapman and Hall/CRC, Taylor and Francis Group.
- Bayes, T. P. (1763). An Essay Towards Solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London, London* 53: 370–418.
- Black, M. (1937). Vagueness. an exercise in logical analysis, *Philosophy of Science* 4(4): 427–455.
- Bocklisch, S. F. & Priber, U. (1986). A parametric fuzzy classification concept, *Proc. International Workshop on Fuzzy Sets Applications*, Akademie-Verlag, Germany: 147–156.
- Bossé, E., Roy, J. & Wark, S. (2007). *Concepts, Models, and Tools for Information Fusion*, Intelligence and Information Operations Library, Artech House, Boston, Mass.  
URL: <http://www.gbv.de/dms/bowker/toc/9781596930810.pdf>
- Campos, F. (2006). *Decision Making in Uncertain Situations: An Extension to the Mathematical Theory of Evidence*, Boca Raton, Florida USA.
- Carnap, R. (1947). Probability as a guide in life, *Journal of Philosophy* 44(6): 141–148.
- Delmotte, F. (2007). Detection of Defective Sources in the Setting of Possibility Theory, *Fuzzy Sets and Systems* 5(158): 555–571.
- Dempster, A. P. (1967). Upper and Lower Probabilities Induced by a Multivalued Mapping, *Annals Mathematics Statistics* 28: 335–339.
- Dubois, D., Esteva, F., Godo, L. & Prade, H. (2001). An Information-based Discussion of Vagueness, *The 10th IEEE International Conference on Fuzzy Systems, Toulouse* 2(3): 781–0784.
- Dubois, D., Foulloy, L., Mauris, G. & Prade, H. (2004). Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities, *Reliable Computing* 10(5): 273–297.
- Dubois, D. & Prade, H. (1978). Operations on Fuzzy Numbers, *International Journal of Systems Science* 9(6).
- Dubois, D. & Prade, H. (1983). Unfair Coins and Necessity Measures Towards a Possibilistic Interpretation of Histograms, *Fuzzy Sets and Systems* 1(10): 15–20.
- Dubois, D. & Prade, H. (1994). Possibility Theory and Data Fusion in Poorly Informed Environments, *Control Engineering Practice* 5(2): 811–823.
- Dubois, D. & Prade, H. (2000). Possibility Theory in Information Fusion, *Proceedings of the Third International Conference in Information Fusion, Paris* 3(1): 6–19.
- Dubois, D. & Prade, H. (2003). Possibility Theory and its Applications: A Retrospective and Prospective View, *The 12th IEEE International Conference on Fuzzy Systems, San Antonio* 5(1): 5–11.

- Dujmović, J. J. (1974). Weighted Conjunctive and Disjunctive Means and their Application in Systems Evaluation, *of the University of Belgrade, EE Dept., Series Mathematics and Physics, Belgrade* 483: 147–158.
- Dujmović, J. J. & Larsen, H. L. (2007). Generalized Conjunction/Disjunction, *International Journal of Approximate Reasoning* 46(2): 423–446.
- Dyck, W., Türke, T. & Lohweg, V. (2007). A Fuzzy-Pattern-Classifer-Based Adaptive Learning Model for Sensor Fusion, *2007 IEEE Workshop on Machine Learning for Signal Processing, Thessaloniki* pp. 282–287.
- Eisenführ, F. (2003). *Rationales Entscheiden*, 3 edn, Springer, Berlin.
- Glock, S., Voth, K., Schaede, J. G. & Lohweg, V. (2011). A Framework for Possibilistic Multi-source Data Fusion with Monitoring of Sensor Reliability, *World Conference on Soft Computing. In: San Francisco, CA, USA, May 23-26, 2011 May 2011*.
- Hacking, I. (1975). *The Emergence of Probability*, Cambridge University Press, Cambridge.
- Hall, D. L. & Llinas, J. (2001). *Handbook of a Multisensor Data Fusion*, The electrical engineering and applied signal processing series, CRC Press, Boca Raton, Fla.  
URL: <http://www.gbv.de/dms/hbz/toc/ht013184121.pdf>
- Keynes, J. M. (1921). *A Treatise on Probability*, Dover Publications, London.
- Klir, G. J. & Wierman, M. J. (1998). *Uncertainty-Based Information: Elements of Generalized Information Theory (Studies in Fuzziness and Soft Computing)*, Physica-Verlag.
- Klir, G. J. & Yuan, Y. (1995). *Fuzzy Set and Fuzzy Logic: Theory and Applications*, Prentice-Hall Inc., New York.
- Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1 edn, Julius Springer, Berlin.
- Larsen, H. L. (1999). Importance Weighted OWA Aggregation of Multicriteria Queries, *18th International Conference of the North American Fuzzy Information Processing Society, New York* pp. 740–744.
- Larsen, H. L. (2002). Efficient Importance Weighted Aggregation Between Min and Max, *9th Conference on Information Processing and Management of Uncertainty in Knowledgebased Systems, Annecy* 1(1): 1203–1208.
- Li, R. & Lohweg, V. (2008). A Novel Data Fusion Approach using Two-Layer Conflict Solving, *IAPR Workshop on Cognitive Information Processing, Santorini* pp. 132–136.
- Lipshitz, R., Klein, G., Orasanu, J. & Salas, E. (2001). Taking Stock of Naturalistic Decision Making, *Journal of Behavioral Decision Making* 14: 331–352.
- Lohweg, V. & Mönks, U. (2010a). *Fuzzy-Pattern-Classifer based Sensor Fusion for Machine Conditioning*, Sensor Fusion, I-TECH Education and Publishing, Vienna.
- Lohweg, V. & Mönks, U. (2010b). Sensor Fusion by Two-Layer Conflict Solving, *The 2nd International Workshop on Cognitive Information Processing, Elba* 2(1): 65–79.
- Murphy, C. K. (2000). Combining Belief Functions when Evidence Conflicts., *Decision Support Systems* 29(1): 1–9.
- O'Hagen, M. (1987). Fuzzy Decision Aids, *Twenty-first Asiomar Conference on Signal, Systems and Computers, Pacific Grove* 21(2): 624–628.
- Popper, K. R. (1934). *The Logic of Scientific Discovery*, Routledge; Reprint edition (October 8, 1992), London.
- Ramsey, F. P. (1960). *The Foundations of Mathematics and Other Logical Essays*, Paterson, N.J., Littlefield, Adams.

- Rogova, G. L. & Nimier, V. (2004). Reliability in Information Fusion: Literature Survey, *Proceedings of the Seventh International Conference on Information Fusion, Stockholm* pp. 1158–1165.
- Salicone, S. (2007). *Measurement Uncertainty - An Approach via the Mathematical Theory of Evidence*, Springer, New York.
- Schubert, J. (2007). An information fusion demonstrator for tactical intelligence processing in network-based defense., *Journal Information Fusion* 8(1).
- Shafer, G. (1976). *A Mathematical Theory of Evidence*, 1 edn, Princeton University Press, Princeton.
- Smarandache, F. & Dezert, J. (2006). *Advances and Applications on DSmtT for Information Fusion: Collected Works 1 (2)*, American Research Press, Rehboth USA.
- Sunita, S. A. (1999). Order Effects and Memory for Evidence in Individual versus Group Decision Making in Auditing, *Journal of Behavioral Decision Making* 12: 71–88.
- Yager, R. R. (1979). Possibilistic Decisionmaking, *IEEE Transactions on Systems, Man and Cybernetics* 9(2): 388–392.
- Yager, R. R. (1987). On the Dempster-Shafer Framework and New Combination Rules, *Information Sciences* 41(2): 93–137.
- Yager, R. R. (1988). On Ordered Weighted Averaging Aggregation Operators in Multi-criteria Decision Making, *IEEE Transactions on Systems, Man and Cybernetics* 18: 183–190.
- Yager, R. R. (1994). Aggregation Operators and Fuzzy Systems Modeling, *Fuzzy Sets and Systems* 67(2): 129–145.
- Yager, R. R. (2004). A Framework for Multi-source Data Fusion, *Information Sciences* 163(1-3): 175–200.
- Yager, R. R. & Kelman, A. (1996). Fusion of Fuzzy Information With Considerations for Compatibility, Partial Aggregation, and Reinforcement, *International Journal of Approximate Reasoning* 15(2): 93–122.
- Yager, R. R. & Liu, L. (2008). *Classic Works of the Dempster-Shafer Theory of Belief Functions*, Springer, New York.
- Zadeh, L. A. (1965). Fuzzy Sets, *Information and Control* 8.
- Zadeh, L. A. (1978). Fuzzy Sets as a Basis for a Theory of Possibility, *Fuzzy Sets and Systems* 1: 3–28.
- Zadeh, L. A. (1986). A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination., *AI Magazine* 7(2): 85–90.
- Zadeh, L. A. (2008). Toward Human Level Machine Intelligence - Is It Achievable? The Need for a Paradigm Shift, *IEEE Computational Intelligence Magazine* 3(3): 11–22.



## **Sensor Fusion - Foundation and Applications**

Edited by Dr. Ciza Thomas

ISBN 978-953-307-446-7

Hard cover, 226 pages

**Publisher** InTech

**Published online** 24, June, 2011

**Published in print edition** June, 2011

Sensor Fusion - Foundation and Applications comprehensively covers the foundation and applications of sensor fusion. This book provides some novel ideas, theories, and solutions related to the research areas in the field of sensor fusion. The book explores some of the latest practices and research works in the area of sensor fusion. The book contains chapters with different methods of sensor fusion for different engineering as well as non-engineering applications. Advanced applications of sensor fusion in the areas of mobile robots, automatic vehicles, airborne threats, agriculture, medical field and intrusion detection are covered in this book. Sufficient evidences and analyses have been provided in the chapter to show the effectiveness of sensor fusion in various applications. This book would serve as an invaluable reference for professionals involved in various applications of sensor fusion.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Volker Lohweg, Karl Voth and Stefan Glock (2011). A Possibilistic Framework for Sensor Fusion with Monitoring of Sensor Reliability, Sensor Fusion - Foundation and Applications, Dr. Ciza Thomas (Ed.), ISBN: 978-953-307-446-7, InTech, Available from: <http://www.intechopen.com/books/sensor-fusion-foundation-and-applications/a-possibilistic-framework-for-sensor-fusion-with-monitoring-of-sensor-reliability>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen