Multi-Modal Sensor Fusion Algorithms for Robotics

eingereichtes
ADVANCED SEMINAR
von

cand. ing. Richard Leibrandt

geb. am 23.07.1986
wohnhaft in:
Friedenheimer Str. 41
80686 München
Tel.: 015156503216

Lehrstuhl für
STEUERUNGS- und REGELUNGSTECHNIK
Technische Universität München

Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss
Univ.-Prof. Dr.-Ing. Sandra Hirche

Betreuer:	M.Sc. Cristian Axenie
Beginn:	24.07.2012
Abgabe:	10.09.2012

Abstract

In order to operate in and interact with an environment, machines need to possess knowledge about this environment. Most of the time this knowledge can not be stored in the machine a priori and thus the machine needs sensors in order to retrieve information about the world during operation. Where one sensor is helpful, many sensors are able to increase the reliability of sensory data or even gain data a single sensor could not have alone. For that sensor fusion is needed.

This work gives a short introduction in sensor fusion. Afterwards two powerful sensor fusion methods are selected: The mathematical/statical method “Dempster-Shafer Evidential Reasoning” and the bio-inspired method “Multi-directional ARTMAP”. Both methods have come a long way and are nevertheless up-to-date and further developed. Both methods are explained in more detail. Afterwards they are compared. Particular attention is paid to how they deal with uncertainty, under which conditions they provide reliable results and implementation aspects. Method explanations and comparison are illustrated on a practical robot identification example. The conclusion can be drawn that both methods are complex yet powerful instruments. Which one should be chosen depends on the problem at hand. The comparison might provide insight on how to choose the right tool.

Zusammenfassung

Um in und mit einer Umgebung operieren zu können, müssen Maschinen Wissen über diese Umgebungen besitzen. Meistens kann dieses Wissen der Maschine nicht im Voraus zugeführt werden. Daher benötigt sie Sensoren um sich die benötigten Informationen während des Betriebes zu gewinnen. Wo ein Sensor hilfreich ist, können mehrere Sensoren die Vertrauenswürdigkeit der sensorischen Daten erhöhen oder sich sogar Wissen erschließen, das mit nur einem Sensor nicht zugänglich gewesen wäre. Dazu ist Sensorfusion notwendig.

Diese Arbeit gibt eine kurze Einführung in Sensorfusion. Anschließend werden zwei leistungsstarke Sensorfusionsmethoden ausgesucht: Die mathematisch-statistische Methode “Evidenztheorie von Dempster und Shafer” und “Multi-direktionales ARTMAP”. Die Grundsteine beider Methoden sind bereits vor einigen Jahren gelegt worden, gehören beide Methoden zum Stand der Technik und befinden sich in der Weiterentwicklung. Beide Methoden werden etwas detaillierter beschrieben. Anschließend werden sie verglichen. Ein besonderes Augenmerk wird dabei darauf gelegt, wie sie mit Unsicherheiten umgehen, unter welchen Bedingungen sie verlässliche Ergebnisse liefern und Aspekte in der Anwendung. Die Methodenbeschreibungen und der Vergleich werden an einem praktischen Roboteridentifizierungsbeispiel illustriert. Das Fazit kann gezogen werden, dass beide Methoden komplexe, aber leistungsstarke Instrumente sind. Die Wahl der Methode sollte abhängig vom Problem getroffen werden. Der Vergleich in dieser Arbeit kann bei der Wahl des geeigneten Werkzeugs helfen.

Contents

Chapter 1

Introduction

1.1 The Challenge, Ideas and Contributions

When we imagine a machine (e.g. a robot) that is supposed to interact with the outside environment, it will need information about this world. The machine might get this information from a model of this world. The world, however, is often unknown to the machine: Its creators might have lacked knowledge about the environment, it might have not been feasible to store the large amount of information in the machine, or unforeseen events might occur in the dynamically changing world. In any case, sensors are necessary for the machine as are for a human. As a human uses various senses combined to find his way through his world, a question is raised: How can multiple sensors be used together to improve the machine's understanding of its environment?

Major benefits were to *gain more reliable information and/or information that might have not been sensed otherwise.* Put in a nutshell, due to the synergistic effect, the value of the combined information is greater than the sum of the value of the information provided by each sensor separately. The advantages can be assessed in four aspects: Redundancy (increase in accuracy, reliability - usually at lower level of representation), complementarity (allowing new additional features to be perceived - usually at a higher/symbolic level of representation or without fusion), timeliness (faster sensing due to use of the fastest available sensor and parallel processing) and cost of information (systems are cheaper) [LK90]. Practical advantages are therefore economic benefits and a reduction of bandwidth for the data transfer.

These benefits can be achieved by using one sensor to guide another one's function (multisensor integration according to [LK90]), or by actually combining different sources of sensory information into one representational form (sensor fusion according to [LK90]). In this work we will concentrate on the later one.

Typical Applications of multi-sensor fusion are automatic target recognition, mobile robot navigation, industrial tasks like assembly, target tracking, or aircraft navigation.

Typical errors that can affect these systems occur 1) in the integration and fusion process,

2) in the sensory information, and 3) in the system operation.

It is difficult to formulate a general-purpose method for sensor fusion, due to the high diversity of sensor capabilities, information types, and system requirements. Instead, useful paradigms, frameworks and control structures from different fields, are commonly combined. This modularity 1) reduces complexity and increases flexibility in the system design, 2) enables distributed processing and hierarchical structures, and 3) allows an efficient representation.

The contributions of this work reside in

1. giving a brief overview about the basics of sensor fusion
2. selecting a mathematical and a bio-inspired method and presenting them,
3. compare the selected methods.

1.2 Basic Sensor Integration Functions

In figure 1.1 the basic integration functions and their relationship to each other are displayed. The *sensors* transfer their signals to the *sensor models*, which represent the uncertainty and error in the data and provide procedures to measure its quality. The probability distribution of the data error is often assumed to be Gaussian distribution, since it is physically plausible and easy to calculate. In the *sensor registration* the data sets are commensurated in both their spatial and temporal dimensions.

After registration, the signal is processed in the *sensory processing*. There it is decided whether 1) the data passes a *fusion* on the abstraction level of *signal level*, *pixel level*, *feature level*, or *symbol level* (s. fig. 1.2); whether 2) it is used for *separate operation*, influencing other sensors only indirectly; or whether 3) it is used for *guiding or cueing* other sensors based on the sensed data. The more abstract the level, the less registration is required and the easier it is to distribute sensors across platforms. Registration of a sensor and fusion of its data can be performed with other sensors or the world model.

The *world model* stores information about the environment of the machine, both a priori information and recently acquired sensory information. Using the results of the sensory processing a *sensor selection* selects the most appropriate sensors. In order to do that, sensor performance criteria had to be established. Approaches of sensor selection are “pre-selection” or “real-time selection”.

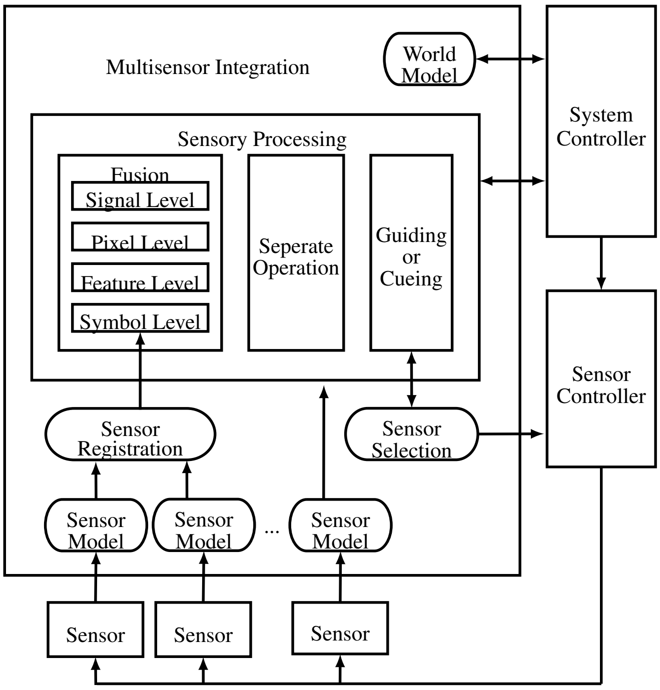

Figure 1.1: Functional diagram of multi-sensor integration and fusion in the operation of a system. The composition is inspired from [LK90].

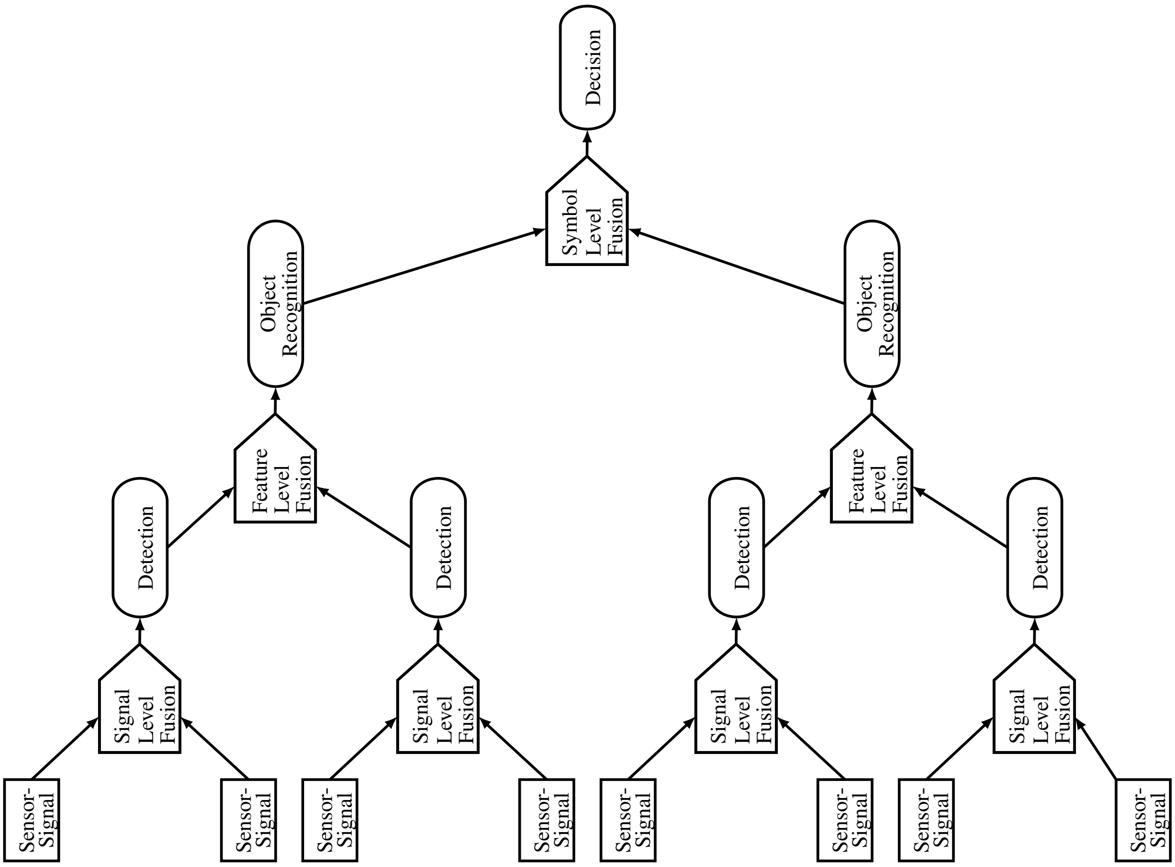

Figure 1.2: Sensor fusion on different abstraction levels and the connecting data processes. The pixel level is regarded equal with the signal level, and left out in this figure.

1.3 Choosing the Compared Methods

In the following chapter, two methods shall be compared with each other. One method is selected from the traditional, mathematical/statical methods and the other is a bio-inspired technique.

In an judgemental comparison of methods, both evaluated methods have to perform a similar task. As explained in the introduction, tasks of sensor fusion methods are to increase the reliability of the information or to derive information through fusion, that could not be gained without it. However, the kind of information dealt with, varies strongly depending on the abstraction level of the fusion method. Thus the methods, which shall be compared, have to be on a similar abstraction level. The abstraction level of the mathematical fusion method can easily be identified. The story is a different one for the bio-inspired method, but although the abstraction level of this cognitive fusion method is ambiguous, it can at least be roughly allocated.

An example for fusion is binaural localization - the biological way of determining a sound source. By calculating the correlation between the audio signals, the time difference of the salient audio signals can be determined. Taking this time delay into account, the source of the salient signal can be localized. Summarized, two auditory signals are fused and so the new information is won, in which radial direction the source is located.

While binaural localization is a bio-inspired fusion method on the signal-level that has the objective to gain new information, mathematical fusion methods on the signal-level commonly have the goal to increase reliability of the signal by the means of redundant information, as for example the Kalman filter. Therefore it is inapplicable to compare those methods with binaural localization.

In visual attention theory (s. [IKN98], [IK99]), sensors mark specific features in an image as interesting. The marked images, so called feature maps, are fused by a linear combination. This retina-inspired method can be allocated to the pixel- or feature-level and be compared with mathematical methods that also try to improve reliability, as for example by means of logical filters or mathematical morphology.

For this work, we are going to compare the mathematical fusion method “Dempster-Shafer Evidential Reasoning” with the cognitive fusion method “Multi-directional ARTMAP”. The main reasons are given next.

1. Goal of this work is to compare multi-modal methods. Multi-modality requires a high abstraction level, thus methods, capable of symbol-level were chosen.
2. The original frameworks of both methods have been well established and proven to be useful in various applications.
3. Both methods are state of the art and in further development.
4. Both methods can be used to increase the reliability of information.

Chapter 2

Analysis and Comparison of two Fusion Methods

2.1 Dempster-Shafer Evidential Reasoning

The Dempster-Shafer Theory ([Sha76]) is a generalization of the Bayesian theory. It is based on the idea of "*belief*" (bel) and "*plausibility*" (pls), which are the lower and the upper probability values of the "*belief interval*". The associated Dempster-Shafer Evidential Reasoning (DSER) can handle ignorance, which is only reduced if supporting evidence is at hand. Every type of ignorance is assigned a certain uncertainty. That way incomplete models or lack of prior information are represented adequately. Therefore less model knowledge about the world and the sensors is required. DSER being able to deal with problems about which little is known. Unlike Bayesian networks, DSER does not have to make assumptions that may not fit well with reality.

Typical applications of DSER are target tracking, gait analysis or fusing GPS-data and IMU-data for navigation [Fol12].

In DSER a "*frame of discernment*" (Θ) is defined as the set of mutually exclusive and exhaustive "*singletons*". A singleton is a hypothesis representing the lowest level of discernment.Imagine an object recognition task to recognize a robot. There exists a blue (B), a green (G) and a yellow (Y) robot, which make different sounds and broadcast unique electromagnetic signatures, thus three singletons. The "*power set* of Θ", denoted as 2^Θ, is composed of all the subsets of Θ, including Θ and the empty set $\varnothing$. Thus the magnitude of 2^Θ is $\|2^\Theta\| = 2^{number-of-singletons}$. In the robot example $\Theta = \{B, G, Y\}$ and the complete 2^Θ is therefore

$$2^\Theta = \Big\{\{B,G,Y\},\{B,G\},\{B,Y\},\{G,Y\},\{B\},\{G\},\{Y\},\varnothing\Big\}.$$

The terms of 2^Θ are called proposition (P) $\left\{P_j \mid P_j \in 2^\Theta\right\}$, for which a sensor i is able to provide direct information and maps P to a probability mass

$$m_i : \left\{P_j \mid P_j \in 2^\Theta\right\} \rightarrow [0,1] \text{ , while } \sum_{P_j \in 2^\Theta} m_i\left(P_j\right) \overset{!}{=} 1 \text{ and } m\left(\varnothing\right) \overset{!}{=} 0.$$

In opposite to conventional probability theory, DSER does not assign the complement of a proposition the opposite probability, meaning $m_i(P_j) = X \not\Rightarrow m_i(\neg P_j) = 1-X$ or $m_i(\neg P_j) \neq 1-m_i(P_j)$. Any probability mass not assigned a proposition P, is included in $m_i(\Theta)$. The belief, the plausibility and the belief interval of a proposition P of sensor i are defined as

$$bel_i(P) = \sum_{P_k \subseteq P} m_i\left(P_k\right) \qquad , \tag{2.1}$$

$$pls_i(P) = 1 - bel(\neg P) \qquad \text{and} \tag{2.2}$$

$$[bel_i(P), pls_i(P)] \qquad , \tag{2.3}$$

respectively. Belief is increased if supporting evidence is available, while plausibility is decreased if contradicting evidence is at hand.

"*Dempster's rule of combination*" fuses proposition P_a from sensor 1 with P_b from sensor 2 to P_c with the formula

$$m_{1,2}(P_c) = \frac{\displaystyle\sum_{P_a \cap P_b = P_c} m_1(P_a) \cdot m_2(P_b)}{1 - \displaystyle\sum_{\substack{P_k \cap P_l = \varnothing \\ \forall P_k, P_l \in 2^\Theta}} m_1(P_k) \cdot m_2(P_l)} \qquad , \tag{2.4}$$

where $P_c \neq \varnothing$, and where $m_{1,2}$ is the orthogonal sum $m_1 \oplus m_2$ and $P_a, P_b \in 2^\Theta$. The denominator normalizes the probability mass, so that their sum equals 1. This is necessary in case of a conflict in which an intersection of the propositions is impossible. If the *degree of conflict*

$$\kappa = \sum_{\substack{P_k \cap P_l = \varnothing \\ \forall P_k, P_l \in 2^\Theta}} m_1(P_k) \cdot m_2(P_l) \qquad \in [0;1] \tag{2.5}$$

of a fusion is low, the decision is more likely to be be accurate, than with a high κ.

In order to recognize the robots we have a visual (👁), an auditory (♬) and a electromagnetic (📡) sensor at our disposal. Let's assume the visual sensor picked up, that the robot is B and the visual sensor is reliable for 70%. The auditory sensor recorded, that it is not the G robot and we rely on it for 40%. The electromagnetic sensor suggests, that the Y robot is not encountered. This sensor is right in 50% of the cases, but also wrong in 20%. First the visual and the auditory sensor are fused, according to "Dempster's rule of combination" (eq. (2.4)). This can be vividly demonstrated with the table

	$m_{\text{👁}}(\{B\}) = 0.7$	$m_{\text{👁}}(\Theta) = 0.3$
$m_{\text{♬}}(\{B, Y\}) = 0.4$	$m_{\text{👁}\text{♬}}(\{B\}) = 0.4 \cdot 0.7 = 0.28$	$m_{\text{👁}\text{♬}}(\{B, Y\}) = 0.4 \cdot 0.3 = 0.12$,
$m_{\text{♬}}(\Theta) = 0.6$	$m_{\text{👁}\text{♬}}(\{B\}) = 0.6 \cdot 0.7 = 0.42$	$m_{\text{👁}\text{♬}}(\Theta) = 0.6 \cdot 0.3 = 0.18$

from which we calculate the beliefs (eq. (2.1)) to $bel(\{B, G\}) = bel(\{B\}) = 0.28 + 0.42 = 0.7$, $bel(\{B, Y\}) = 0.28 + 0.42 + 0.12 = 0.82$, $bel(\{G, Y\}) = bel(\{Y\}) = bel(\{G\}) = bel(\varnothing) =$

0 and $bel(\Theta) = 1$. Afterwards the fused data $m_{\text{👁♬}}(\{...\})$ is fused (eq. (2.4)) with the electromagnetic sensor, depicted in the table

	$m_{\text{📡}}(\{B, G\}) = 0.5$	$m_{\text{📡}}(\{Y\}) = 0.2$	$m_{\text{📡}}(\Theta) = 0.3$
$m_{\text{👁♬}}(\{B\}) = 0.7$	$m_{\text{👁♬📡}}(\{B\}) = \frac{0.35}{0.86}$	$m_{\text{👁♬📡}}(\emptyset) = 0.14$	$m_{\text{👁♬📡}}(\{B\}) = \frac{0.21}{0.86}$
$m_{\text{👁♬}}(\{B, Y\}) = 0.12$	$m_{\text{👁♬📡}}(\{B\}) = \frac{0.06}{0.86}$	$m_{\text{👁♬📡}}(\{Y\}) = \frac{0.024}{0.86}$	$m_{\text{👁♬📡}}(\{B, Y\}) = \frac{0.036}{0.86}$
$m_{\text{👁♬}}(\Theta) = 0.18$	$m_{\text{👁♬📡}}(\{B, G\}) = \frac{0.09}{0.86}$	$m_{\text{👁♬📡}}(\{Y\}) = \frac{0.036}{0.86}$	$m_{\text{👁♬📡}}(\Theta) = \frac{0.054}{0.86}$

.

Since $m_{\text{👁♬📡}}(\emptyset)$ has to equal 0, every entry had to be divided by $\left(1 - \sum m_{\text{👁♬📡}}(\emptyset)\right) = 0.86$. We receive the belief (eq. (2.1)) and plausibility (eq. (2.2)) values

$$bel(\{B\}) = \tfrac{0.35+0.06+0.21}{0.86} = 0.72093... \qquad pls(\{B\}) = 1 - \tfrac{0.024+0.036}{0.86} = 0.93023...$$

$$bel(\{G\}) = 0 \qquad pls(\{G\}) = 1 - \tfrac{1-(0.09+0.054)}{0.86} = 0.00465...$$

$$bel(\{Y\}) = \tfrac{0.024+0.036}{0.86} = 0.06976... \qquad pls(\{Y\}) = 1 - \tfrac{0.35+0.06+0.09+0.21}{0.86} = 0.17441...$$

and so the rounded belief intervals (eq. (2.3)) B: [72%; 93%], G: [0%; 0.5%], Y: [7%; 17%]. The low degree of conflict $\kappa = \sum m_{\text{👁♬📡}}(\emptyset) = 0.14$ suggests a high accuracy of the results. Thus we would assume, that the robot in question is of type B.

Benefits of DSER are first of all the increase of reliability of decisions.

2.2 Associative Memory

In the nervous system of insects, data fusion is an associative process [WW06]. Since the modalities differ in dimensions, the associative process is realized in a hierarchal architecture. In this hierarchy a single highest neuron indicates what kind of object is perceived. The network architecture encompasses all features encoded from different modalities and abstraction levels. Whether an object is perceived does not only depend on the bottom-up sensory information, but also on the top-down expectations, stored as classes of objects. Those may be very specific or rather general. A technical implementation of an associative memory are ARTMAPs, based on the Adaptive Resonance Theory.

a) Adaptive Resonance Theory

The Adaptive Resonance Theory (ART) [CG02] deals with the information processing and storage in the brain and led to various versions of ART networks. The goal of an ART network is to determine to which class an input pattern belongs to. An abstract representation is displayed in figure 2.1. F1 is called the “short-term memory” or “comparison field” (neurons represent attributes) and F2 the “long-term memory” or “recognition field” (neurons represent classes). The input pattern is saved in F1 and is compared to the weights of the synapses connecting F1 with F2. The more a set of weights matches the input pattern, the more the according corresponding neuron in F2 inhibits the output of the other neurons in F2. The similarity of the input pattern in F1 and the weights is determined by the signal value $T_j = \frac{|I \cap w_j|}{\alpha + |w_j|}$, where T_j is the signal value, I the input vector, w_j the weight vector of the j-th F2-neuron, and α the signal rule parameter. With each neuron in F2 representing one object class, the winning F2-neuron determines to which class the input pattern in F1 most likely belongs to and selected for a resonance test. The resonance test decides that an input pattern in F1 does indeed belong to the class of the winning F2-neuron, if the match criterion $\frac{I \cap w_j}{I} \geq p$ is met. The "vigilance parameter" p has the important role of determining how detailed (high p leads to many fine-grained categories) or how general (low p leads to few, more-general categories) the classes are (s. 2.3. d)).

If no F2-pattern is found, two unsupervised learning options are available: Either a new F2-neuron is assigned to represent a new class and the F1-pattern is stored in its corresponding weights or the most similar F2-pattern is updated. For the update two unsupervised learning principles are available: In slow learning, the weights of the winning F2-neuron are updated with a small amount. In fast learning method is to let the input pattern just fall within the memory node's boundaries according to $w_j^{new} = I \cap w_j^{old}$.

b) Multi-directional ARTMAP

Building upon the ART network, so called ARTMAPs [CGR91] have been developed, that feature supervised learning, many-to-one and one-to-many learning. They use a second ART network and a "map field" in order to teach the first ART network correct patterns. Also they are able to compare the results of different ART networks to each other and draw conclusions from the cumulative pattern estimations.

One of the most advanced ARTMAPs is the Multi-directional ARTMAP (MdARTMAP) [Sch10], which features many-to-many learning. In it ART networks are hierarchically linked by associative learning. The concept is displayed in figure 2.3. For sensor fusion each ART network is assigned to exactly one modality. The multiple ART networks are connected via one "*map field memory*" F3, which contains the map field neurons that represent the classes for object recognition. The F2-neurons of the ART networks are connected with the F3-neurons in the map field, suggesting to which class the object belongs to. The connection between a F2-neuron and a F3-neuron are strengthened, when at least two ART networks recognize their input pattern as the same object class. The connections strengthen according to the Hebbian learning rule. On the decision whether and which object is perceived, ART network classes, which are strongly connected with each other, give a higher activation value to the respective F3-neuron, than weakly connected classes. However, more weakly connected ART network classes have always a greater influence than fewer strongly connected ones (see example at [Sch10, page 36,37]).

The most difficult task of unsupervised learning is to find the optimal vigilance parameters p, which is done in an empirical process. Since different modalities differ in resolution and variance and thus reliability, the vigilance parameters have to be tuned accordingly. Additionally the ART networks can be ranked and steer each other's search.

Benefits: MdARTMAP is able to do both increase the reliable of decisions and generate knowledge. An orange object (visual sensor 👁), makes a specific sound (auditory sensor ♫). A signal port is the third modality. Just with the visual and the auditory sensors, a robot is able to tell whether it can receive a status update from the object.

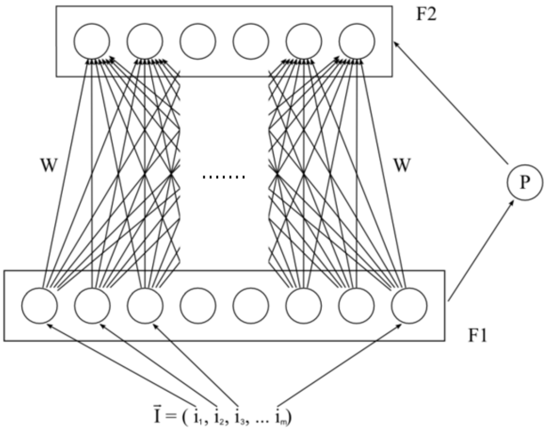

Figure 2.1: ART network.

$$Act(X) = |W| + \frac{\sum_{i}^{|W|} w_i}{|ART\ nets|} \tag{2.6}$$

$$\begin{matrix} winning \\ node \end{matrix} = \underset{X}{\operatorname{argmax}}(Act(X)) \tag{2.7}$$

Figure 2.2: Map Field node activation function (eq. (2.6)) and map field winning node selection (eq. (2.7)). W is a vector with all active connections to map field node X. $w_i \in [0, 1]$ is the normalized connection strenght between the ART networks *ART nets* and the map field nodes.

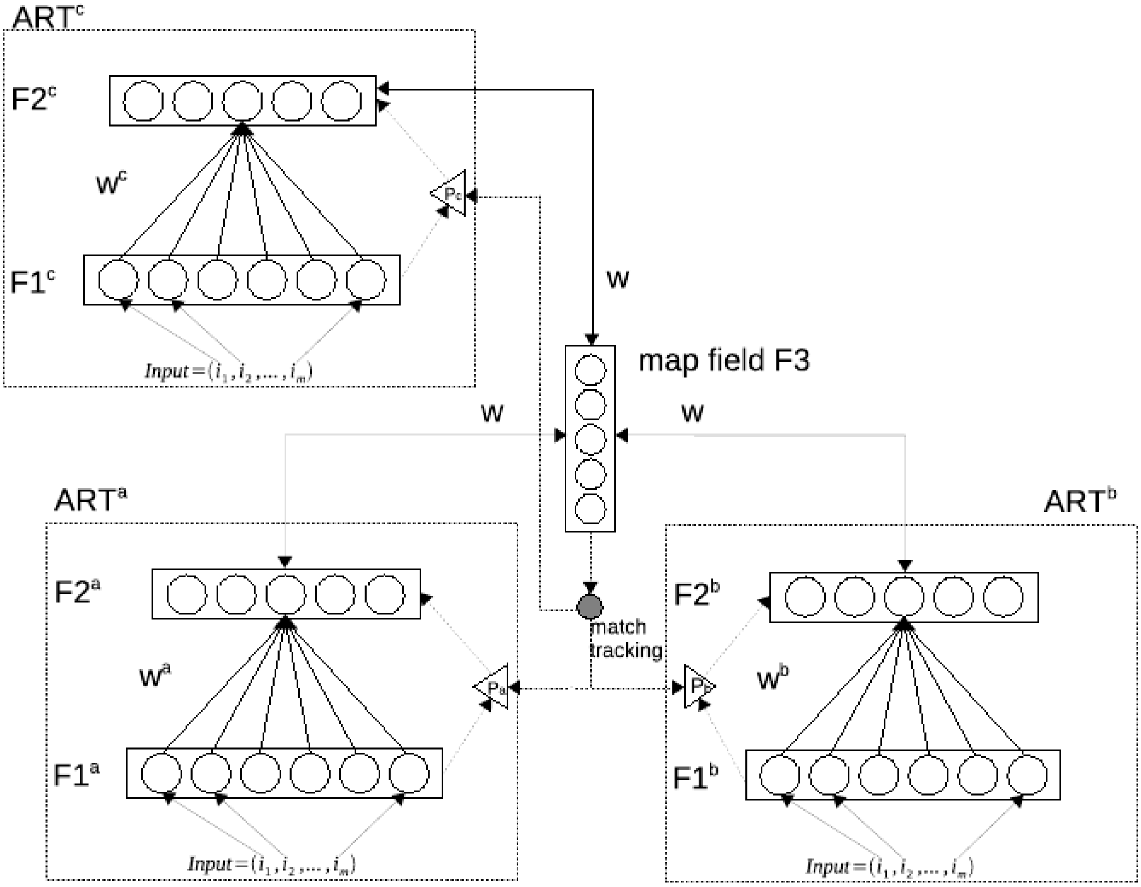

Figure 2.3: Multi-directional ARTMAP, taken from [Sch10]. All F2-neurons are connected to all F3-neurons. When input patterns (e.g. B, G, Y for visual modality) are presented to the ART networks (e.g. ART$^{a}\sim$ 👁, ART$^{b}\sim$♬, ART$^{c}\sim$📱), their output classes F2, activate the associated neurons in the map field (associative network - eq. (2.6)). The F3-neuron with the higest activation value (eq. (2.7)) is selected as winner.

2.3 Comparison

DSER and MdARTMAP have advantages and disadvantages, some of which set them apart from each other and some of which they share, despite their very different way of functioning. Their most important advantages are their capabilities to deal with uncertainty and that they can be used on different abstraction levels. They differ on what they depend on to be reliable. Reliability is the most important topic of sensor fusion, since it is one of the main reasons for fusion in the first place.

a) Abstraction level

DSER has been used on pixel, feature, and symbol level, but only one level at a time, deciding what kind of choice to trust on that particular abstraction level. For instance, for assessing the structural integrity of materials, DSER outperformed nearly all other methods in fusing eddy current and infrared thermo-graphic data on pixel level [GBT99, table 2]. Even so, since DSER deals with hypotheses, it is used on symbol level in most examples of literature, e.g. [Hae05]. Several examples are given in this work.

In contrast to that, MdARTMAP is able to cover several abstraction levels at the same time. For instance, the input pattern of an ART network could be a picture, where each pixel (pixel level) is feed into one F1-neuron. Depending on the architecture of the system, the output of the MdARTMAP could be the decision what kind of feature (feature level) or what kind of object class (symbol level) is at hand.

b) Dealing with uncertainty and noise

The most mentioned advantage of DSER is its effectiveness in dealing with a lack of knowledge.

DSER is able to *represent uncertainties*: First, it is able to represent ignorance (the lack of evidence), in contrast to contradictory information (negative evidence) ([Sar00]). It does not assign unassigned probability to the opposite hypothesis, which would affect other hypotheses without evidence [GP96]. Second, DSER can assign probabilities to both hypotheses corresponding to single classes and propositions (multiple hypotheses) corresponding to unions of classes [Fol12], [GP96]. Third, probabilities are expected within “belief interval” and not set to a specific value [Kle04].

DSER is able to *accept incomplete models*: Not all probabilities and likelihood functions have to be known a priori [Kle04]. If not available, information is not assumed that might not fit with the real data later . By using propositions instead of hypotheses, it can handle imprecise models, but it can not handle a complete absence of them - some knowledge about sensory behavior is required.

While DSER uses different types of ignorance, ARTMAPs generalize on basis of training data sets. Input patterns similar to those sets are recognized as such by the ART networks. The uncertainty between training data and input data is dealt with [GRG+00].

Unlike DSER, ARTMAPs are able to handle a complete lack of a model, since they learn everything from data during operation.

Both DSER and MdARTMAP are able to deal with *ignorance regarding classes* - solution: multi-modality. In the case of two types of red robots, a color sensor can not decide which red robot is at hand. A fusion with sensors of different modalities provides clarity.

Also both methods are able to deal with *ignorance regarding one sensor* - solution: taking sensor limitations into account. For example a color sensor might not be able to tell a red and a pink robot apart, but it is sure not to sense a blue one. To deal with this, DSER assigned evidence to the proposition {'Red','Pink'}, but none to {'Red'} or {'Pink'}. Other than MdARTMAP, which generalizes the color "red" and feeds this more general color as feature to the memory field F3. One might say, DSER deals with this in a "crisp way" (here on a discrete symbol-level), and MdARTMAP in a "soft way" (here on a continuous signal-level).

c) Reliability

Both DSER and MdARTMAP give accurate results even when the information is scarce, which is a very big advantage - see part b) . Still, the reliability of DSER and MdARTMAP depends on certain parameters and key factors:

As mentioned, DSER requires at least some information about underlaying likelihood functions, though less than other methods. The more model-knowledge is available, the more accurate the predictions. In summary, *DSER is model dependent.*

In contrast, MdARTMAP requires data samples for the training. The more representative data samples are available, the more accurate the predictions. In summary, *MdARTMAP is data dependent*, regarding the quantity and quality of the samples.

A very important key factor for DSER's accuracy the magnitude of the degree of conflict κ. To illustrate: Two detection systems are used to detect an interesting object's position (left 'L', right 'R', or center 'C'): a visual saliency map (s. [IKN98], [IK99]) (👁) and an auditory saliency map (s. [KPLL05]) combined with binaural localization (♬). In a first case the systems consider the salient object to be left with $m_{\text{👁}}(\{'L'\}) = 0.1$, in the center with $m_{\text{👁}}(\{'C'\}) = 0.9$ and $m_{\text{♬}}(\{'C'\}) = 0.8$, or right with $m_{\text{♬}}(\{'R'\}) = 0.2$. After fusion and normalization (eq. (2.4)), the object is believed (eq. (2.1)) to be in the center with $bel(\{'C'\}) = 1$ and $\kappa = 0.28$ (eq. (2.5)). In a second case the masses $m_{\text{👁}}(\{'L'\}) = 0.9$, $m_{\text{👁}}(\{'C'\}) = 0.1$, $m_{\text{♬}}(\{'C'\}) = 0.2$, and $m_{\text{♬}}(\{'R'\}) = 0.8$ are assigned, resulting in the surprising belief $bel(\{'C'\}) = 1$ and the high $\kappa = 0.98$. This example[1] shows that a high degree of conflict results in a strong belief in a unlikely hypothesis and demonstrates, that DSER depends on a low κ in order to produce reliable results. We conclude that the accuracy of *DSER depends on the evidence* provided by the sensors.

[1] This type of example is also known as "Zadeh's paradox". With it, Zadeh criticized DSER to produce counterintuitive and erroneous results (e.g. in [Zad86]), which is DSER's biggest criticism ([Fol12]). It is worth mentioning that several authors regard Zadeh's criticism as unjustified, e.g. [Xio08], [Hae05].

The parameter with a big influence on MdARTMAP's reliability is the vigilance parameter (p). Whereas κ should be minimized in DSER, p is a trade off between reliability and informative value. With too general classes (low p), the correct class is chosen, but its expressiveness is very low. With too fine-grained classes (high p), the decision is not reliable. In the above example, a very low p might result only in the two classes 'left' and 'right'. If the robot had to follow the detected object, when facing it, the robot would not go straight forward, but sway left and right. A very high p might lead to the classes 'very left', 'left', 'center-left', 'center', etc. and result in unnecessary computational expense. Finding the optimal p and assessing it is the most difficult task of design (s. d)).

d) Implementation: Adjustment, Computational Expenses, Applications

In order to gain the underlaying likelihood functions for DSER, experiments can be made previously. Often Gaussian error distributions are assumed (s. 1.2). Thus two parameters, namely variance and expected value, have to be determined experimentally per likelihood function. In MdARTMAP the vigilance parameters have to be determined through experiments - one per ART network. While the Gaussian parameters are to be determined before operation, the ps are determine during operation. While the determination of the Gaussian parameters can be rather straightforward, determining the vigilance parameters is one of the most difficult tasks in designing a MdARTMAP [Sch10].

Due to its complexity, DSER is comparably slow, that is why e.g. Buede and Girardi [BG97], prefer simpler methods like Bayesian theory. DSER gets exponentially slower to the amount of singletons, since the amount of propositions rises exponentially [Sar00].

MdARTMAP is especially slow during training, fast during operation. An increase of neurons in ART networks increases their number of weights exponentially and thus its computational expenses. An increase of modalities increases the memory field F3's work load only linear, but the calculation of new vigilance parameters are expensive (s. c)). With time, the amount of classes in F3 is expected to rise logarithmic in an unsupervised setting, and with it the computational cost. ARTMAPs are very suitable for parallelization. On appropriate hardware, ARTMAPs can provide extremely fast and fault tolerant processing [GRG+00].

For DSER, applications include detection of signals, objects, even intrusions in computer systems, and pattern, object, or target recognition [Fol12]. MdARTMAP was developed for the swarm mirco -robots in the Replicator Project, but applications for MdARTMAP may be similar.

2.4 Summary and Conclusion

The goals of sensor fusion were stated and an introduction given. The two methods DSER and MdARTMAP were chosen to be explained in more detail and compared. Finally some applications were mentioned. We conclude that no method is by default better and to choose, the details of the problem and the methods, as outlines, have to be considered.

Bibliography

[BG97] D.M. Buede and P. Girardi. A target identification comparison of bayesian and dempster-shafer multisensor fusion. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 27(5):569 –577, Sep. 1997.

[CS02] Conradt, J., Simon, P., Pescatore, M., and Verschure, PFMJ. (2002). Saliency Maps Operating on Stereo Images Detect Landmarks and their Distance, Int. Conference on Artificial Neural Networks (ICANN2002), p. 795-800, Madrid, Spain.

[CG02] G.A. Carpenter and S. Grossberg. Adaptive resonance theory. 2002.

[CGR91] G.A. Carpenter, S. Grossberg, and J.H. Reynolds. Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, (4):565–588, 1991.

[Fol12] Bethany G. Foley. A dempster-shafer method for multi-sensor fusion, Mar. 2012.

[GBT99] X.E. Gros, J. Bousigue, and K. Takahashi. NDT data fusion at pixel level. *NDT&E International*, 32(5):283 – 292, 1999.

[GP96] Harald Ganster and Axel Pinz. Active fusion using dempster-shafer theory of evidence. 1996.

[GRG+00] Eric Granger, Mark A. Rubin, Stephen Grossberg, Pierre Lavoie, and Bethany G. Foley. A what-and-where fusion neural network for recognition and tracking of multiple radar emitters. Dec. 2000.

[Hae05] Rolf Haenni. Shedding new light on zadeh's criticism of dempster's rule of combination. In *8th International Conference on Information Fusion*, volume 2, page 6 pp., Jul. 2005.

[IK99] Laurent Itti and Christof Koch. A comparison of feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, 10:161–169, 1999.

[IKN98] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. volume 20, 1998.

[Kle04] Lawrence A. Klein. *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*. SPIE Press, 2004.

[KPLL05] Christoph Kayser, Christopher I. Petkov, Michael Lippert, and Nikos K. Logothetis. Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biologyg*, 15:1943–1947, Nov. 2005.

[LK90] R.C. Luo and M.G. Kay. A tutorial on multisensor integration and fusion. In *Industrial Electronics Society, 1990. IECON '90., 16th Annual Conference of IEEE*, volume 1, pages 707 – 722, Nov. 1990.

[Sar00] M. Sarkar. Modular pattern classifiers: a brief survey. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 4, pages 2878 –2883 vol.4, 2000.

[Sch10] T. P. Schmidt. Robotics: Environmental awareness through cognitive sensor fusion, Mar. 2010.

[Sha76] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[WW06] J. Wessnitzer and B. Webb. Multimodal sensory integration in insects: towards insect brain control architectures. volume 1, 2006.

[Xio08] Wei Xiong. Analyzing a paradox in dempster-shafer theory. In *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on*, volume 5, pages 154 –158, Oct. 2008.

[Zad86] L. Zadeh. A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination. volume 7, Summer 1986.