Sensor Management for Identity Fusion on a Mobile Robot

Thomas Dall Larsen, Nils A. Andersen, & Ole Ravn
Department of Automation, Technical University of Denmark
Building 326, DK-2800 Lyngby, Denmark
e-mail: {tdl, naa, or}@iau.dtu.dk

Abstract

This paper addresses the problem of identity fusion, i.e. the problem of selecting one of several identity hypotheses concerning an observed object. Two problems are considered. Firstly the problem of representing and fusing measurements relating to identity hypotheses is treated. Here it is shown that if the hypotheses are chosen to be concerned with object properties rather than object identity, Bayes' combination rule outperforms Dempster-Shafers. Further some points are made concerning the numerical properties of the normalization in Dempster-Shafers combination rule, and of the use of a priori knowledge in identity fusion. Secondly the problem of selecting the most appropriate sensing action is addressed. A number of methods from the literature is described and compared, and a simple and computationally cheap metric is proposed and shown to be very efficient.

1 The Identity Fusion Problem

Consider a mobile robot navigating in an uncertain environment that needs to be able to identify certain objects in its path. Specifically it needs to be able to determine if an object in front of it is a door. The frame of discernment for this problem can be chosen as:

$$\Theta = \{H_1, H_2, H_3, H_4, H_5\}, \tag{1}$$

where the mutually exclusive and exhaustive set of hypotheses are:

H_1: The object is a closed door
H_2: The object is an open door
H_3: The object is a window
H_4: The object is a person
H_5: The object is something else

For this, the robot is equipped with a suite of sensors. Using these, the robot can then perform a sequence of measurements each stating the identity of the object and fuse these using for instance Bayes' or Dempster-Shafers rules. To obtain identity measurements like $z_s = H_1$, however, the sensor data will need to be preprocessed. In this processing some sensory information can be lost. The sensor data might be more ambiguous or uncertain but then be forced to choose a favorite hypothesis to provide a nice input to the identity fusion.

2 Using Object Properties

Instead of letting measurements declare identity, more information can be preserved in these by relating them to object properties. A number of properties that separates a door from its surroundings could for instance be:

z_1: Squareness (a door is a fairly square object)
z_2: Color separation with wall
z_3: Depth (especially when open)
z_4: Dimensions (within certain min-max limits)
z_5: Proximity to floor (unlike windows)
z_6: Inanimate (unlike persons)

The five objects the robot must be able to discern, all match these six characteristic properties differently. If therefore the measurements focus on these, they will not have to determine identity - only look for properties. The great advantage being that these are a lot closer to what can be sensed. A sonar for instance sees depth, and different vision algorithms can recognize one or more features using a variety of methods (like for instance intensity diagrams or edge detectors). The different features can then be fused to determine the identity of the object. In the sequel a sensing action, z_i, will denote a sensor and an algorithm combination that investigates the corresponding property defined above.

2.1 Bayesian Fusion

In order to fuse measurements using Bayes' rule, it is necessary to know or estimate the conditional probability distributions $P(z|\Theta)$ for the sensing actions, that is, the probabilities that the properties are observed, provided the different hypotheses are true. These distributions can be found by moving the mobile robot around and perform a large number of sensing actions and observe how often the different sensing actions identifies the features of the objects discerned by Θ. Such an experiment will result in a set of distributions[1] as the one

[1] The array of probabilities, $P(z_i|\Theta)$, are here referred to as probability distributions though they are not normalized. Strictly only $\frac{P(z|\Theta)}{\sum_k P(z|P(H_k))}$ are distributions.

shown in table 1[2].

$P(z_i\|H_j)$	H_1	H_2	H_3	H_4	H_5
z_1	0.95	0.95	0.95	0.20	0.30
z_2	0.95	0.70	0.60	0.70	0.50
z_3	0.10	0.98	0.30	0.20	0.20
z_4	0.95	0.95	0.10	0.10	0.10
z_5	0.90	0.90	0.10	0.90	0.90
z_6	0.10	0.10	0.10	0.80	0.20

Table 1: Conditional probability distributions.

Given an a priori probability distribution $P(\Theta)$ and an observation that the object has property z_i, the resulting a posteriori distribution can be obtained from Bayes' rule:

$$P(\Theta|z_i) = \frac{P(z_i|\Theta)P(\Theta)}{\sum_k P(z_i|P(H_k))P(H_k)} \tag{2}$$

When using Bayesian fusion it is necessary to know *all* the conditional probabilities $P(z_i|H_j)$. It is not possible for a sensor to express evidence for just a few of the hypotheses and ignore the rest, each hypothesis must be assigned a probability although the sensor may be totally ignorant of the truth of these. This inadequate representation of uncertainty is one of the main criticisms of Bayesian fusion and gave birth to the Dempster-Shafer method.

2.2 Dempster-Shafer Fusion

When Dempster-Shafer reasoning is used, the sensors assign probability masses to *propositions*, i.e. sets of hypotheses the particular sensor is unable to distinguish between. For the frame of discernment Θ, there exists $2^{|\Theta|}$ such propositions, why this set is referred to as 2^{Θ}.

For instance the sensor action, z_1, investigating the squareness property cannot distinguish doors from windows. If a square property is observed, a proposition consisting of a disjunction of these objects will therefore be assigned one common mass. Using the values in table 1, the mass distribution, $m_1(\Theta)$, from sensor action z_1 is calculated:

$$m_1(A_1) = \frac{\sum_{H_j \in A_1} P(z_1|H_j)}{\sum_j P(z_1|H_j)} = 0.8507$$

$$m_1(H_4) = \frac{P(z_1|H_4)}{\sum_j P(z_1|H_j)} = 0.0597$$

$$m_1(H_5) = \frac{P(z_1|H_5)}{\sum_j P(z_1|H_j)} = 0.0896$$

where: $A_1 = \{H_1, H_2, H_3\}$.
The masses from the different sensors are combined using Dempster-Shafers rule of combination:

$$m(A) = K \sum_{A_1 \cap A_2 = A} m_1(A_1)m_2(A_2) \tag{3}$$

$$K^{-1} = 1 - \sum_{A_1 \cap A_2 = \emptyset} m_1(A_1)m_2(A_2) \tag{4}$$

The *support* of a proposition, A_i, in a given mass distribution, $m(\Theta)$, can be found from:

$$S(A_i) = \sum_{A_i \in A} m(A) \tag{5}$$

By combining the support with the *plausibility* i.e. the lack of support for its negation $(1 - S(\neg A))$ an evidential interval can be formed which shows the amount of data that supports and contradicts the proposition. In Dempster-Shafer theory the evidential intervals are the outputs and in general it is not possible to extract probabilities from a mass distribution (unless no ambiguity is present). When a single probability-like measure is needed nontheless, the ambiguous mass can be distributed equally over the relevant hypotheses. Probability is then defined by:

$$P(A_i) = \frac{\sum_{A_i \in A} m(A)}{|A|} \tag{6}$$

2.2.1 Some Numerical Properties

In the literature Dempster-Shafer fusion is often described using very simple examples where only a few mass distributions are fused. When the method is used recursively with a wide variety of sensors some interesting problems surface which are unreported in the literature. Firstly, the normalization of the masses in equation (4) uses the new conflicting evidence to ensure that $\sum_{A \in 2^{\Theta}} m(A) = 1$. The first time two distributions are combined this normalization will have an accuracy of a few times the machine precision. At every subsequent fusion the uncertainty will grow in an exponential manner as all the factors in equation (3)-(4) are imprecise. The growth rate of the error naturally depends on the number of factors in the two summations, but eventually the sum of the probability mass will deviate seriously from unity. Here simulations in Matlab showed that the uncertainty was more than doubled after each fusion, meaning that errors starting off at machine accuracy (10^{-16}) after 30 fusions had reached a level of several percents (after 40 fusions the mass distribution was sheer noise). This serious problem can be fixed by adding an additional normalization after equation (4):

$$m(\Theta) = \frac{m(\Theta)}{\sum_{A \in 2^{\Theta}} m(A)} \tag{7}$$

Another point that is important to make is that in the mass assignment in equation (3), propositions that do not occur as subsets of propositions with non-zero masses in both distributions, $m_1(\Theta)$ and $m_2(\Theta)$, will be assigned a zero mass. If one of the mass distributions are non-zero only for the singletons $(H_1, H_2, \ldots, H_N)$ all other masses in $m(\Theta)$ but the singleton masses will become zero. Once this has happened it is easy to see that

[2] Note that the conditional probabilities in table 1 are estimated and thus *not* found experimentally.

from this point on only singletons will have non-zero masses. With the variety of distributions for the different sensors a lot of masses will quickly become (and stay) zero. If this is not considered in the implementation a lot of computation will be wasted.

2.3 Dempster-Shafer vs Bayes

Often it is claimed that an important difference between the Dempster-Shafer fusion method and the Bayesian one, is the treatment of a priori knowledge. It is argued that Bayesian fusion can be performed if and only if the a priori distribution has been determined and that Dempster-Shafer fusion in contrast does not require and is not able to utilize a priori knowledge. Consequently, when comparisons are made, the Bayesian method is given a priori knowledge and the Dempster-Shafer method not. Often then, the presence of a priori knowledge in the Bayesian fusion is what really makes the results from the two methods differ and what has been compared is therefore not Dempster-Shafer vs Bayes but rather whether a priori knowledge should be used. To overcome this problem, firstly the use of a priori knowledge will be examined, and then Dempster-Shafer and Bayesian fusion will be compared without using a prioris.

2.3.1 Using A Priori Knowledge

If the relative occurencies of the objects (the *a priori* probability distribution) are known, this can be used as the initial probability distribution, the first incoming measurement will be fused with. If for instance the a prioris are known to be:

$$P(\Theta) = [0.60\ 0.10\ 0.20\ 0.05\ 0.05], \tag{8}$$

and a sensor reports equal evidence for H_1 and H_4, the fact that H_1 is known to be more probable makes this the preferred hypothesis. In contrast if the Bayesian method is started of with a non-informative prior (i.e. the a priori probability distribution is uniform), the first posterior distribution derived from fusing the first measurement will equal the (normalized) conditional of this. If the Bayesian method therefore is used with a non-informative prior it considers only evidence from measurements. Likewise, an a priori distribution can be fused in the Dempster-Shafer method as any other distribution. Both methods can therefore use or ignore a priori knowledge.

The impact of using a priori knowledge will now be tested in simulations. Five different objects are simulated to show properties with probabilities according to table 1. Sensing actions are then performed until a hypothesis has reached a probability of more than 95% (how to choose the sensor actions will be treated in section 3 - in this section the sensing action that differentiates the best between the two most likely hypotheses are chosen). A total of 20.000 identifications are made, half of them using a non-informative prior and the other half using the priori distribution in (8). The results are shown in table 2. From the table it can be seen that

Use prior	Object of type 1	2	3	4	5
No	5.8	5.5	5.7	7.9	10.9
	(98%)	(98%)	(97%)	(97%)	(97%)
Yes	5.7	10.2	9.1	15.6	22.0
	(100%)	(96%)	(92%)	(91%)	(91%)

Table 2: Bayesian fusion with/without a prioris. The numbers denote the average number of actions before the identification converge. The numbers in parentheses are the percentage of successful identifications.

the use of a priori knowledge increases the convergence speed and percentage for the likely hypotheses and conversely decreases them for hypotheses with low prior probability. The sensory evidence has to "fight" the a prioris, and as can be seen from the table it does not always succeed (the erroneous identifications is predominantly objects being taken for the most likely object - a closed door).

Clearly, most object types converge much better with a non-informative prior, when the prior distribution is as in (8). But as it is the most common objects that benefits the most from using prior information, a comparison should be made of the average number of actions and correct identifications for all the objects (table 3). The

Use prior	Average number of Actions	Success
No	6.1	97.8%
Yes	8.2	97.2%

Table 3: Bayesian fusion with/without a prioris.

percentage of identifications that converge to the correct object are about the same whether or not a priori knowledge is used, but the total number of actions used to identify the 10.000 objects increases by 30% when a priori knowledge are used. Using a priori knowledge therefore has the unfortunate property that more sensory evidence is required. Therefore it is chosen *not* to use a priori knowledge in the remainder of this paper.

2.3.2 Comparison of Dempster-Shafers and Bayes' combination rules

If all conditional probabilities are known precisely (as in table 1) nothing can be gained by using the Dempster-Shafer method. However, due to its better representation of uncertainty, Dempster-Shafer might outperform Bayes if the statistics are not known precisely. Consider the case where all the conditionals are not known, but instead only incomplete knowledge such as $P(z_i|H1 \vee H_2)$ is possessed. In the Bayesian framework this cannot be represented accurately and it has to be assumed that the conditional probabilities for the disjunction of hypotheses equals that of the individual hypotheses (as in table 1). The true probabilities could instead be as in table 4. As the mass distributions calculated

$P(z_i\|H_j)$	H_1	H_2	H_3	H_4	H_5
z_1	0.98	0.98	0.89	0.20	0.30
z_2	0.95	0.75	0.60	0.65	0.50
z_3	0.10	0.98	0.30	0.25	0.15
z_4	1.00	0.90	0.05	0.05	0.20
z_5	0.85	0.95	0.10	0.88	0.92
z_6	0.08	0.12	0.10	0.80	0.20

Table 4: True conditional probability distributions.

from table 1 also can represent the conditional probability distribution in table 4 the Dempster-Shafer framework, models the (limited) knowledge precisely and does not introduce any additional errors solely due to the representation.

The two fusion methods are now tested in simulations. Five different object types are simulated to show properties with probabilities according to table 4 but modelled (incorrectly) by table 1 for the Bayesian method and (correctly) for the Dempster-Shafer method. The stop criterion is that a hypothesis has reached a probability of more than 95%. For each object and fusion method a thousand simulations were made. The results are shown in table 5.

Fusion Method	Object of type 1	2	3	4	5
Bayes	5.5	6.1	5.9	6.5	10.7
	(99%)	(97%)	(98%)	(98%)	(92%)
DS	6.3	5.6	9.2	8.8	14.0
	(99%)	(99%)	(96%)	(99%)	(93%)

Table 5: Dempster-Shafer vs Bayesian fusion.

From table 5 it is observed that all success ratios are above 95% except for objects of type five. Hence, persons, windows and open or closed doors are all observable objects using the six available sensing actions, while the more loosely defined object, O_5, occasionally is falsely identified. By averaging the convergence rate, success ratio, and the computational complexity for the five objects, table 6 emerges.

Fusion Method	Average number of Actions	Flops	Success
Bayes	6.9	$2 \cdot 10^2$	97%
DS	8.8	$2 \cdot 10^4$	97%

Table 6: Dempster-Shafer vs Bayesian fusion (averaged for the five object types).

From table 6 it follows that the Dempster-Shafer method is considerably more computationally demanding[3] than the suggested Bayesian fusion and that it actually performs worse despite the fact that it models the sensor readings more accurately. It can of course not be ruled out that in certain conditions the Dempster-Shafer method will perform better. However, the more accurately the sensors are modelled, the better the Bayesian fusion will be and the less the motivation for experimenting with Dempster-Shafer fusion will be.

[3] No attempts has been made to optimize the Dempster-Shafer code though. The mass distributions contain all possible propositions, i.e. 32 ($= 2^{|\Theta|}$) masses, which requires a lot of computation.

3 Sensor Management

The sensor management problem is here defined as the problem of choosing sensor actions among the six possible to obtain a quick identification of the object the mobile robot is confronted with. When a probability distribution for the five hypotheses and conditional probability distributions for each of the sensing actions (as defined in table 1) are known, the a posterioris resulting from each of the measurements can be calculated without actually performing the measurement. Choosing the optimal sensor action can then be reduced to finding a metric for, which of the a posteriori distributions is the most desirable - i.e. finding a utility metric. Three such metrics will be considered.

3.1 Minimising Entropy (E)

As the purpose of sensing is to gather information, an obvious utility function would be to use the *information* content of the posterior distributions. Information can be defined in various ways, but for discrete distributions, it is often defined as the opposite of entropy. The entropy of a distribution is here (as in [1]) defined by:

$$\begin{aligned} h(P(\Theta)) &= E\{-\ln(P(\Theta))\} && (9)\\ &= -\sum P(\Theta)\ln(P(\Theta)), && (10) \end{aligned}$$

The sensor action, z_i^*, that yields the posterior distribution wich minimizes the entropy (or conversely maximizes the information content) can then be chosen as the optimal action:

$$z_i^* = \arg\max_{z_i}(h(P(\Theta))) \quad (11)$$

3.2 Discrimination Gain (DG)

In [2] different methods for allocating a single sensor to investigate different areas (cells) for the presence of a target is compared in simulations. A method developed by the author that optimizes the discrimination between the a priori and the a posteriori probabilities is found to be superior. The discrimination (or cross-entropy), D, of the a posteriori probability distribution $P(\Theta|z_i)$ after a measurement z_i given the a priori distribution $P(\Theta)$ is defined by:

$$D(z_i) = \sum_j P(H_j|z_i)\ln\left(\frac{P(H_j|z_i)}{P(H_j)}\right) \quad (12)$$

The expected discrimination after a measurement z_i with possible outcomes $\{Z\}$ is given by:

$$E\{D|z_i\} = \sum_{z_i \in \{Z\}} D(z_i) P(z_i | P(\Theta)) \quad (13)$$

The probability of getting the measurement z_i is:

$$P(z_i | P(\Theta)) = \sum_j P(z_i | P(H_j)) P(H_j) \quad (14)$$

The discrimination gain method selects the sensor action that maximizes (12) and thereby ensures that the a posteriori distribution will differ as much as possible from the a priori, meaning that as much new information as possible has been added.

3.3 Distincting the Two Likeliest (D)

Though minimizing the discrimination gain maximizes the total information gain, it is not necessarily optimal as some information can be more valuable than other. This distinction between relevant and less relevant information of course makes the framework less rigorous but if the potential benefit is a better performance this could be a price worth paying. When doing identity fusion the goal of sensing is to find the most probable hypothesis in a discrete set of hypotheses. An obvious approach would therefore simply be to select the sensing action that discriminates the two most likely hypotheses j and k the best. This can be done without calculating the a posterioris simply by finding the sensor action z_i^* where the difference between $P(z_i|P(H_j))$ and $P(z_i|P(H_k))$ is the biggest:

$$z_i^* = \arg\max_{z_i} \left(|P(z_i|P(H_j)) - P(z_i|P(H_k))| \right) \quad (15)$$

3.4 Comparison of Information Metrics

The three management methods are now tested and compared in Monte Carlo simulations. A robot seeking information about an object chooses the next sensor action, z_i, by:

1. optimizing discrimination gain (DG)
2. distincting two most likely hypotheses (D)
3. minimizing entropy (E)
4. random (RND)

When entropy is minimized it is necessary to know the posterior distribution after fusing the different measurements. As the outcome of the measurement is unknown (e.g. is the object square or not) the *expected* posteriors are found by assuming the most likely hypothesis is true. Simulations are now performed using Bayesian fusion with a non-informative prior and the same objects and stop criterion as in section 2.3. The results are as shown in table 7.

Man.	Average number of		
Method	Actions	Flops	Success
DG	8.6	726	98%
D	7.1	12	97%
E	7.0	260	98%
RND	16.2	1	97%

Table 7: Comparison of management methods.

From the table it follows that choosing the sensor actions intelligently either by optimizing the discrimination gain (*DG*), the entropy (*E*) or the difference in conditional probabilities between the two most likely hypothesis (*D*), reduces the number of necessary sensor actions required to identify an object reliably by a factor of two. This alone of course promotes some effort to be made in choosing appropriate utility metrics. The three information based metrics leads to quite similar performances but as the *D* method is significantly cheaper computationally, this must be the test winner.

4 Conclusions

A comparison between Bayes' and Dempster-Shafers rule of combination showed that Bayes' rule leads to a faster convergence, is more reliable and is less computationally demanding. Dempster-Shafer theory still provides a better representation of ignorance but in the example with ambiguousness studied here, Bayesian fusion still outperformed Dempster-Shafer fusion. Further, it has been shown that simply choosing the sensor action that differentiates the most between the two most probable hypotheses performs as well (or better) than many other more advanced approaches, but is computationally much more feasible.

References

[1] J. M. Manyika and H. F. Durrant-Whyte. An information-theoretic approach to management in decentralized data fusion. In *Proceedings of the SPIE - The International Society for Optical Engineering*, 1992.

[2] K. Kastella. The search for optimal sensor management. In *Proceedings of the SPIE - The International Society for Optical Engineering*, 1996.

[3] H. Xu and J. Vandorpe. Perception planning in mobile robot navigation. In *1994 IEEE International Conference on MFI'94. Multisensor Fusion and Integration for Intelligent Systems*, Las Vegas, NV, USA, 1994.

[4] M. Abdulghafour, J. Goddard, and M. A. Abidi. Non-deterministic approaches in data fusion - a review. In *Proceedings of the SPIE - The International Society for Optical Engineering*, 1990.