ELSEVIER

A Bayesian approach to fusing uncertain, imprecise and conflicting information

Simon Maskell

QinetiQ, St. Andrews Road, Malvern, Worcestershire, WR14 3PS, UK

Abstract

The Dezert–Smarandache theory (DSmT) and transferable belief model (TBM) both address concerns with the Bayesian methodology as applied to applications involving the fusion of uncertain, imprecise and conflicting information. In this paper, we revisit these concerns regarding the Bayesian methodology in the light of recent developments in the context of the DSmT and TBM. We show that, by exploiting recent advances in the Bayesian research arena, one can devise and analyse Bayesian models that have the same emergent properties as DSmT and TBM. Specifically, we define Bayesian models that articulate uncertainty over the value of probabilities (including multimodal distributions that result from conflicting information) and we use a minimum expected cost criterion to facilitate making decisions that involve hypotheses that are not mutually exclusive. We outline our motivation for using the Bayesian methodology and also show that the DSmT and TBM models are computationally expedient approaches to achieving the same endpoint. Our aim is to provide a conduit between these two communities such that an objective view can be shared by advocates of all the techniques.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Information fusion; Bayesian; Uncertainty; Imprecision; Conflicting information; Transferable belief model; Dezert–Smarandache theory; Dempster–Shafer theory

1. Introduction

In information fusion applications, it is the representation of uncertainty that is the key enabler to extracting information from multi-sensor data (both co-modal data from multiple sensors of the same type and cross-modal data from sensors of different types). The development of all information fusion algorithms is critically dependent on using an appropriate method to represent uncertainty. A number of different paradigms have been developed for representing uncertainty and so performing data and information fusion, which are now briefly discussed:

- Fuzzy logic [1] represents belief through the definition of a mapping between quantities of interest and belief functions.

- Bayesian probability theory [2] articulates belief through the assignment of probability mass to mutually exclusive hypotheses.
- Dempster–Shafer theory (DST) [3] generalises Bayesian theory to consider upper and lower bounds on probabilities.
- The transferable belief model (TBM) [4] and Dezert–Smarandache theory (DSmT) [5] are further generalisations (over DST) of Bayesian theory. The TBM and DSmT represent uncertainty over the assignment of probability to mutually exclusive hypotheses by instead assigning probability to a power set of mutually exclusive hypotheses.
- Recently, a further generalisation, involving assignment of mass to a hyper-power set of hypotheses has been proposed [6].

Advocates of Bayesian theory make reference to a proof that Bayesian inference is the only way to consistently

E-mail address: s.maskell@signal.qinetiq.com

manipulate belief relating to a set of hypotheses [7]. Conversely, advocates of DST, the TBM and DSmT motivate their approaches by the fact that given a set of hypotheses, Bayesian inference is unable to satisfactorily manipulate uncertain, imprecise and conflicting information [3–5]. This paper aims to act as a conduit between these two extreme viewpoints and the associated information fusion research communities. The hope is that this paper acts as a catalyst for the cross-fertilisation of ideas between these communities. The paper is intended to complement related work that has considered how one can subsume DST, the TBM and DSmT into a Bayesian approach [8] and approaches based on robust Bayesian inference [9]; this paper differs in that we explicitly consider how to devise Bayesian models that have the same emergent properties as analysis with DST, the TBM and DSmT.

The approach that is adopted is to accept that an initial application of Bayesian theory to fusion problems involving uncertain, imprecise and conflicting information is unable to satisfactorily manipulate such information. However, rather than attempt to redefine the method for manipulating belief on a given set of hypotheses, we choose to change the model definition and so the definition of the hypotheses. We show that, by exploiting recent advances in the Bayesian analysis of complex data (e.g. the recent development of, for example, particle filters [10] and Markov chain Monte-Carlo algorithms [11]), one can devise a rigorous Bayesian approach to fusing uncertain, imprecise and conflicting information. Furthermore, this approach has the same emergent properties as the TBM and DSmT, which can therefore be regarded as computationally efficient (although approximate) implementation strategies of this Bayesian approach.[1]

It should be noted that, as identified by the Bayesian community [12], model design is a critical component of a fusion system. Strong advocates of Bayesian inference will advocate the Bayesian methodology on the basis that this model design is made explicit. While making this explicit is useful, the problem of understanding how to design fusion systems remains whether model design is an implicit or explicit part of this process!

This paper is a rejection of the hypothesis that a Bayesian approach cannot solve certain problems involving the fusion of uncertain, imprecise and conflicting information. However, the author accepts that, while this paper demonstrates that an axiomatically consistent and robust Bayesian approach can be devised for such problems, specific system level constraints may dictate that approximations (such as those employed in the TBM and DSmT) should be used. The conclusions from any comparison is highly specific to the application being considered. So, this paper does not attempt to consider such comparisons, but aims to demonstrate that Bayesian approaches can and should be included in such comparisons in the future.

The paper begins in Section 2 with a description of how this Bayesian approach is devised. Section 3 considers several examples of how this approach is capable fusing uncertain, imprecise and conflicting information. Finally, Section 4 concludes.

2. Bayesian approach

2.1. Belief

Suppose an event has an outcome, x, that is one of a number of mutually exclusive hypotheses, $x \in X$. Furthermore, suppose one of these hypotheses is true, while the others are all false.

From a Bayesian (not frequentist) perspective, probability quantifies belief. To avoid confusion with belief functions, the term probability will be used from this point hence where appropriate. The probability associated with a hypothesis, $p(x)$, is a number that represents which of the mutually exclusive hypotheses we believe to be true. This probability is always non-negative and sums to unity across the hypotheses[2]:

$$p(x) \geqslant 0 \tag{1}$$

$$\sum_{x \in X} p(x) = 1 \tag{2}$$

Unfortunately, the true event is often very complex and cannot be modeled exactly. In such scenarios one must consider a model, which is an approximation to the real world. This approximation is chosen to be high enough fidelity that it captures the complexity of the event in terms of the parameters of interest but low enough fidelity that the probability can be calculated. It is this model complexity that is the key to the development of a Bayesian approach to fusing uncertain, imprecise and conflicting information (as shown in Section 3.2).

This model is the *prior*; it articulates the anticipated outcome of the event before any measurements are received. The choice of prior makes explicit all relevant knowledge of the system under consideration. Implicit consideration of prior knowledge as part of (for example) maximum likelihood modeling, is often equivalent to a specific explicit model of prior knowledge. However, there is a danger with implicit prior knowledge modeling that one unintentionally can introduce strong prior knowledge implicitly, as a result of parameterisation for example; one cannot be simultaneously ignorant of all parameterisations of a variable[3].

---

[1] The implication is that since TBM and DSmT approximate the only consistent way to manipulate beliefs, there will be scenarios where these approximations degrade performance significantly. Conversely, there will be scenarios where these approximations do not impact performance and are vital in facilitating real-time processing. Understanding which class of scenarios includes a given scenario remains an open research question.

[2] Open and closed worlds will be considered shortly.

[3] As a simple example, consider a point in a 2D plane. If one assumes all cartesian position of the point are equally likely, this puts a non-uniform prior on points when defined in polar co-ordinates. So, an uninformative prior on one parameterisation is not uninformative in another parameterisation.

This disparity between the true system and the model can lead to the model covering a subset of the potential outcomes of the event and naturally leads to the distinction between a *closed world* assumption and an *open world* assumption. In a closed world, one makes the strong assumption that the subset of events that the model caters for are a large subset of the total set of events. Conversely, in an open world, one admits the possibility that the true outcome of the event is not part of the model.

It is possible to articulate an open world in a Bayesian model. To do this, one must consider the fact that the (closed world) model is not a complete description of the true system as part of the open world model. More specifically, one must extend the model to include a hypothesis or set of hypotheses that represent the assumption of a closed world model being incorrect. These hypotheses do not need to be carefully defined, but simply need to articulate knowledge of the anticipated order of magnitude of variables (as is considered in Section 3.3).

2.2. Ignorance

One often has a number of decisions, $d \in D$, that can be made and a reward[4] associated with making each decision in the case that each hypothesis is true, $R(d|x)$. An optimal decision, $d^\star$, is then defined as one that maximises the expected reward:

$$d^\star = \arg\max_d \sum_{x \in X} R(d|x)p(x) \qquad (3)$$

The decisions can have labels and these labels can be associated with the outcome of the event. However, there is no requirement for the labels to be mutually exclusive or for there to be the same number of labels as there are hypotheses.

So, one can have decisions with labels that relate to multiple hypotheses being true. Given the rewards and the probability, the optimal decision can then be to select a decision with a label that relates to a sets of hypotheses (as considered in Section 3.3). Using such a formulation a Bayesian approach can decide to claim ignorance.

It is worth noting the similarity to ideas in the TBM and DSmT literatures (such as the pignistic transform [4]) that involve transforming belief masses associated with elements of the power set of hypotheses to decisions relating to the mutually exclusive hypotheses.

2.3. Uncertain belief

If we receive two independent measurements, y_1 and y_2, and wish to know how to update our probability about x, $p(x)$, given these measurements, we can apply Bayes rule as follows:

---

4 Such reward functions could be defined by an expert or could be estimated from historic data.

$$p(x|y_1, y_2) = \frac{p(y_1|x)p(y_2|x)p(x)}{p(y_1, y_2)} \qquad (4)$$

where $p(x|y_1, y_2)$ is the updated *posterior* probability and we have assumed knowledge of how likely the measured data was given any assumed known state, x, is articulated in the *likelihoods*, $p(y_1|x)$ and $p(y_2|x)$. Note that $p(y_1, y_2)$ is just a normalising constant and not a function of x and that the assumption of independent measurements has been exploited in deriving (4).

Eq. (4) is true if $p(y_1|x)$ and $p(y_2|x)$ are exact. However, typically, these quantities are calculated by integrating over some other parameters, θ_1 and θ_2:

$$p(x|y_1, y_2) = \frac{p(y_1|x)p(y_2|x)p(x)}{p(y_1, y_2)} \qquad (5)$$

$$= \int p(y_1, \theta_1|x)\, d\theta_1 \int p(y_2, \theta_2|x)\, d\theta_2$$

$$\times \frac{p(x)}{p(y_1, y_2)} \qquad (6)$$

If the integrals in (6) are not analytically tractable then they must be approximated and therefore the resulting application of Bayes rule is also approximate. If one of the approximations is less accurate than the other then the associated term in Bayes rule will be more uncertain. The result is that these errors have an adverse effect on a fusion process that assumes $p(y_1|x)$ and $p(y_2|x)$ to be exact.

To cater for this in a Bayesian framework, one can represent the error in the integrals by considering a number of hypotheses for the error process and so a number of hypotheses for the true likelihood. The diversity of the sampled likelihoods then conveys the imprecise nature of the probability and one can fuse the hypotheses by considering trajectories through the space of samples. This use of the diversity of a set of samples to convey imprecise information is illustrated in Section 3.4 using a simple variant of a particle filter [10].

2.4. Conflict

Another form of imprecise information is that resulting from conflicting information. The imprecise nature of the probability is manifested as there being multiple different explanations for the data that result in very different probabilities about some quantity of interest. This conflict needs to be represented if later data are to be able to refine the probability over which explanation is most likely.

To articulate this conflicting information in a Bayesian context, one can consider multiple hypotheses that explain the data, where each hypothesis has associated with it a probability about the quantity of interest. The conflict is then represented through the diversity of these hypotheses, which, in the case of conflicting, rather than imprecise information, will typically result in very different probabilities about the quantity of interest. This is exemplified in Sections 3.1 and 3.5.

Table 1
Table of Experts' fused conclusion as a function of their probability of making an error given the data, Y, as discussed in Section 3.1

$P(e)$	$P(\bar{e}_1, \bar{e}_2|Y)$	$P(\bar{e}_1, e_2|Y)$	$P(e_1, \bar{e}_2|Y)$	$P(e_1, e_2|Y)$	$P(M|Y)$	$P(T|Y)$	$P(C|Y)$
0.01	0.0013	0.4731	0.4731	0.0525	0.4859	0.4859	0.0282
0.001	0.1303	0.4347	0.4347	0.0003	0.4305	0.4305	0.1391
0.0003	0.3333	0.3333	0.3333	0.0001	0.3300	0.3300	0.3400
0.0001	0.6000	0.2000	0.2000	0.0000	0.1980	0.1980	0.6040

Table 2
Parameter values for Identification fusion considered in Section 3.2

Model	Mean	Variance
747	0	1
Fighter model 1	0	1
Fighter model 2	0	100

3. Examples

We now consider some examples for which a simple application of Bayesian inference encounters difficulties, but where, through refining the model, we are able to resolve these issues without departing from the Bayesian paradigm.

In this paper, the aim is to demonstrate that one can represent uncertainty over probability in a Bayesian context. To model this uncertainty in a way that is easily articulated necessitates the use of specific algorithms in the context of the exemplar applications. It is anticipated that other Bayesian algorithms would be better suited to these applications. These other algorithms would be equivalent to modeling the uncertainty over probability. However, these other algorithms would not be well suited to demonstrating that a Bayesian approach can represent uncertainty over probability. This is the motivation for the models, algorithms and parameter values used in this section.

3.1. Zadeh's example

This example was proposed by Zadeh [13], and has been used as motivation to extend Dempster–Shafer reasoning to consider conflict and demonstrated to be solved using the TBM [14] and DSmT [15] theories. The discussion is reminiscent of that proposed by other authors (for example in [16]), but the focus here on demonstrating that the issue identified by Zadeh can be resolved without a departure from a Bayesian context.

3.1.1. Zadeh's problem

Two experts are consulted about a patient. The experts diagnose the patient into three classes, (M)eningitis, (C)oncussion and Brain (T)umor. One expert states that, "I am 99% sure it's meningitis, but there is a small chance of 1% that it's concussion". The other expert states that, "I am 99% sure it's a tumor, but there is a small chance of 1% that it's concussion".

3.1.2. Solutions to Zadeh's problem

A straightforward application of a naïve Bayesian (or Dempster–Shafer) approach results in a fused output of there being a 100% probability of the patient having concussion.

Zadeh argues that this is counter-intuitive and asks how both experts could be so wrong. The author asserts that if one trusts the experts' abilities to calculate these probabilities, then this fused output is correct. However, intuition indicates that one of the experts got something wrong.
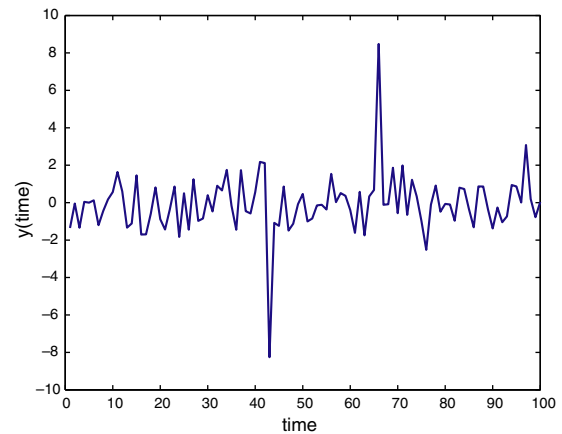

Fig. 1. Exemplar data for scenario 1 considered in Section 3.2.
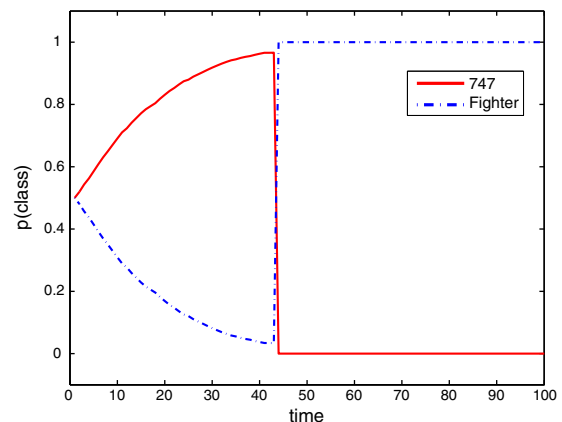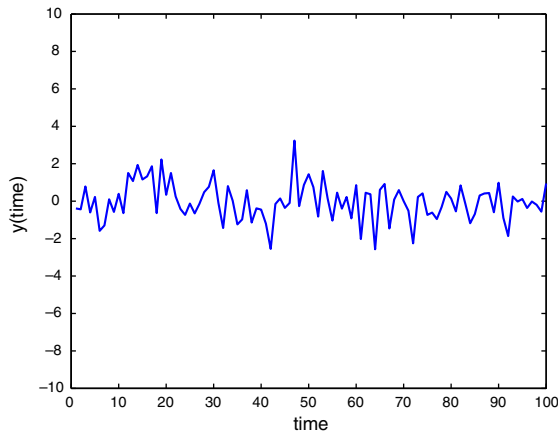

Fig. 2. Sequential classification output for exemplar data for scenario 1 considered in Section 3.2.

From a Bayesian perspective, this indicates that the model is insufficiently complex to consider factors that intuition indicates are important. Specifically, there is a need to model the fact that the experts may have made an error. It is straightforward to extend the hypothesis space to consider the experts making such errors.

Denote e_i for the hypothesis that the ith expert makes an error and $\bar{e}_i$ for the hypothesis that the ith expert does not make such an error. One assumes each expert was in error with a prior probability of $P(e_i) = P(e)$. If an expert was in error, then the classification probabilities for that expert are taken to be uniform across the three classes.

One can then simply apply Bayes rule to calculate the fused classification and the posterior probability that the experts was in error. More specifically, one can consider each of the four combinations of experts being in error and not. For each combination, one can calculate a fused classification result (normalised to unity) and a weight for that combination (equal to the sum of the unnormalised product of the experts' classification probabilities multi-

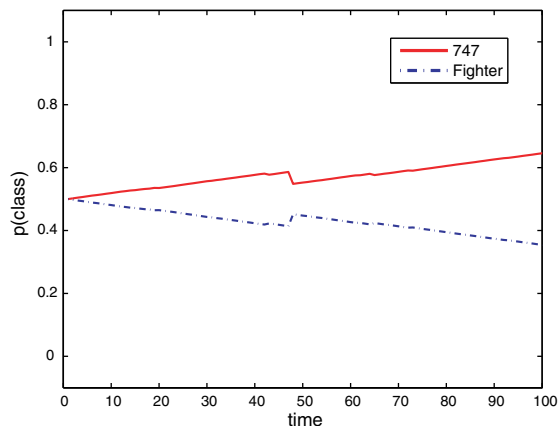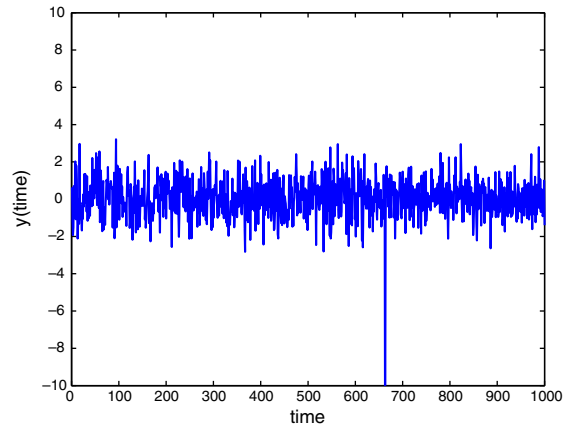Fig. 5. Exemplar data for scenario 3 considered in Section 3.2.

Fig. 6. Sequential classification output for exemplar data for scenario 3 considered in Section 3.2.

Fig. 3. Exemplar data for scenario 2 considered in Section 3.2.
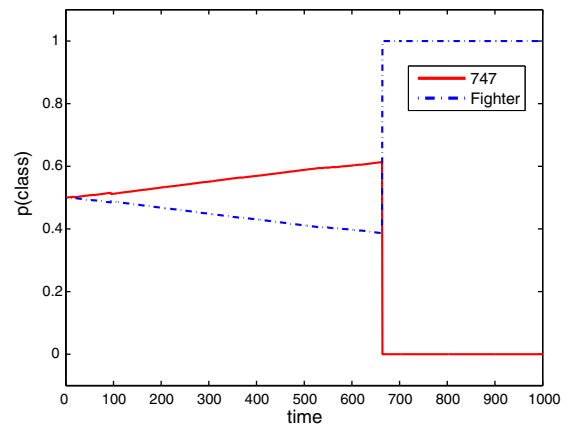

Fig. 4. Sequential classification output for exemplar data for scenario 2 considered in Section 3.2.

plied by the priors on whether the experts were in error). One can then calculate the fused classification as a weighted sum of the fused classification results. Table 1 shows this fused classification result and the posterior probabilities of the different combinations of experts' errors, for each of a number of values for $P(e)$.

It is evident that $P(e)$ needs to be very small for this approach to draw the same conclusion as the naïve Bayesian fusion approach[5]; one needs to place a surprisingly large amount of trust in the experts' opinions (ie. that one expects less than 3 in 10,000 experts to be wrong a priori) for the most probable conclusion to be that the patient has concussion. For values of $P(e)$ judged to be in accordance with the author's intuition, the posterior indicates

---

[5] This example emphasises that a probability of zero is a very informative input; if one expert calculates the probability of a hypothesis to be zero, no weight of evidence from other experts can make this the most likely hypothesis. Zero and nearly-zero are therefore very different probabilities in terms of their effect on a naïve Bayesian fusion algorithm.
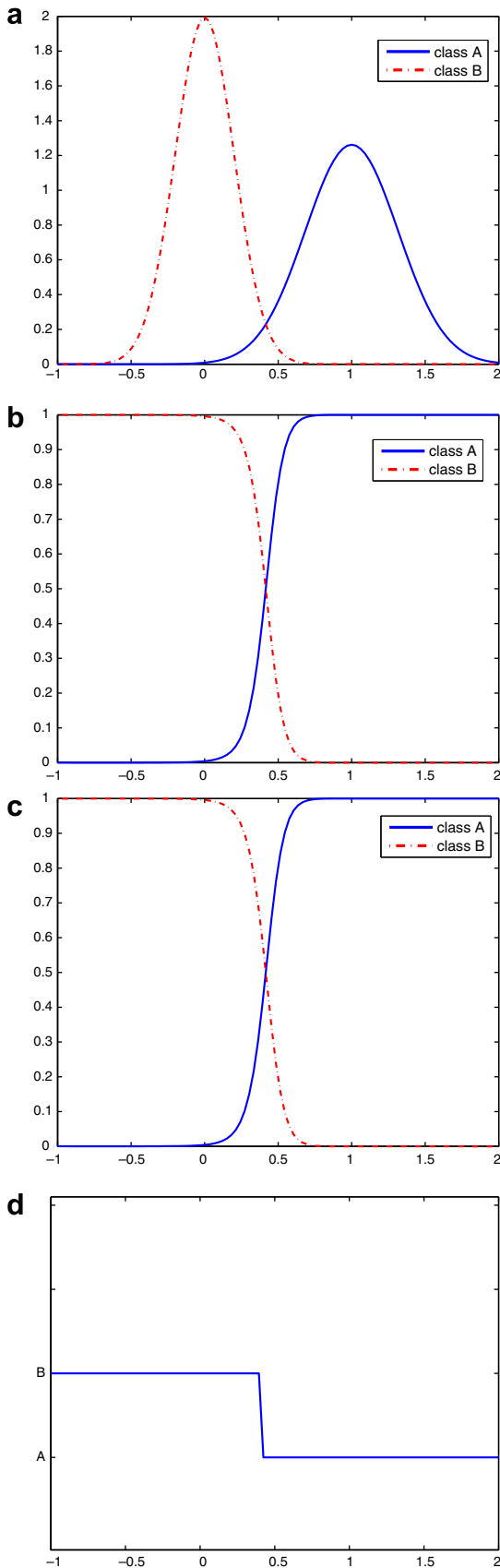
Fig. 7. Risk averse classification scenario 4 considered in Section 3.3: (a) likelihood; (b) classification probabilities; (c) expected cost; (d) decisions.

Table 3
Costs for scenario 4 considered in Section 3.3

Decision	Class	
	A	B
A	1	0
B	0	1

Table 4
Parameter values for decision making scenarios (scenarios 4–7) considered in Section 3.3

Class	Mean	Variance
A	1	0.1
B	0	0.04
∅	0.5	1

Table 5
Costs for scenario 5 considered in Section 3.3

Decision	Class	
	A	B
A	1	0
B	0	0.01

that one of the experts was in error and that the other expert's judgement was correct.

Note that this shows that by extending the hypothesis space, one can consider problems with conflict in a Bayesian context. It is also worth noting that, in this example, the same effect could be considered by simply modifying the expert's probabilities before applying a naïve Bayesian fusion approach. Such an approach would not take onboard the author's perception of the point Zadeh was making in his paper; the experts both believe they are correct!

3.2. Identification fusion

Motivated by some previous work used to motivate the TBM [17], we consider the classification of an air target into one of two classes: fighter jet and 747. We observe accelerations and have two models, one for fighter jet and one for 747. Crucially and in contrast to [17], we use models that agree with our intuition: for the bulk of the time, a fighter jet and 747 have accelerations that are drawn from the same Gaussian distribution. However, the fighter jet occasionally has high accelerations. We model this with a component with small weight in a Gaussian mixture for the fighter jet's model, such that the only difference is that the model for the fighter jet's acceleration has heavier tails than that for the 747. The parameter values are shown in Table 2.
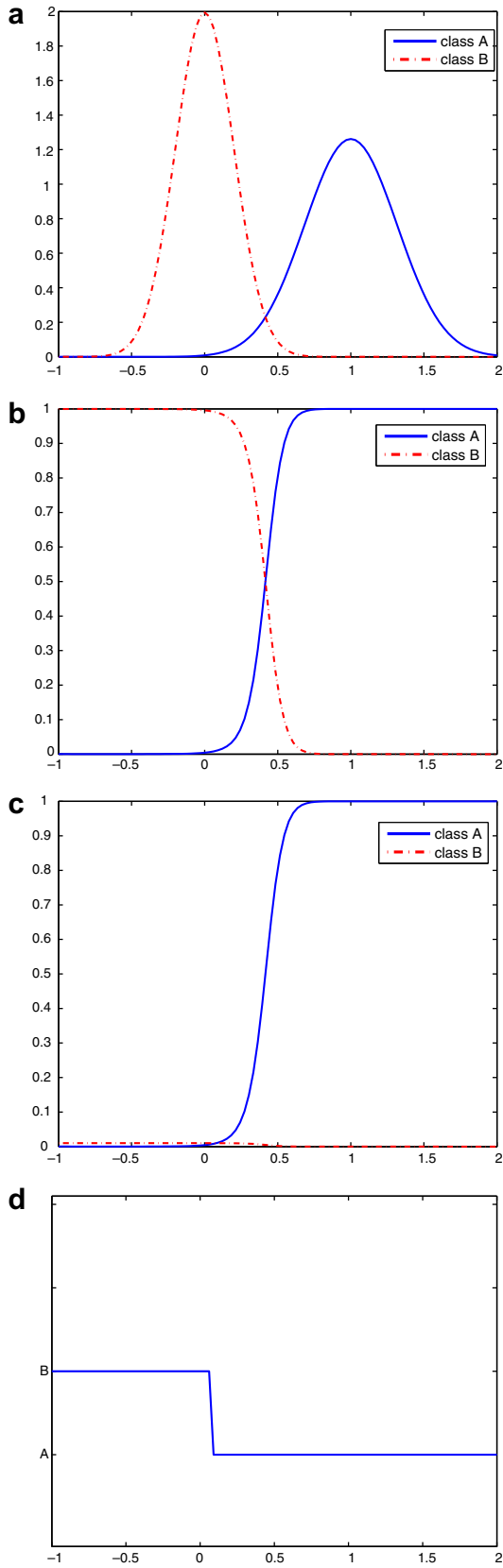
Fig. 8. Risk averse classification scenario 5 considered in Section 3.3: (a) likelihood; (b) classification probabilities; (c) expected cost; (d) decisions.
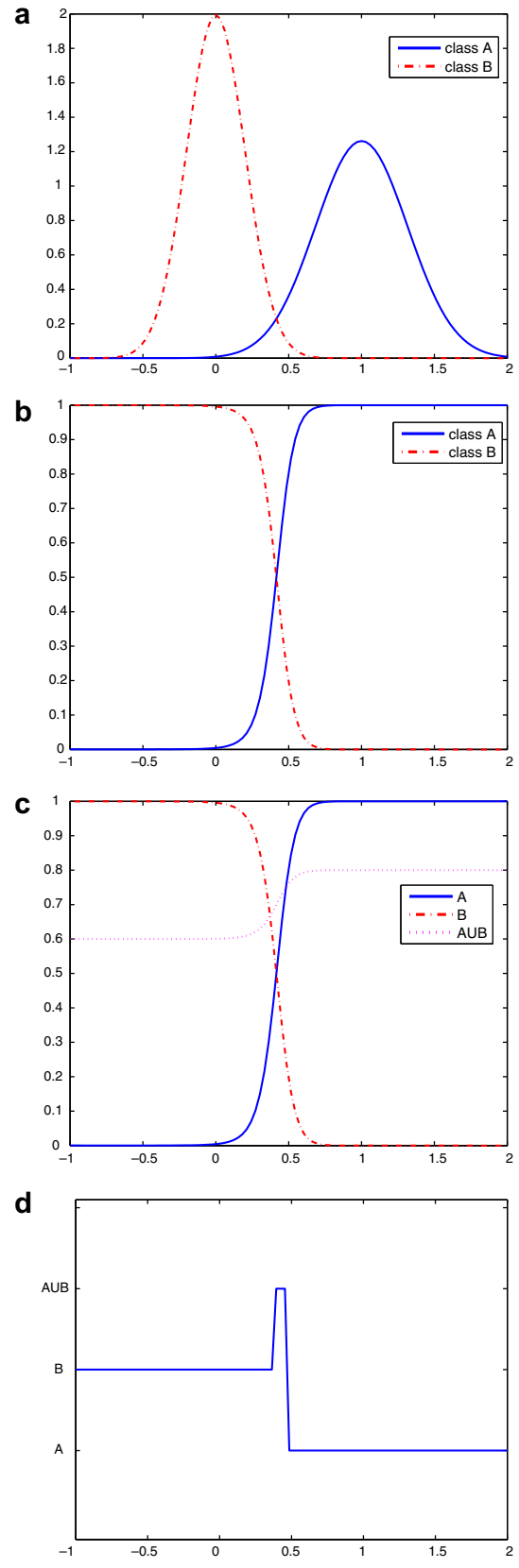

Fig. 9. Risk averse classification scenario 6 considered in Section 3.3: (a) likelihood; (b) classification probabilities; (c) expected cost; (d) decisions.

Table 6
Costs for scenario 6 considered in Section 3.3

Decision	Class	
	A	B
A	1	0
B	0	1
$A \bigcup B$	0.8	0.6

Table 7
Costs for scenario 7 considered in Section 3.3

Decision	Class		
	A	B	∅
A	1	0	0
B	0	1	0
∅	0.4	0.4	1

3.2.1. Scenarios 1, 2 and 3

We use bank of filters [18] to fuse data over time.[6] We consider three scenarios: the weight on the large variance component (Fighter Model 2) in the Gaussian mixture is respectively 0.1, 0.01 and 0.001. Exemplar data (generated by simulating from the fighter jet model) are shown in Figs. 1, 3 and 5. The associated classification output as a function of time is shown in Figs. 2, 4 and 6. Note that the time scales are different for scenario 3 (since the average time between outliers is significantly longer than in scenario 1).

From an initially equal classification probability, it can be seen that the classification output evolves towards a probability that favours the 747 until a large amplitude measurement is received, at which point the target is classified as a fighter jet. The evolution is at a rate that decreases as the heavy tailed component's weight reduces.

3.3. Risk averse classification

Motivated by the desire to illustrate the ability of Bayesian analysis to consider an open world and articulate ignorance, we consider a two class problem.

3.3.1. Scenario 4

The two classes, A and B, have likelihoods relating to a scalar parameter as shown in Fig. 7a; the likelihoods are Gaussian with parameters tabulated in Table 4. From this, were one to observe a value of this parameter, the classification probabilities would be as shown in Fig. 7b. Given the same reward for correctly classifying and misclassifying

---

[6] Note that a related Bayesian approach can be used to make the filter efficient by adapting in response to the number of likely classes [19]. This approach is perceived by the author to meet the same design aims as the Transferable Belief Model, which uses the transfer of belief to achieve this efficiency.
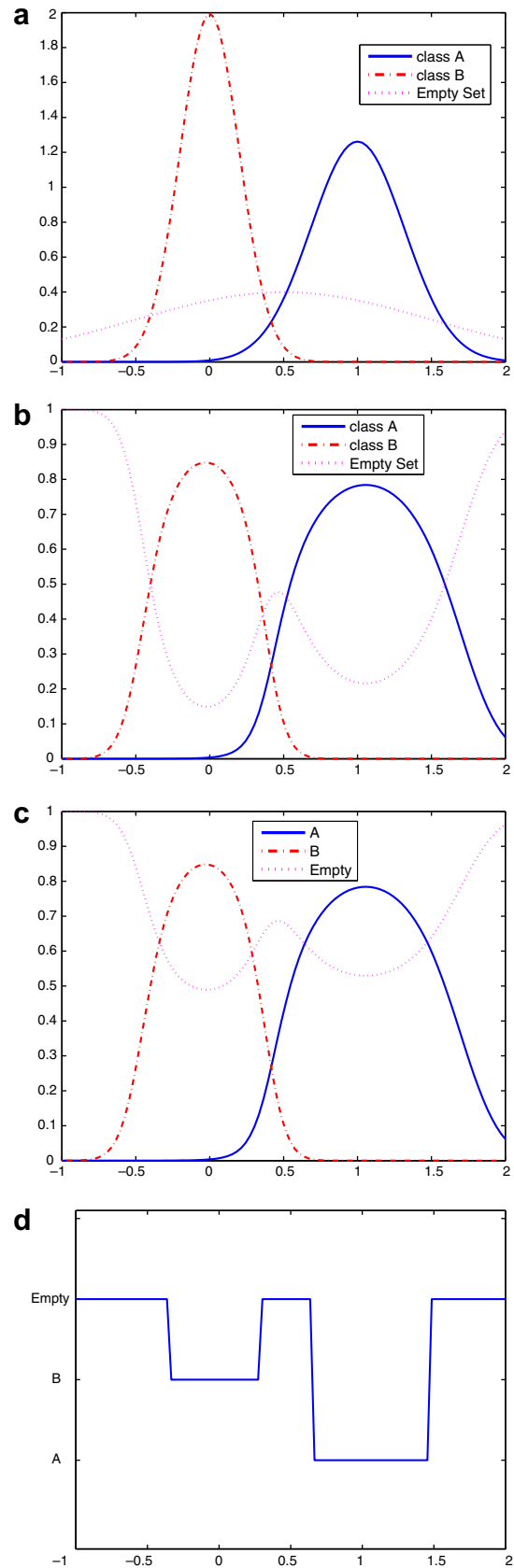

Fig. 10. Risk averse classification scenario 7 considered in Section 3.3: (a) likelihood; (b) classification probabilities; (c) expected cost; (d) decisions.
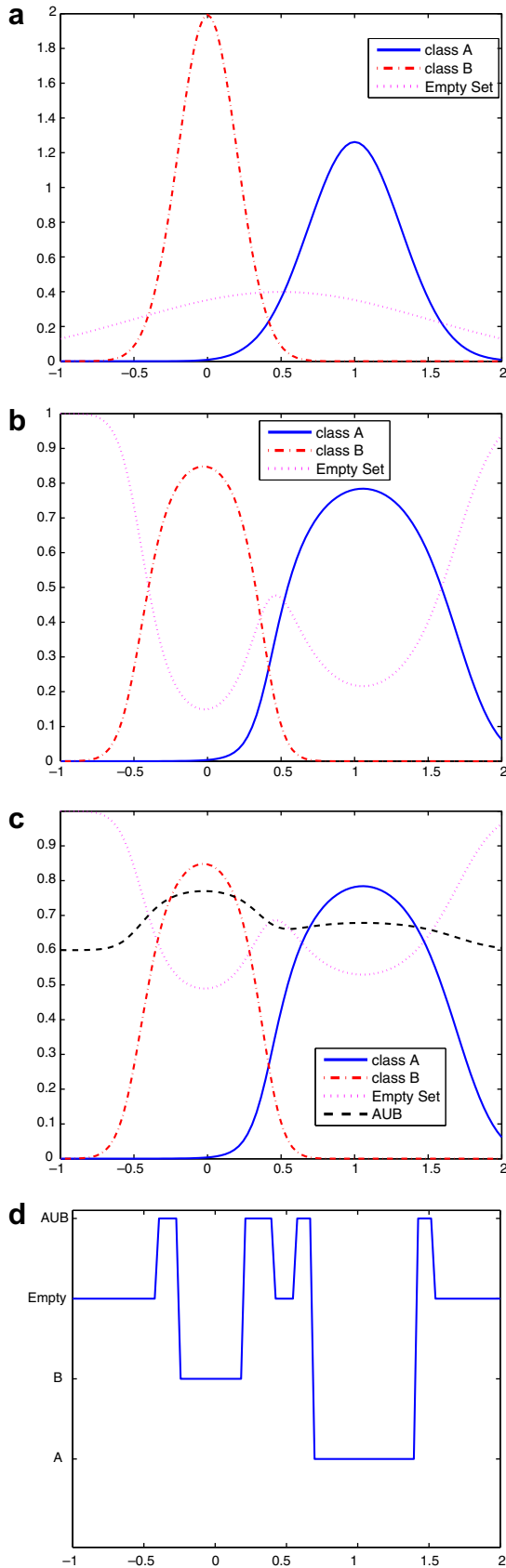
Fig. 11. Risk averse classification scenario 8 considered in Section 3.3: (a) likelihood; (b) classification probabilities; (c) expected cost; (d) decisions.

Table 8
Costs for scenario 8 considered in Section 3.3

Decision	Class		
	A	B	$\emptyset$
A	1	0	0
B	0	1	0
$\emptyset$	0.4	0.4	1
$A \bigcup B$	0.7	0.8	0.6

Table 9
Table of parameter values used in Section 3.4

Parameter	Classifier 1		Classifier 2	
	A	B	A	B
Mean	0	1	1	0
Variance	0.1	0.1	0.1	0.1
Variance of mean	0.5	0.5	0.00001	0.00001

Table 10
Classification naïve output considered in Section 3.4

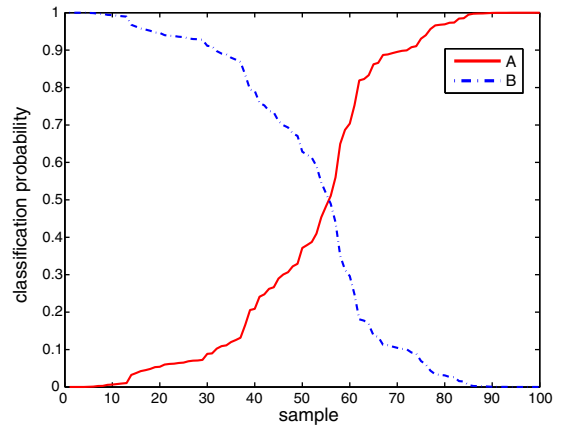| Class	Classifier 1	Classifier 2	Fused output
A	0.9526	0.0474	0.5
B	0.0474	0.9526	0.5

Fig. 12. Classification outputs for each of 100 samples in one Monte-Carlo run considered in Section 3.4.

a target of each type (the rewards are tabulated in Table 3), the expected cost for two decisions, A and B are as shown in Fig. 7c. Hence, the optimal decision for different observed parameters is as shown in Fig. 7d. Note that there is a boundary to one side of which the optimal decision is
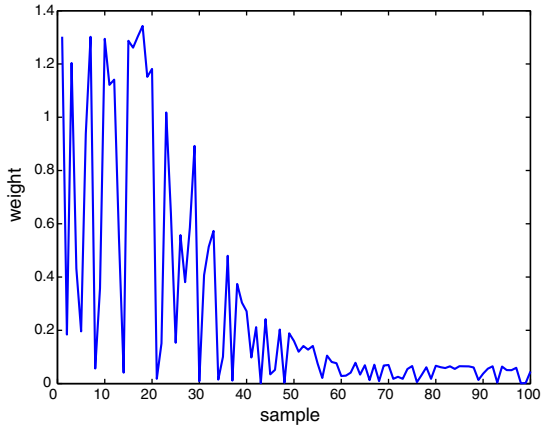
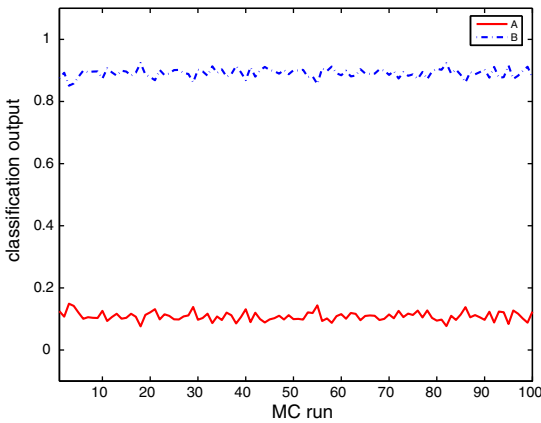Fig. 13. Weights for each of 100 samples in one Monte-Carlo run considered in Section 3.4.

Fig. 14. Outputs of 100 Monte-Carlo runs illustrating the fusion of imprecise information considered in Section 3.4.
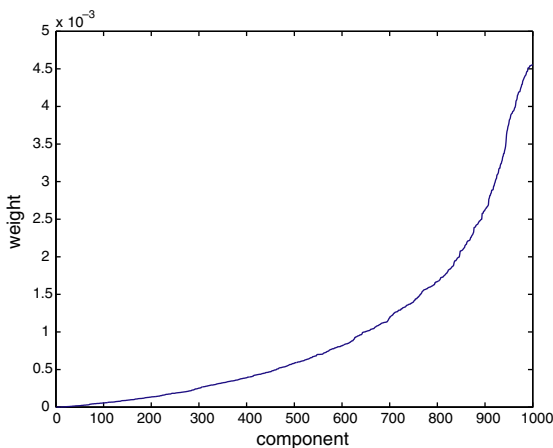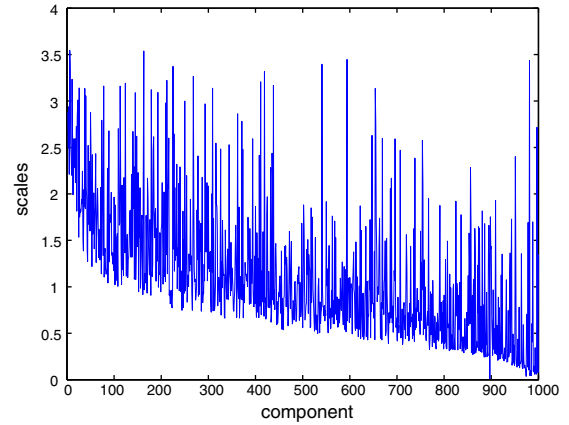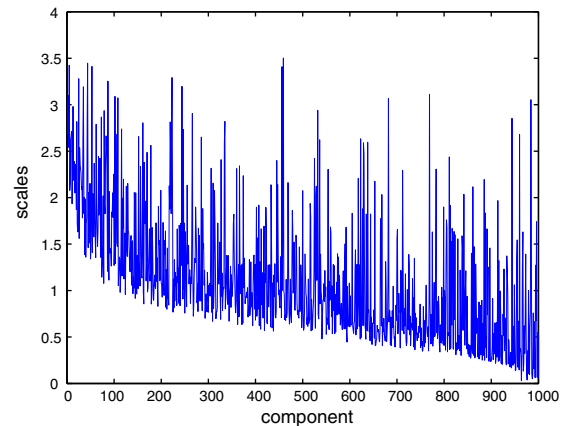

Fig. 15. Components' mixture weights sorted in order of increasing weight, as discussed in Section 3.5.

(a) x

(b) e

Fig. 16. Components' scales sorted in order of increasing weight, as discussed in Section 3.5.

that the target is a member of class A and to the other side of which the optimal decision is that the target is a member of class B.

3.3.2. Scenario 5

If one changes the reward structure to that shown in Table 5 such that there is a different reward for correctly classifying one target type than the other then the decision boundary moves, as illustrated in Fig. 8.

3.3.3. Scenario 6

To cater for ignorance, as discussed in Section 2.2, rather than consider an alternative methodology for manipulating probability, one can introduce another decision with a label of $A \bigcup B$. As shown in Fig. 9, by defining appropriate rewards (given in Table 6), this decision (that one is ignorant) is then optimal when certain observations are received.

3.3.4. Scenario 7

Furthermore, by introducing an open world model, $\emptyset$, which (as defined in Table 4) is a vague prior on the param-
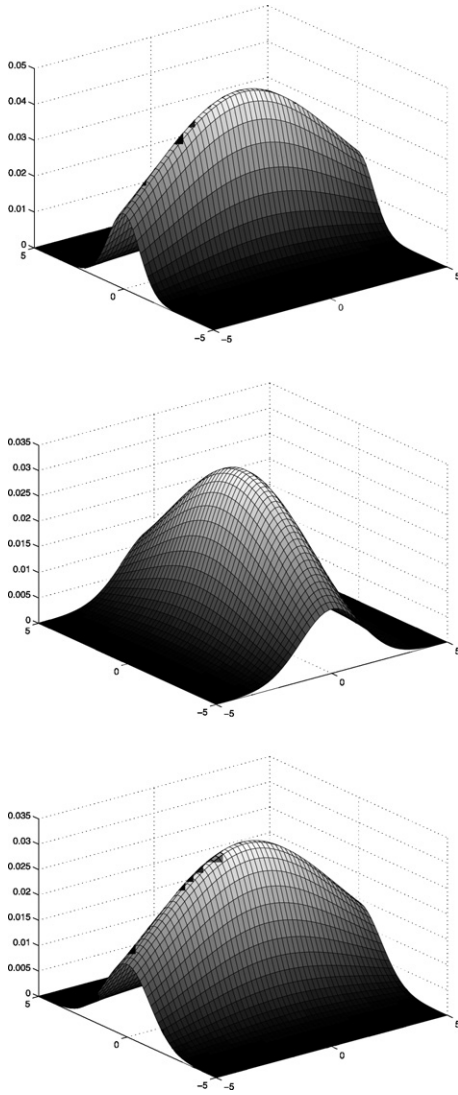
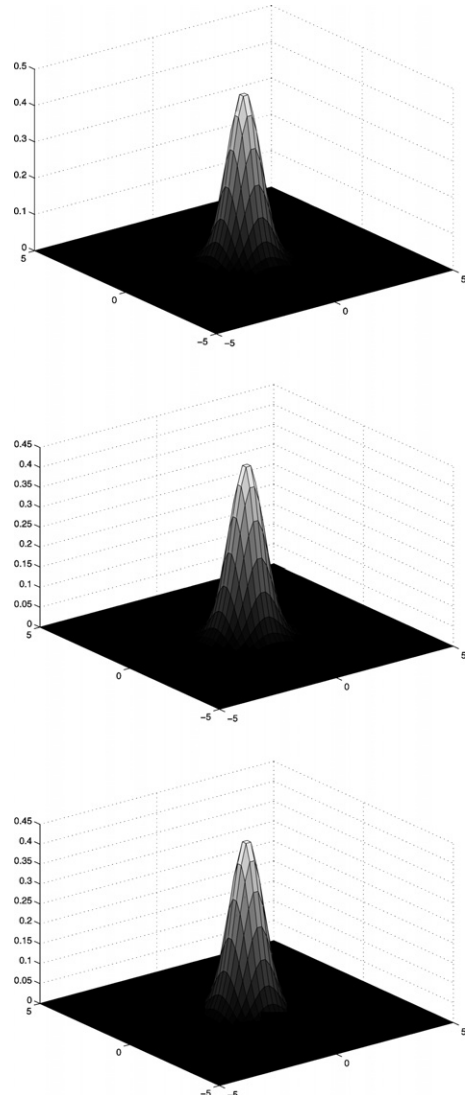Fig. 17. Distributions associated with three components with largest weights, as discussed in Section 3.5.

Fig. 18. Distributions associated with three components with smallest weights, as discussed in Section 3.5.

eter value[7], one can define rewards (shown in Table 7) such that the optimal decision given certain observations is to classify the target as not a member of *A* or *B*. This is illustrated in Fig. 10.

3.3.5. Scenario 8

Finally, one can combine these concepts to devise a Bayesian approach to decision making that adopts an open world model and can decide one is ignorant. This is exemplified in Fig. 11, which is based on the costs shown in Table 8.

---

[7] The definition of the open world model needs to make explicit any implicit knowledge of the order-of-magnitude of the parameters. This process of explicitly articulating this knowledge is potentially non-intuitive. However, this knowledge must exist if one can entertain the possibility that a closed world assumption is not valid.

3.4. Fusion of imprecise classification information

We now consider an example of fusing the output of two Gaussian classifiers, each of which has a model for classes
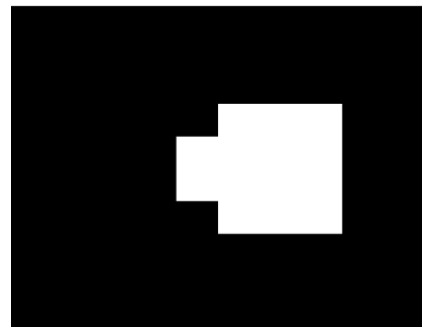

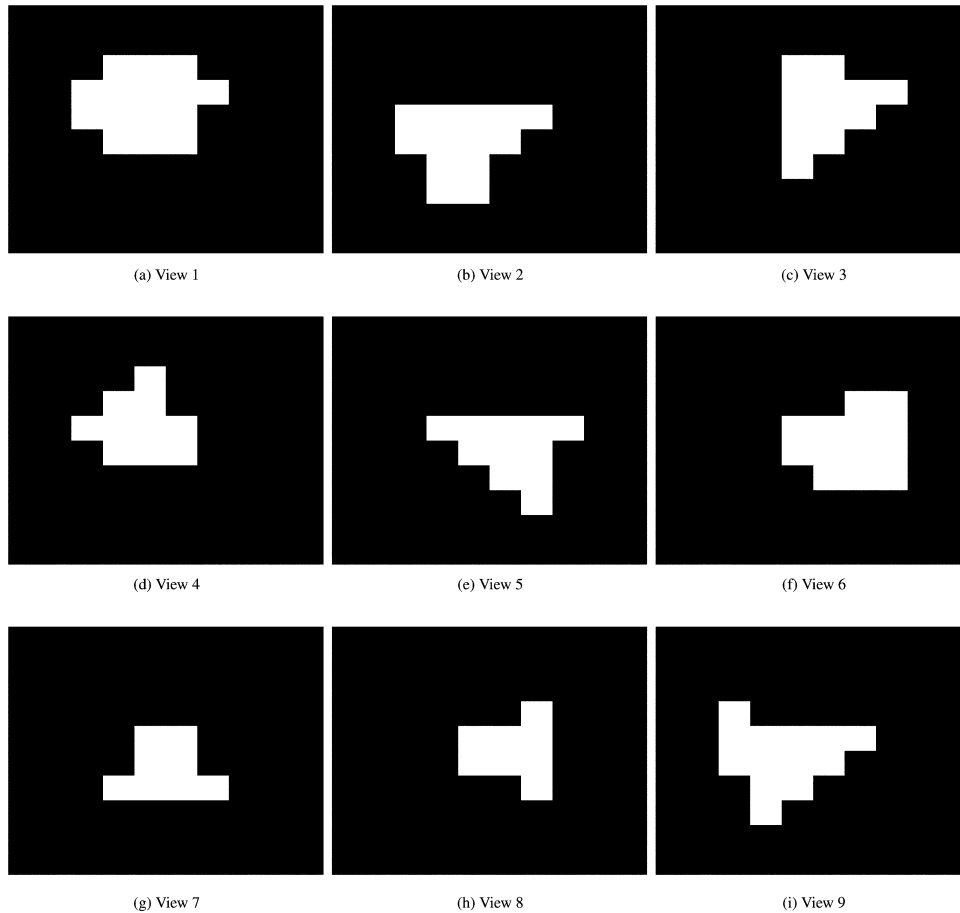Fig. 19. Received Image discussed in Section 3.6.

(a) View 1   (b) View 2   (c) View 3

(d) View 4   (e) View 5   (f) View 6

(g) View 7   (h) View 8   (i) View 9

Fig. 20. Templates for cone considered in Section 3.6.

A and *B*. We assume one of the classifiers is more imprecise; the estimates of its parameters have a larger variance (perhaps due to the availability of less training data for this classifier). The mean and variance for the models in the classifiers together with the variance of the mean are shown in Table 9.[8]

The two classifiers make a measurement of 0.2. If we use the estimated parameter values, the classifiers output the probabilities shown in Table 10. A naïve Bayesian fusion of these two classification outputs results in the fused output shown. Note that the fused output is midway between the two classifiers' output, whereas, since we know that the parameter values for classifier 2 are more accurate, one might expect the fused output to be biased towards the output of classifier 2.

We represent the uncertainty over the classifiers' parameters through the diversity of 100 samples. More specifically, we employ importance sampling (a full particle filter, with resampling, is not necessary here since we are only considering two outputs[9]). We sample 100 samples of the means for the two classes and the two classifiers. For each sample, we calculate the importance weight, which (since we have sampled from the prior) is just the likelihood (integrating over the classes) and a classification output.

The classification outputs for one Monte-Carlo run are shown in Fig. 12 (sorted in decreasing order of probability of class *B*). The weights are shown in Fig. 13 (sorted in the same order as Fig. 12). It is clear that the samples with high weights all have fused outputs that have a high classification probability for class *B*.

We calculate an output by using a weighted average of the samples' classification outputs. The resulting output from each of 100 Monte-Carlo runs are shown in Fig. 14. It is clear that this output is in agreement with intuition and is accounting for the imprecision of the information

---

[8] Note that, in this specific case, one could analytically integrate the uncertainty over the mean estimate. However, the aim here is to devise an exemplar illustration of how a Bayesian analysis can be used to fuse imprecise information and the specifics of the example are chosen primarily to be straightforward to understand by the target audience.

[9] A particle filter would be necessary if we were considering the fusion of many classifier outputs. The model design would need to explicitly consider whether the errors in the parameter estimates were assumed static or could be modeled as independent errors at each timestep. If the parameters are assumed static, then more sophisticated techniques (such as [20] and more recent related developments) will be needed to avoid the degeneracy issues that are encountered when naïvely applying particle filters to such problems.
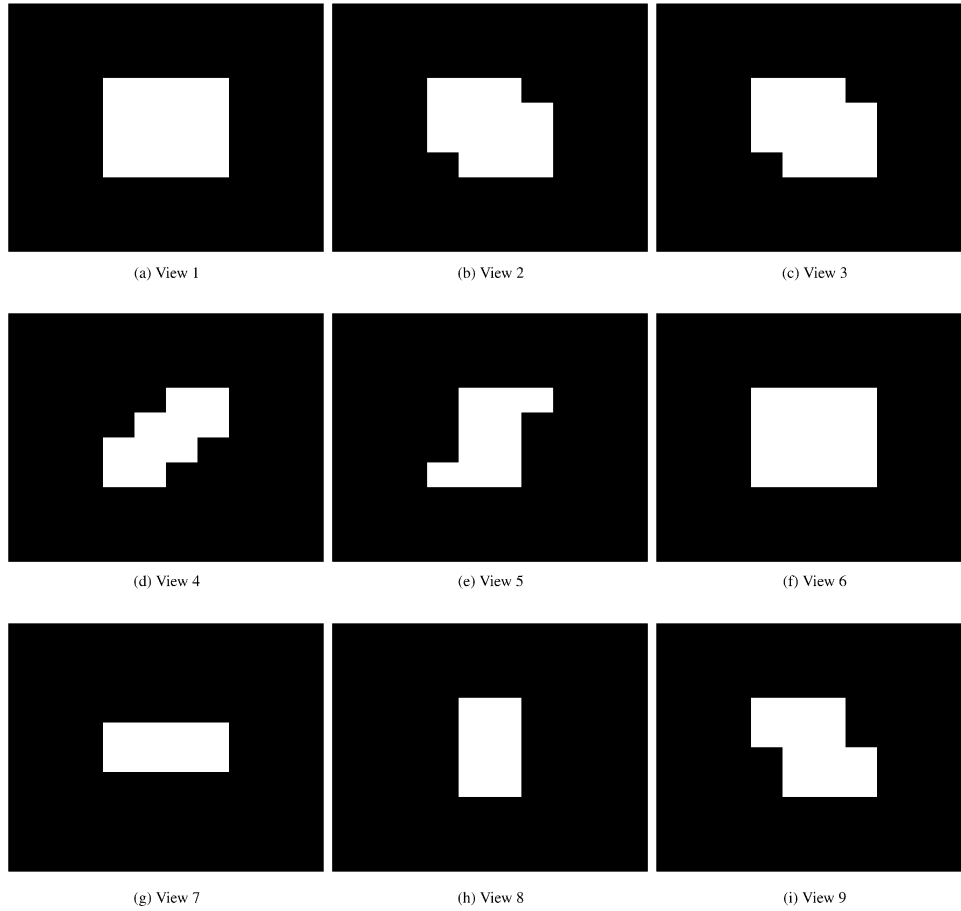
Fig. 21. Templates for cylinder considered in Section 3.6.

in the fusion of the identification outputs; the fused classification output is evidently closer to that output from classifier 2.

Note that this example has assumed that we have the ability to consider the imprecise classification output as the result of unknown parameters of the classifier, which we can sample. There is an argument for considering scenarios in which the classifier operates as a black box. One could then consider the observed classification output as a measurement and use likelihoods in the classification space to model the imprecision. However, the author has a strong preference for explicitly considering the parameters of the classifier and this has motivated the example chosen.

3.5. Conflict over belief of vector-valued continuous variables

There has been recent interest in the extension of the TBM to consider uncertainty over real-valued quantities [21]. In this example, we demonstrate that a Bayesian approach to such problems is straightforward to develop and that it trivially extends beyond the scalar real-valued quantities considered before to representation of uncertainty over vectors of real-valued quantities. This example

also demonstrates the ability of the Bayesian approach to handle conflict.

We consider a scenario where we are interested in some state, x. We observe y, which is the sum of x and some measurement noise, e:

$$y = x + e \tag{7}$$

Both x and e are heavy tailed so an outlier for y can be the result of either the process generating x or that generating e. We wish to infer the values of x and e from a single outlying measurement of y.

We choose to represent the heavy tailed distributions for x and e using a scale mixture of Normals [22]:

$$p(x) = \int p_x(\sigma_x) N(0, \sigma_x^2) \, d\sigma_x \tag{8}$$

$$p(e) = \int p_e(\sigma_e) N(0, \sigma_e^2) \, d\sigma_e \tag{9}$$

where $N(\mu, \sigma^2)$ is a Normal distribution with mean μ and variance σ^2 and we choose $p_x(\sigma) = p_e(\sigma) = Ga(\sigma, \ldots)$ such that $p(x)$ and $p(e)$ are Student-T distributed.

Our approach is to sample values for the vector valued quantity, $[\sigma_x, \sigma_e]$ from their priors such that, conditional on this sample, we have a Normal distribution for the vector valued quantity $[x, e]$. We can then represent the uncer-
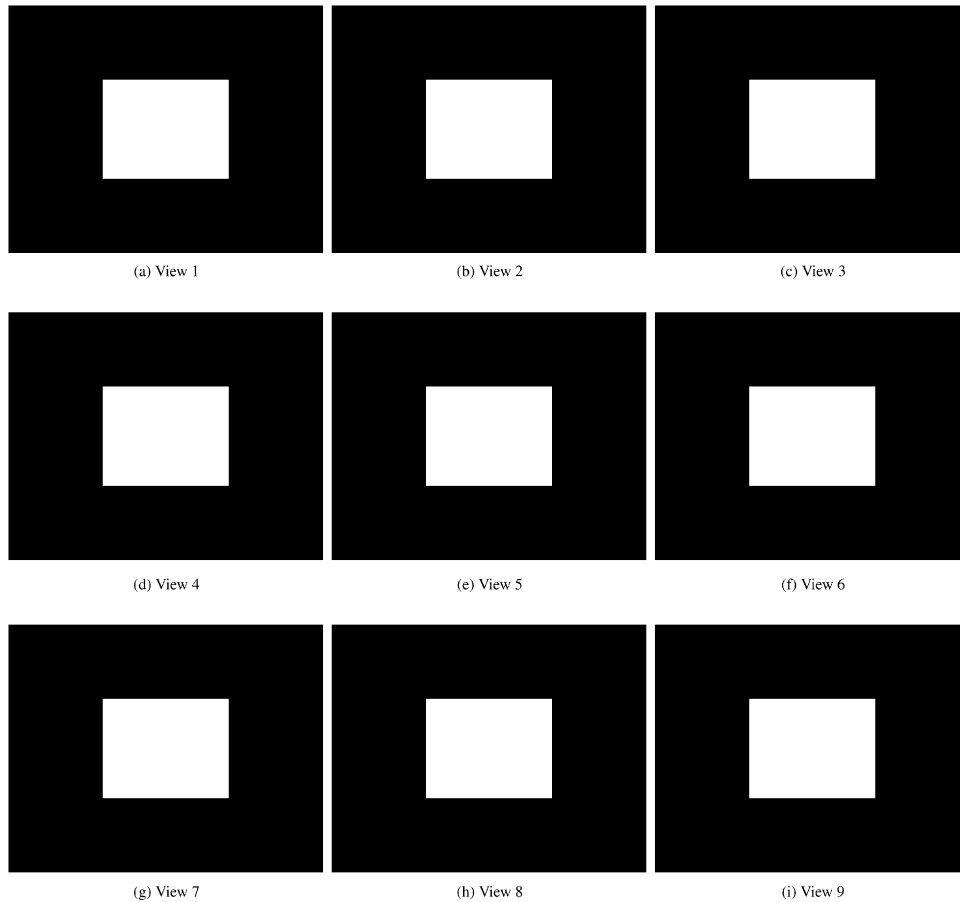
Fig. 22. Templates for sphere considered in Section 3.6.

tainty over the probability associated with this vector real-valued quantity using a mixture of these Normal distributions. The weights of the mixture components and the posterior values for the mean and variance of the components are then calculated using a standard Kalman filter.

Fig. 15 shows the components' mixture weights sorted in order of increasing weight. Fig. 16 shows the values of σ_x and σ_e for these components (sorted in the same order). Note that there is a trend for the components with the high weight to have smaller scales but that there is not a strong preference as to which process caused the outlier; the conflict regarding the potential causes for the outlier is represented. Finally, to emphasise that this process is representing uncertainty over vector real-valued quantities, Fig. 17 and 18 respectively show the prior distribution over the joint space of $[x, e]$ for the three components with the highest weights and the three components with the lowest weights.[10] Note that the components with high weight have a large variance in one direction and that the components with low weight all have low variances in both directions.

3.6. ATR fusion

In the last set of results, we consider a challenging unclassified automatic target recognition (ATR) task similar to that considered previously [24]. We observe imagery (silhouettes) of a target that is one of: cone, hemisphere, sphere or cylinder. There are viewpoints where all four classes project to a circle on the image plane. We assume we know the azimuth, elevation and range of the target and that the classes are such that the objects project to the same circle at these viewpoints. This scenario is designed such that given imagery of a circle, we cannot identify the target: it is only when the target changes orientation that we can potentially identify the target.

We generate nine points uniformly over the surface of a unit sphere[11] and use the resulting points to define look directions. For each look direction and for each class, we

---

[10] The careful reader will note that, while in this specific example, the posterior is nonzero on a scalar subspace of the vector since the posterior is nonzero on the line $y = x + e$, this is a feature of the specific example and the approach can readily be used in higher-dimensional vector valued problems such as those previously considered in a tracking context [23].

[11] We sample the points randomly over the surface of a sphere and then iteratively adjust the positions of the points. Each pair of points mutually repel one another with a force that is aligned with the vector between them and decays with the square of the distance that the points are apart. The points are constrained to move on the unit sphere. The procedure terminates when the distance moved by any point is less than a given small distance.

(a) View 1   (b) View 2   (c) View 3

(d) View 4   (e) View 5   (f) View 6

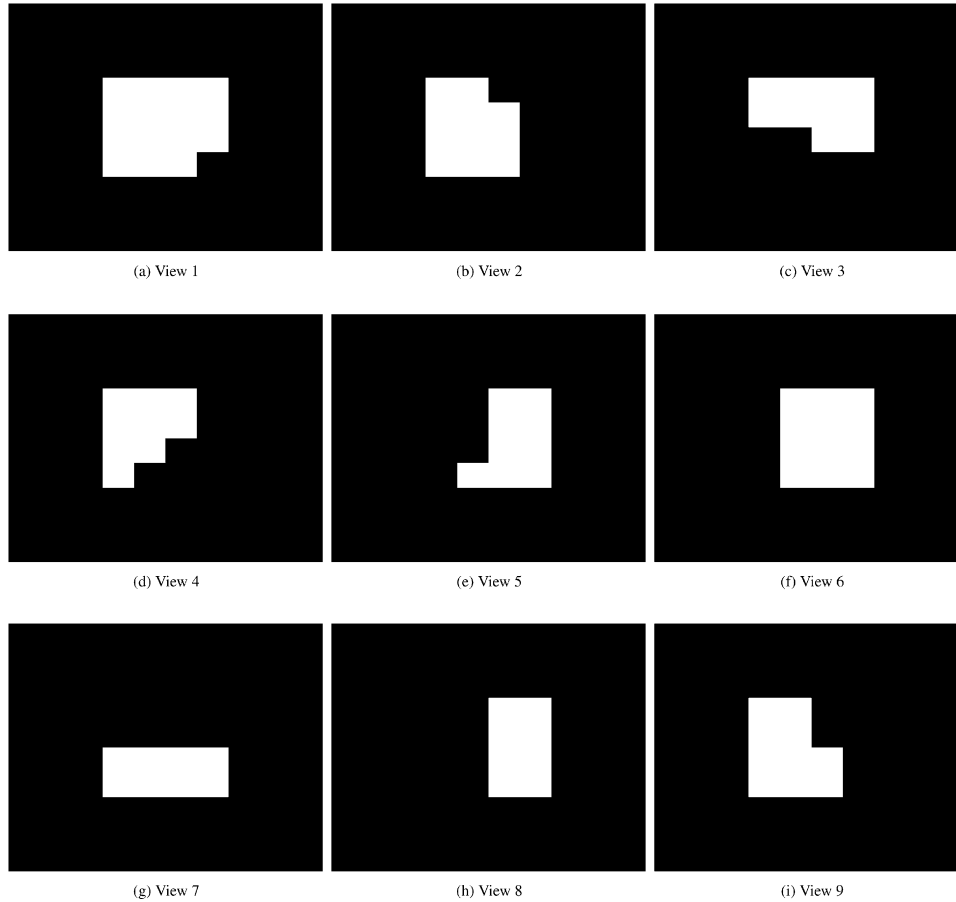(g) View 7   (h) View 8   (i) View 9

Fig. 23. Templates for hemisphere considered in Section 3.6.

generate a template silhouette. This template library is available to the classification algorithm. Exemplar templates used are shown in Figs. 20–23.

We do not model the error process using the sum squared difference for all pixels in the silhouette proposed in [25] and used in [24]. Instead, we assume that the vector of pixel values comprising the image, d, is a non-linear function of the look angle, θ, but a linear function of the derived template silhouette, G_θ plus zero-mean Gaussian noise, e:

$$d = AG_\theta + e \qquad (10)$$

From this model, by putting a uniform prior on A and a Jeffrey's prior on the variance of e, one can derive the following posterior:

$$p(\theta|d) \propto \frac{1}{|G_\theta^{\mathrm{T}} G_\theta|} \times (d^{\mathrm{T}} d - d^{\mathrm{T}} G_\theta (G_\theta^{\mathrm{T}} G_\theta)^{-1} G_\theta^{\mathrm{T}} d)^{-\left[\frac{N^2}{2}+1\right]} \qquad (11)$$

where d is assumed to be the vector of pixel values for an $N \times N$ pixel image. We find the principled derivation of the likelihood appealing and have found experimentally that it outperforms the sum squared difference approach.

Note that if the templates G_θ are normalised to have unit energy (such that $G_\theta^{\mathrm{T}} G_\theta = 1$) then (11) is maximised at the same point as a correlator that calculates $d^{\mathrm{T}} G_\theta$. However, in contrast to such correlators, (11) can be considered to be a likelihood (with respect to θ), making it possible to fuse independent measurements by simply multiplying the likelihoods.

One way to perform ATR in this scenario is to consider a hidden Markov model (HMM) with hidden states that relate to each of the sampled look directions.

We consider an application where the system is provided a sequence of identical images which are all low-noise. The image used, simulated from a cone viewed from an angle near to those that would project to a circle, is shown in Fig. 19.

The results obtained from applying a HMM to this problem of fusing data from the 10 time steps are shown in Fig. 24 where nine Monte-Carlo runs are shown (with different template sets). The elements of the transition matrix used are calculated from considering each state to correspond to a point on the surface of a sphere. A random walk over this surface is then used to calculate the transition probabilities. The intensity of the random walk is such
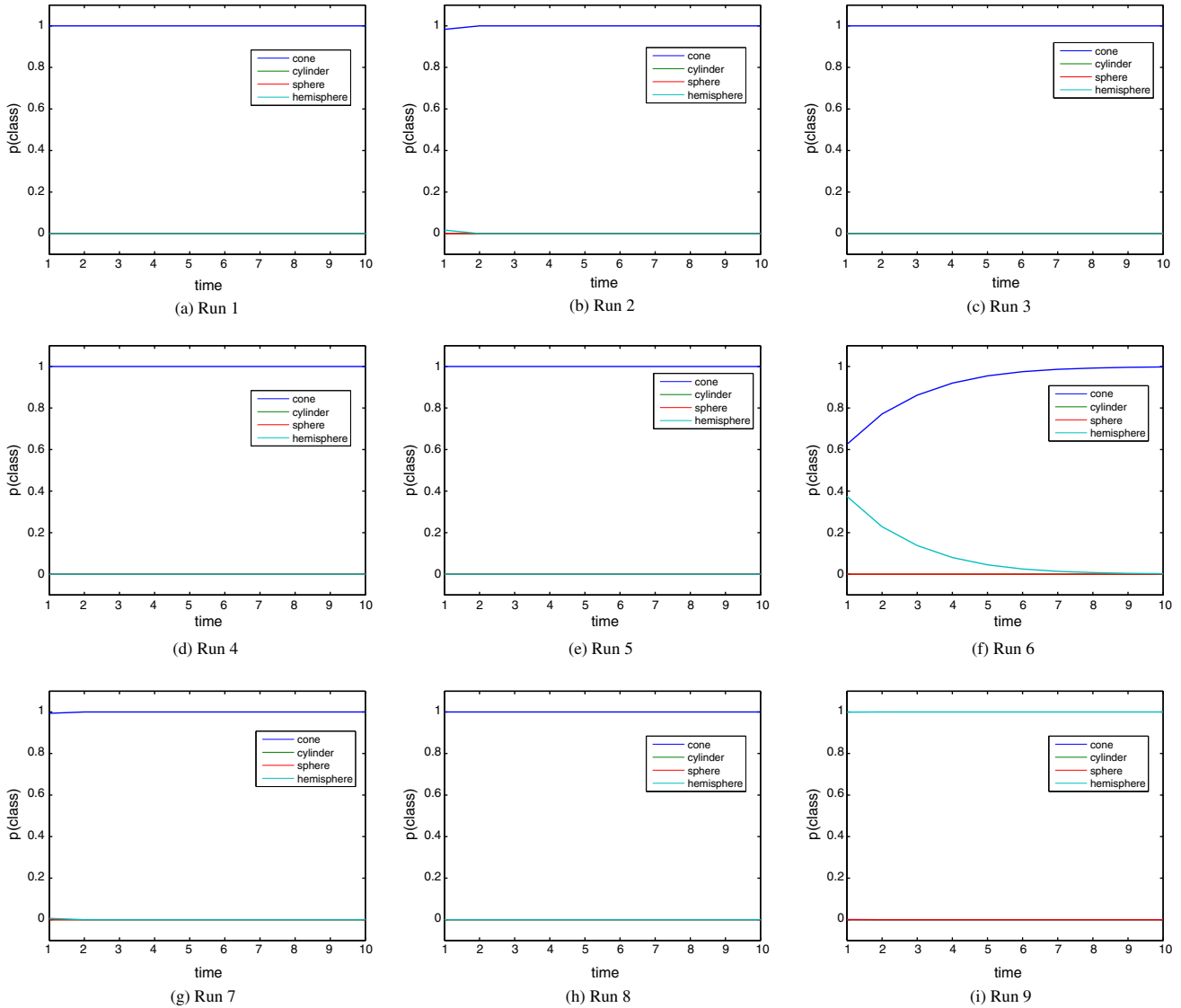
Fig. 24. Fusion output using a HMM considered in Section 3.6.

that the standard deviation of the change in viewing angle between each of the 10 time steps is 0.1°.

Note that, in run 6, the fusion of the images results in an increasingly confident classification output. This contradicts intuition since there is little new information contained in the last nine images. This comes about because the HMM is approximating the difference between the observed image and the templates as noise. In fact, in this scenario, the errors are dominated by the disparity between the look directions for which the templates are defined and the look direction associated with the imagery.

To model this quantisation error, we consider the look direction to be a continuous (multivariate) variable defining the look direction, rather than the discrete variable used in the HMM approach. We consider each of the template silhouettes as being associated with a value of this continuous variable. We therefore pose the problem in

terms of a regression from look direction to observed imagery.

To model the dependence of the imagery on the look direction, we use a Gaussian process [26]. A Gaussian process is simply a generalisation of a multivariate Gaussian distribution to an infinite set of variables, each of which is associated with a continuous value of the look direction. The covariance structure of the variables is then parameterised succinctly. In this specific application, the joint distribution of a pixel value for two look directions, g_{θ_1} and g_{θ_2} is:

$$p(g_{\theta_1}, g_{\theta_2}) = \mathcal{N}\left(\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, s \begin{bmatrix} 1 & \exp\left(-\frac{|\theta_1-\theta_2|}{\sigma}\right) \\ \exp\left(-\frac{|\theta_1-\theta_2|}{\sigma}\right) & 1 \end{bmatrix}\right)$$

(12)

where $\mathcal{N}(\mu, \Sigma)$ is a multivariate Gaussian with a mean of μ and a covariance of Σ, σ is a scaling parameter in distance
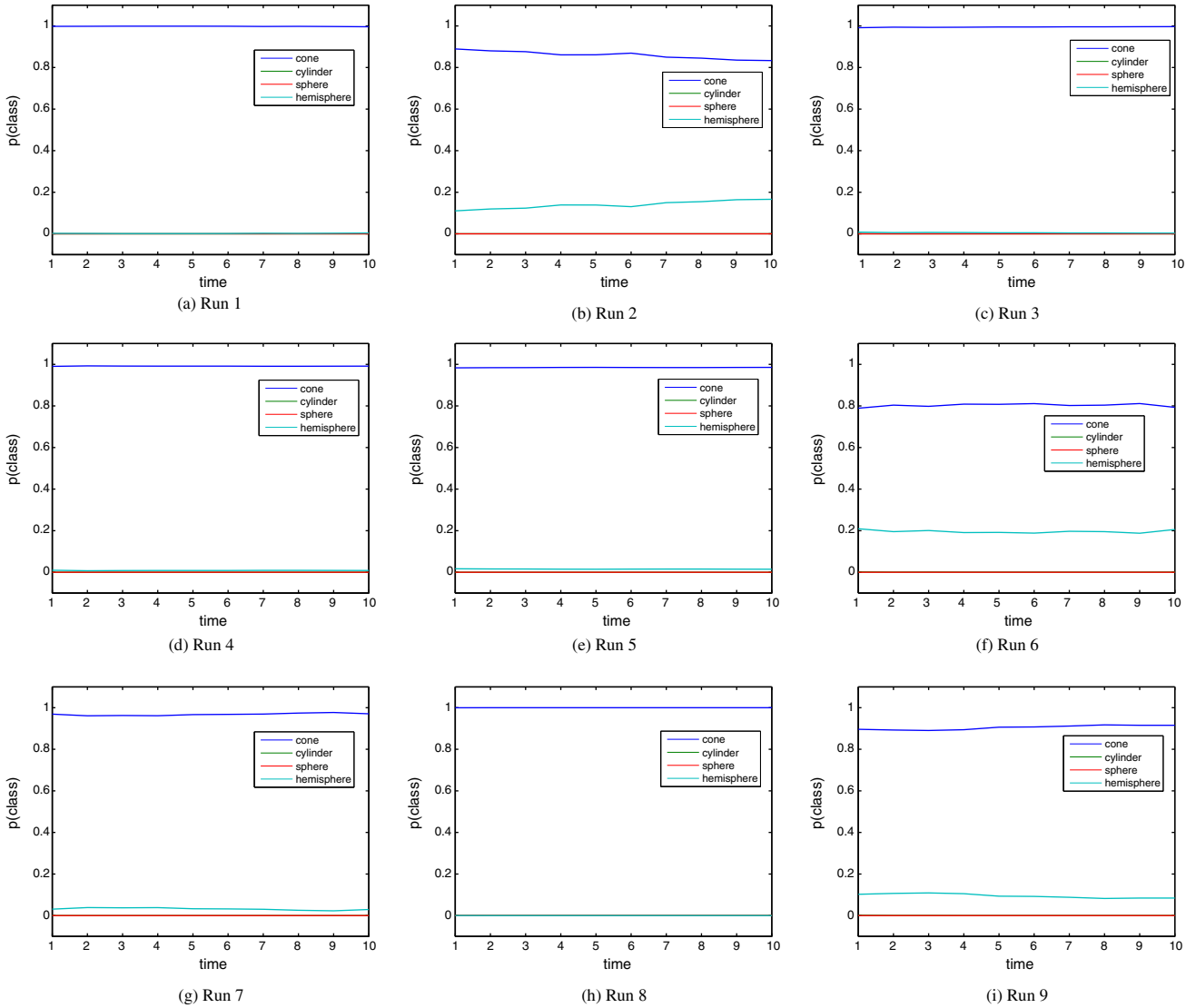
Fig. 25. Fusion output using a particle filter considered in Section 3.6.

between points and $|\Delta|$ is the size of the vector Δ (calculated as the angle between the two look directions). The images that have been shown all have zero entries where the image is black and one where the image is white. s is chosen such that, in the presence of no other knowledge, the prior for the pixel in the image has a covariance equal to that of the template images.

We can then form a joint distribution on an unknown g_θ and the template silhouette's pixels $g_{\theta_1} \ldots g_{\theta_N}$. Hence, we can produce a distribution for $p(g_\theta | g_{\theta_1} \cdots g_{\theta_N}, \theta_1 \cdots \theta_N, \theta)$. So, given the templates, their associated look angles and an unseen look angle, we can produce a distribution on the template for this unseen look angle. Note that the pixels comprising the image are modelled as being the result of independent Gaussian processes (with the same spatial statistics) and that we apply a nonlinear map (based on hyperbolic tangent) to image intensities (to cater for the fact that the intensities are binary in the templates.

We apply this technique with a value for σ of 45°; this is the scale of look directions over which we assume the images are constant and is much bigger than our assumed change in aspect between images.

We use an SIR particle filter to perform inference with 100 particles with the likelihood defined by (12) and the same dynamics as used to define the HMM.[12]

Brief pseudo-code (assuming we have the same template library, as used by the HMM, i.e. templates, $T_L(c)$, for each class, c, associated with each of a number of known look directions, θ_L) for the particle filter implementation is as follows:

---

[12] The reader interested in understanding the details of how to implement a particle filter with such models is referred to [10] and the many other tutorials on the subject.

- FOR each particle, $i = 1 \ldots P$
  - Initialise particle's look directions with θ_0^i (uniformly distributed over sphere)
  - Initialise particle's template library, $T_0^i = \emptyset$
  - FOR each class, $i = 1 \ldots C$
    * Initialise weights, $w_0^{i,j} = (PC)^{-1}$
  - END FOR

- END FOR
- FOR each timestep, $t = 1 \ldots T$
  - FOR each particle, $i = 1 \ldots P$
    * Sample look direction, $\theta_t^i \sim p(\theta_t | \theta_{t-1}^i)$
    * Sample a class, $c^\star$, uniformly
    * Form Gaussian process Covariance using (12) for template seen from θ_t^i given $\theta_{1:t-1}^i$ and θ_L
    * Sample $T^\star$ from Gaussian process (using covariance, T_{t-1}^i and $T_L(c^\star)$)
    * Augment template library: $T_t^i = \{T^\star, T_{t-1}^i\}$
    * FOR each class, $c = 1 \ldots C$
      - Evaluate Gaussian process, p_c, for $T^\star$ given $T_L(c)$ and T_{t-1}^i
      - Calculate likelihood, l, using (11)
      - Calculate weight as $w_t^{i,j} = w_{t-1}^{i,j} \frac{p_c l}{p_{c^\star}}$
    * END FOR

  - END FOR
  - Normalise weights
  - Output classification probabilities as $p(c|y_{1:t}) \approx \sum_i w_t^{i,j}$
  - Resample if necessary
- END FOR

The results are shown in Fig. 25. It is clear that the technique addresses the concern with the HMM; a Bayesian approach that models the quantisation error does not change its classification probabilities so significantly as new measurements with little new information content are received.

4. Conclusions

It has been shown that a Bayesian approach can fuse uncertain, imprecise and conflicting information. Examples have emphasised the importance of model definition.

References

[1] L. Zadeh, Fuzzy logic and approximate reasoning, Synthese 30 (1975) 407–428.

[2] T. Bayes, An essay toward solving a problem in the doctrine of chances, Philos. Trans. Roy. Soc. Lond. 53 (1764) 370–418.

[3] L. Zhang, Representation, independence, and combination of evidence in the Dempster–Shafer theory, in: R.R. Yager, J. Kacprzyk, M. Fedrizzi (Eds.), Advances in the Dempster–Shafer Theory of Evidence, John Wiley and Sons Inc., New York, 1994, pp. 51–69.

[4] Ph. Smets, R. Kennes, The transferable belief model, Artif. Intel. 66 (2) (1994) 191–234.

[5] F. Smarandache, J. Dezert (Eds.), Applications and Advances of DSmT for Information Fusion, Am. Res. Press, Rehoboth, 2004.

[6] F. Smarandache, Unification of fusion theories, Int. J. Appl. Math. Stat. 2 (2004) 1–14.

[7] R.T. Cox, Probability, frequency, and reasonable expectation, Am. J. Phys. 14 (1946) 1–13.

[8] R. Mahler, Can the Bayesian and Dempster–Shafer approaches be reconciled? yes, in: Proceedings of International Fusion Conference, 2005.

[9] A. Gelman, The boxer, the wrestler, and the coin flip: a paradox of robust bayesian inference and belief functions, The American Statistician 60 (2006) 146–150.

[10] A. Doucet, J.F.G. de Freitas, N.J. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Springer, New York, 2001.

[11] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 1999.

[12] A. O'Hagan, J. Oakley, Probability is perfect, but we can't elicit it perfectly, Reliab. Eng. Syst. Safe. 85 (2004) 239–248.

[13] L. Zadeh, On the validity of Dempster's rule of combination of evidence, Memo M 79/24, 1979.

[14] Ph. Smets, The nature of the unnormalized beliefs encountered in the transferable belief model, in: Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence, 1992, pp. 292–297.

[15] J. Dezert, Foundations for a new theory of plausible and paradoxical reasoning, Inform. Security, Int. J. 9 (2002).

[16] R. Haenni, Shedding new light on Zadeh's criticism of Dempster's rule of combination, in: Proceedings of International Fusion Conference, 2005.

[17] B. Ristic, Ph. Smets, Kalman filters for tracking and classification and the transferable belief model, in: Proceedings of International Fusion Conference, 2004.

[18] N. Gordon, S. Maskell, T. Kirubarajan, Efficient particle filters for joint tracking and classification, Proc. SPIE 4728 (2002) 439–449.

[19] S. Maskell, Joint tracking manoevring targets and classification of their maneovrability, EURASIP JASP 15 (2004) 2339–2350.

[20] N. Chopin, A sequential particle filter method for static models, Biometrika 89 (2002) 539–552.

[21] Ph. Smets, Belief functions on real numbers, Int. J. Approx. Reason. (2004).

[22] D.F. Andrews, C.L. Mallows, Scale mixtures of normal distributions, J. Roy. Stat. Soc., Ser. B 36 (1974) 99102.

[23] S. Maskell, G. Gordon, N. Everett and M. Robinson, Tracking manoeuvring targets using a scale mixture of normals, in: Proceedings of Signal and Data Processing of Small Targets, SPIE, 2004.

[24] P. Minvielle, A. Marrs, S. Maskell, A. Doucet, Joint target tracking and identification part II: Shape video computing, in: Proceedings of International Fusion Conference, 2005.

[25] J. Deutscher, A. Blake, I. Reid, Articulated body motion capture by annealed particle filtering, in: Proceeedings of CVPR, 2000.

[26] D.J.C. MacKay, Introduction to Gaussian processes, in: C.M. Bishop (Ed.), Neural Networks and Machine Learning, NATO ASI Series, vol. 168, Springer, Berlin, 1998, pp. 133–165.