Comparison of Fusion Methods for Successive Declarations of Radar Range Profiles

T. Bieker
Research Institute for High Frequency Physics and Radar Techniques (FHR)
Research Establishment for Applied Sciences (FGAN)
Neuenahrer Straße 20, Wachtberg-Werthhoven, Germany
bieker@fgan.de

Abstract - Classification of High-Range-Resolution profiles is a viable method of non-cooperative target identification. In order to increase the reliability and robustness of the classification result, methods of decision-level identity fusion can be applied. Different approaches have been used for a cumulative fusion of declarations of successively recorded radar range profiles. Besides probabilistic techniques such as the Bayesian fusion, non-probabilistic methods based on Dempster-Shafer or voting algorithms have come into focus. In this paper these different approaches are compared for typical situations which can arise in aircraft identification scenarios.

Keywords: Target identification, Evidential Reasoning, Bayes fusion, Voting fusion.

1 Introduction

In recent years there has been an increasing request for multi-sensor data fusion for target identification in order to guarantee an identification result of high confidence and to become more independent of the failure-free function of the single sensors. But even for a single sensor it is reasonable to stabilize the classification result by fusing successive reports. Instead of $N$ different sensors one has the declarations of a classifier at $N$ time steps. In this paper we consider the successive fusion of classifier decisions for one-dimensional high-resolution radar signatures of air targets, so-called high-resolution range profiles (short HRR profiles), for non-cooperative target identification. A HRR profile represents a projection of the target’s radar scattering structures onto the radar line of sight. It contains information about the magnitude of the backscattered radar radiation as a function of range and aspect angle. With an appropriate and extensive database, they are most suitable for discriminating between different fighter aircraft types. We use a simple 1-Nearest-Neighbor classifier [1] which is based on calculating the correlation coefficient for a given test profile and all reference profiles in the data base. The maximum of the correlation coefficient determines the class, to which the given test profile is assigned. However, since we consider decision-level fusion, the results of this investigation apply to all classifiers which provide a rating of classes as a result. As fusion methods we apply Evidential Reasoning, Bayes fusion as well as Voting fusion. These methods have been subject to various investigations in the past [2, 3, 4], however, the focus of these papers was primarily on convergence comparisons of the different methods. An interesting aspect is to analyze the behavior of the fusion algorithms in critical situations. Typical examples for critical situations are given if

1. if there are several competing classes. How easily is a trend established by the fusion algorithms?

2. misclassifications occur. How stable are the fused reports?

3. the classifier changes its overall preference from one class to another. How long does it take for the fusion algorithm to adapt to the new situation?

It is an essential task to find a fusion technique which is robust against single misclassifications on the one hand, but on the other hand reacts in time if there is a change of class preference, i.e. the fusion report has to reflect the classifier output on an optimum time scale. For this purpose we have tested many time series of range profiles of our data base. However, for a systematical investigation of the behaviour of the different fusion methods it is not sufficient to apply the algorithms to “typical” time series of declarations as they are produced by applying the classifier to range profiles contained in our data base. In order to analyze the behavior in critical situations and the influence of parameters as, for example, the likelihood matrix, it is necessary to construct series of mass numbers and declarations as well as likelihood matrices.

In Section 2 we give an overview of the different fusion methods and explain how they are applied to the correlation classifier results for HRR profiles. Section 3 provides a comparison of the fusion algorithms for certain typical situations which occur in HRR profile classification; in the first part we present fusion examples for series of profiles of our data base and in the second part we investigate the fusion of constructed declaration series. These results are summarized in Section 4.
2 Fusion Methods

2.1 Bayes Fusion

In the framework of Bayes theory, probability measures a degree of belief that a certain situation will occur [5, 6]. The fundamental assumption of Bayes theory is that we have a probability distribution, i.e. a probability measure of all events which can occur in the situation under consideration. The atoms of this sample space are pairwise incompatible, and together they describe the complete situation. The well-known Bayes rule for the probability that hypothesis \( H \) is true, given evidence \( E \), reads

\[
P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)},
\]

where \( P(E|H) \) is the likelihood of the occurrence of evidence \( E \), given hypothesis \( H \), \( P(H) \) is the a-priori probability (prior) of the hypothesis \( H \), and the total probability of evidence \( E \) is \( P(E) \). For a sequence of \( N \) pieces of evidence, \( E_1, \ldots, E_N \), and a number of \( M \) hypotheses \( H_1, \ldots, H_M \), this becomes

\[
P(H_i|E_1, \ldots, E_N) = \frac{P(E_1, \ldots, E_N|H_i) \cdot P(H_i)}{P(E_1, \ldots, E_N)}. \tag{2}
\]

Note that in this formulation of the Bayes rule the stochastic independence of the pieces of evidence, \( E_1, \ldots, E_N \) as well as the conditional independence of the pieces of evidence, \( E_1, \ldots, E_N \), given the hypothesis \( H_i \), are not required. If we do assume both independencies as given, equation 2 reads

\[
P(H_i|E_1, \ldots, E_N) = \frac{P(H_i) \prod_{n=1}^{N} P(E_n|H_i)}{\sum_{m=1}^{M} P(H_m) \prod_{n=1}^{N} P(E_n|H_m)}. \tag{3}
\]

2.1.1 Application of Bayes theory to the classification of HRR profiles

Transferred to the framework of target identification, the declaration \( D_i \) of a HRR profile of the target type \( i \) by the classifier to class \( j \) at time step \( n \) represents an evidence \( E_n \), while the hypothesis \( H \) denotes the actual target type. If we have \( N \) successive HRR profiles, and thus \( N \) class declarations, we can combine these pieces of evidence under the assumptions that the target under consideration remains the same (hypothesis \( H_i \) is constant), and that the evidences \( E_1, \ldots, E_N \) are independent, by equation 3. Even if the range profiles of two successive time steps are considered to be decorrelated, the assumption of conditional independence is still a sensitive point. The maximum posterior probability of a target \( i \) is 

\[
L_{ij} = P(E_i|D_j),
\]

which contains the likelihoods that a target of type \( i \) is declared as a member of class \( j \) by the classifier. One must bear in mind, though, that in principle \( L_{ij} \) has to be computed for each aspect angle. We have computed this matrix from our HRR data base of fighter aircraft in order to obtain typical values for \( L_{ij} \) and for azimuth angle windows of 10°.

A further point to consider is the occurrence of zero elements in \( L_{ij} \), which causes problems in the application of equation 3. A zero element of \( L_{ij} \) means that a declaration of a target \( i \) as class \( j \) is considered as impossible. As a result, each \( P(H_i|E_1, \ldots, E_N) \), where at least one "impossible" declaration occurs, should equally become zero. This is a consequence of the assumption of conditional independence, which replaces the expression \( P(E_1, \ldots, E_N|H_i) \) by the product \( P(E_1|H_i) \cdot \ldots \cdot P(E_N|H_i) \). Since the likelihood matrices computed from our database are mainly occupied by the diagonal elements and have actually many zero off-diagonal elements, many declaration series would be cut-off after a few time steps. In order to avoid the cut-off of declaration series by zero \( L_{ij} \) elements, we choose a more careful approach by keeping a very low basis likelihood for any declaration. Thus \( P(H_i|E_1, \ldots, E_N) \) will approach (but not become equal to) zero for successively unlikely declarations and can be increased again by declarations of higher likelihood. After adding a basis likelihood of 0.01 to the computed likelihoods, the new likelihood matrix has to be renormalized in order to yield a column sum of one.

2.2 Dempster-Shafer method

Although with the Bayes approach we have a classical and simple fusion method, it has several shortcomings: The hypotheses have to be mutually exclusive and they have to describe the given scenario completely. This implies that uncertainty cannot be expressed in this framework, because \( P(x) + P(\neg x) = 1 \) for any \( x \). The main disadvantage is, however, that with the knowledge of the complete likelihood matrix it requires a huge amount of input information which is almost impossible to gather in reality. For practical purposes this represents a serious limitation. An approach to circumvent these difficulties is given by Evidential Reasoning, which has been proposed as an alternative to Bayes theory by Shafer [7], based on ideas first developed by Dempster [8]. Similar to Bayes theory, the Dempster-Shafer (DS) approach uses degrees of belief in order to express uncertainty numerically. The difference in DS theory is that belief is assigned to propositions which are built by sets of hypotheses, and that these sets of hypotheses are not required to have non-zero intersections.

Only the atoms of the considered scenario are represented by mutually exclusive hypotheses, which build together the frame of discernment \( \Theta = \theta_1, \ldots, \theta_N \). The total belief on \( \Theta \) can be partitioned into portions and assigned to the propositions, i.e. the subsets of \( \Theta \). Since the number of all subsets of \( \Theta \) is \( 2^\Theta \), a mass distribution function \( m: 2^\Theta \to [0, 1] \) is defined by

\[
m(\emptyset) = 0, \tag{4}
\]

\[
\sum_{A \subseteq \Theta} m(A) = 1, \tag{5}
\]

where \( m(A) \) is the mass assigned to the proposition \( A \).
which builds a basic belief assignment (bba). The bba number \( m(A) \) is the measure of belief that is associated with the subset \( A \) and that cannot be subdivided any further. All subsets \( A \) of \( \Theta \) with \( m(A) \neq 0 \) are called focal elements of the belief function. A special proposition is given by \( \theta_1 \lor \theta_2 \lor \ldots \lor \theta_N \). If we assign belief to the disjunction of all elements of \( \Theta \), this belief is called uncommitted and expresses that we know that the truth is somewhere in the frame of discernment but we cannot assign it to any of its subsets, i.e. we cannot specify our belief any further. This gives us the possibility to express a general level of uncertainty.

The function \( \text{Bel} : 2^\Theta \rightarrow [0, 1] \)

\[
\text{Bel}(A) = \sum_{B \subseteq A} m(B) \tag{6}
\]

defines the \textit{Belief} in a subset \( A \) as a sum of all the basic belief masses that support its constituents. As a dual measure to the Belief in \( A \), the function

\[
\text{Pl}(A) = 1 - \text{Bel}(\neg A) = \sum_{B \cap A \neq \emptyset} m(B) \tag{7}
\]

is defined, which is called \textit{Plausibility} and gives the extent to which proposition \( A \) might be true. The functions Bel and Pl define an interval \([\text{Bel}(A), \text{Pl}(A)]\) which is considered as a measure of ignorance about \( A \).

Two pieces of evidence with mass numbers \( m_1 \) and \( m_2 \) can be combined by Dempster’s combination rule

\[
m(C) = k^{-1} \sum_{A \cap B = C} m_1(A) \cdot m_2(B) \tag{8}
\]

with

\[
k = \left[ 1 - \sum_{A \cap B = \emptyset} m_1(A) \cdot m_2(B) \right]^{-1} \tag{9}
\]

If the two pieces of evidence do not contradict each other, these are said to be concordant, and we have \( k = 1 \), otherwise they are conflicting, which implies \( k > 1 \).

For our purpose we consider simple belief functions which build a special class of belief functions. A simple support function supports one and only one subset of the frame of discernment. \( s \) is a simple support function focused on \( F \subseteq \Theta \) such that \( s(\emptyset) = 1 \) and

\[
s(A) = \begin{cases} s, & \text{if } F \subseteq A, \ A \neq \emptyset \\ 0, & \text{otherwise} \end{cases} \tag{10}
\]

With this definition we have \( m(F) = s(F) \) and \( m(\emptyset) = 1 - s(F) \) for the basic belief assignment.

### 2.2.1 Application of Dempster-Shafer theory to the classification of HRR profiles

A problem in DS theory is the determination of a suitable basic belief assignment. By using a 1-NN classifier, the test profile \( x \) is assigned exactly to that class whose prototype has the minimum distance to \( x \) in feature space, which is equivalent with the maximum correlation with the test profile. In this way we obtain the correlation coefficient \( r_i \) of a test profile \( x \) and the prototype of class \( C_i \) as a natural measure of similarity between the test profile and the profiles of all classes in the database. Investigations of our extensive HRR data base of fighter aircraft have shown that the correlation coefficient \( r_i \) is always very high, i.e. \( r_i > \alpha \), so we map it to the interval \( \tilde{r}_i = (r_i - \alpha)/(1 - \alpha) \). The DS combination is performed in two distinct steps. First, we have to combine the class declarations for one time step, since we have for each class a piece of evidence in the form of a correlation coefficient. We set \( m_i = m\{C_i\} = \tilde{r}_i \) as the basic belief, that the profile \( x \) belongs to class \( C_i \). The basic belief that the profile \( x \) belongs to any of the \( N \) classes, i.e. the uncommitted belief, is then given by \( m\{\{C_i\}\} = m_{\{i\}} \). If we combine these simple support functions for one time step and \( N \) classes according to Equation 8, we obtain:

\[
M(C_i) = \frac{m_{\{i\}}}{1 + \sum_{k=1}^N m_{\{k\}}} \tag{11}
\]

\[
M(C) = \frac{1}{1 + \sum_{k=1}^N m_{\{k\}}} \tag{12}
\]

For the fusion results of the bbas of two time steps \( n \) and \( n - 1 \) one has (The superscripts denote the time step index):

\[
\tilde{M}^{n+1}(C_i) = \frac{\sum_{A \cap B = C_i} M^{n+1}(A) \cdot M^{n-1}(B)}{\sum_{A \cap B \neq \emptyset} M^{n+1}(A) \cdot M^{n-1}(B)} \tag{13}
\]

where

\[
\sum_{A \cap B = C_i} M^{n+1}(A) \cdot \tilde{M}^{n-1}(B) =
\]

\[
M^{n+1}(C_i) \tilde{M}^{n-1}(C_i) + M^{n}(C_i) \tilde{M}^{n-1}(C_i) + M^{n}(C) \tilde{M}^{n-1}(C_i) \tag{14}
\]

and

\[
\tilde{K}^n = \sum_{i=1}^N M^n(C_i) \left( \tilde{M}^{n-1}(C_i) + \tilde{M}^{n-1}(C_i) \right) + M^n(C) \tag{15}
\]

The Belief and Plausibility function read

\[
\text{Bel}^n(C_i) = \sum_{B \subseteq C_i} \tilde{M}^n(B) = \tilde{M}^n(C_i) \tag{16}
\]

\[
\text{Pl}^n(C_i) = \sum_{B \cap A \neq \emptyset} \tilde{M}^n(B) = \tilde{M}^n(C_i) + \tilde{M}^n(C) \tag{17}
\]

One can show easily that the difference between Bel\(^n\) and Pl\(^n\) decreases in every time step in which a non-zero mass is assigned to a class. Since Pl\(^n\)(C\(_i\)) − Bel\(^n\)(C\(_i\)) = \tilde{M}^n(C), the difference can be computed
iteratively:

\[ \tilde{M}^n(C) = \frac{M^n(C)M^{n-1}(C)}{K^n} \]
\[ = M^1(C) \cdot \prod_{i=1}^{n} \frac{M^i(C)}{K^i}, \quad (18) \]

where \( M^1(C) \) denotes the total mass that is assigned to the frame of discernment in the first timestep. Inserting the expression for \( K^i \), equation ... yields

\[ \tilde{M}^n(C) = M^1(C) \cdot \prod_{i=1}^{n} \frac{1}{1 + Z^i} \quad (19) \]

with

\[ Z_i = \frac{1}{M^i(C)} \sum_{i=1}^{N} M^i(C_i) \left( \tilde{M}^{i-1}(C_i) + \tilde{M}^{i-1}(C) \right). \quad (20) \]

The expression \( Z^i \) is positive and becomes zero if \( M^i(C_i) = 0 \) \( \forall i = 1(1)N \). Thus for any time step, where \( M^i(C) < 1 \), one has \( Z^i > 0 \), so that Bel and Pl are approaching each other.

2.3 Voting fusion

A third, very simple fusion method is given by treating the fusion problem like a democratic process. The identity decisions from \( N \) sensors are simply counted as votes. For the cumulative evaluation of the classifier decision of HRR profiles the number \( N \) of the sensors is replaced by the number of time steps \( N \), over which the

votes for each class are counted. For a memory of only one time step this amounts to evaluating only the actual HRR profile and the result is the same as the actual classifier decision. The average number of votes \( c_m \) for class \( j \) and a memory of \( N \) time steps is given by

\[ c_m(t, N) = \frac{1}{N} \sum_{n=0}^{N-1} d_m(t - n), \quad t \geq N, \quad (21) \]

where \( d_m(t) \) is 1 if in time step \( t \) the target is declared as class \( m \) and 0 otherwise.

3 Fusion results

3.1 Results for exemplary time series of declarations of measured range profiles

The section of our HRR data base appropriate for testing fusion algorithms contains range profiles for eight different classes of fighter aircraft. For each aircraft type there is data obtained in radar measurements of different flights so that a given test range profile can be classified with reference data from all flights excluding the flight of the test profile itself. In this way independence of test and reference data is guaranteed. We select series of successive test profiles for a selected target type from one flight, so that the azimuth spacing of the profiles is about 1°. The correlation coefficients are calculated for all combinations of one test profile with all reference profiles and the maximum value for each target class is taken. As described in section 2.2.1, we scale the correlation coefficient to the interval \([0, 1]\) in order to obtain mass numbers for the Dempster-Shafer method. For the voting and the Bayes method only a declaration is required, i.e. the class for which the correlation coefficient is globally maximum. From the mass numbers we compute belief and plausibility according to Equation 16 and 17 and from the declarations we get both the number of hits for each class for the voting method and the posterior for the Bayes method. In the latter case, we use the likelihood matrix suitable for the azimuth value of the given test profile.

In the following diagrams we show (always for those three classes with the highest mean massnumber) i) the mass numbers as well as the results of ii) the Dempster-Shafer approach, iii) the Bayes method and the iv) Voting fusion.

![Figure 1: Mass numbers and fusion results for target F (1)](image)
Figure 2: Mass numbers and fusion results for target D (1)

this collapse is sharper and more pronounced than for the other two fusion algorithms. For the voting technique, the curves of the different classes are not as well-separated as for the other methods. This is due to the arbitrarily long memory time (in this case 44 time steps) which has been chosen in order to make the different techniques comparable.

In Figure 5 the correct target type is class B, which is not obvious from the mass numbers. There is an initial preference for class H which is reflected by all three fusion techniques. After time step 20, the preference for class H becomes weaker and a concurrence between classes A and B emerges. While the Dempster-Shafer algorithm supports class A for 13 time steps, the Bayes result prefers class A only for four time step and the voting algorithm does not support class A at all.

Figure 3: Mass numbers and fusion results for target D (2)

Figure 4: Mass numbers and fusion results for target A (1)

The striking point in Figure 4 is that there are misclassifications between timesteps 20 and 35, which affect only the fusion results of the Dempster-Shafer and the Voting algorithm, while the Bayes fusion result sticks to class A without any doubt.

Figure 5: Mass numbers and fusion results for target B (1)
3.2 Results for constructed time series of declarations

Although the first part of our results has shown that all three fusion techniques work well in principle, we have so far analyzed their behavior only by examples. However, since a more systematic examination is required, in this part we discuss the results for representative constructed mass number and declaration series. In order to keep this analysis as concise as possible we restrict our investigations to three classes.

3.2.1 Competing Classes

As a typical likelihood matrix we use

\[
L = \begin{pmatrix}
0.90 & 0.05 & 0.05 \\
0.05 & 0.90 & 0.05 \\
0.05 & 0.05 & 0.90
\end{pmatrix}.
\]

(22)

The mass numbers shown in Figure 6 vary randomly between 0.6 and 0.7, and there is no obvious preference for one of the three classes. The averages of the mass numbers are 0.655, 0.658, 0.650 for class 1, 2, 3, respectively, so that class 2 gains the highest average number of hits. The results of the fusion algorithms seem to reflect this fact, because in all three cases class 2 is preferred for the largest number of time steps.

Figure 7: Random mass numbers and fusion results (2)

With a likelihood distributed more homogeneously between all elements, this effect becomes weaker.

3.2.2 Misclassifications

The investigation of our HRR profile data base have shown that even in a series of many correct declarations misclassifications can occur. This raises the question, how stable the fusion algorithms are against misclassifications. Figures 8 and 9 show situations where class 3 is assumed to be the correct class and where nine erroneous assignments to class 1 occur.

Figure 8: Mass numbers and fusion results for a declaration series which contains nine misclassifications at time steps 10 to 18.

In the first case the target is classified correctly for the first nine time steps. The misclassifications lead to a collapse of the fusion curves for class 3, but not to a switch to class 1. In the second situation, however, the target has been declared correctly as class 3 for 27 time steps.

Figure 9: Misclassification results (3)
steps, before it is misclassified. In this case both the Dempster-Shafer as well as the Bayes results remain unaffected, whereas the fusion result of the voting algorithm reacts immediately with a slow decrease of the class 3 curve and an equally slow increase of the class 1 curve. Obviously the impact of misclassifications depends strongly on the number of correct classifications which preceded the wrong declarations.

### 3.2.3 Change of class preference

This leads to the further question, how long it takes for the fusion algorithms to notice that the classifier has changed its preference.

Considering Figure 10, class 3 is preferred for the first 15 time steps, after which preference switches to class 1 for the next 16 steps and then back to class 3 again for the rest of the time series. The Dempster-Shafer fusion result is not affected by the temporary class preference change apart from a very shallow decrease of the class 3 curve around time step 30. However, both the Bayes and the voting fusion results for class 3 decrease down to 0.5, but no cross-over to class 1 takes place.

This is different, if the temporary preference change is slightly longer (18 time steps) than the time for which class 3 has been preferred (Figure 11). A cross-over to class 1 occurs for three time steps for the voting and the Bayes result, while the Dempster-Shafer
result is still hardly affected. The Dempster-Shafer result shows for three time steps a cross-over of the class 1 and class 3 curves for the given mass numbers, if the change exceeds 21 time steps (Figure 12). For lower mass numbers, the cross-over disappears again (Figure 13).

This and further investigations show that in general a preference change in the Bayes and voting fusion results will not occur unless the number of declarations for the class preferred after the change exceeds the number of classifications for the class preferred before the change. However, for the Dempster-Shafer algorithm, this reaction time depends also on the relative level of the mass numbers of the competing classes.

4 Summary

In this paper we have investigated the behavior of Dempster-Shafer, Bayes and voting fusion techniques for successive declarations of HRR profiles. We have analyzed typical declaration series of real measured HRR profiles in an extensive data base of fighter aircraft profiles. The declarations are produced by a 1-Nearest-Neighbor classifier based on the maximum correlation coefficient of a test HRR profile and the prototype profile of each class. The mass numbers for the Dempster-Shafer algorithm are obtained by a linear transformation of the maximum class-specific correlation coefficient. For the Bayes fusion algorithm, we use an aspect-angle dependent likelihood matrix. This examination shows that for the largest part of measured HRR profile series, a clear discrimination between a preferred class and competing classes is established within a few time steps by all three fusion techniques. However, we encounter several critical situations, where the fusion results have to be evaluated with some caution.

In order to study these situations systematically, we have constructed special test declaration series. For closely competing classes, we find that trends are easily established by all three fusion methods, even if the declarations are distributed randomly. A second critical situation is given if misclassifications occur. The fusion results are stable against single misclassifications, if the correct class has been preferred for many time steps before the misclassifications. Once a class A has been preferred for \( N \) time steps, after a change of the classifier’s preference to a class B the fusion algorithms are locked to class A for at least further \( N \) time steps; especially for the Dempster-Shafer approach this reaction time also depends on the value of the mass numbers. Concerning the identification result, i.e. the class which is supported by the fusion algorithms, the three techniques do not differ appreciably. However, in comparison to Bayes and Dempster-Shafer method, voting fusion turns out to be more sensitive to declaration changes. The average number of votes is a measure which reacts immediately if the classifier changes its preference, but without switching its support quickly to the new class.

References