

Effective ATR Algorithms Using Information Fusion Models

Arnaud Martin

ENSIETA, E³I²
2 rue François Verny
29806 Brest, Cedex 9
France

Arnaud.Martin@ensieta.fr

Emanuel Radoi

ENSIETA, E³I²
2 rue François Verny
29806 Brest, Cedex 9
France

Emanuel.Radoi@ensieta.fr

Abstract – Several types of classifiers have been developed in order to extract the information for the automatic target recognition (ATR). We have noted that these performances are different according to the classifier and the radar target. We propose in this article three approaches of information fusion in order to outperform three radar target classifiers. These three techniques of fusion are the Sugeno's fuzzy integral, the possibility theory and the Dempster-Shafer theory. In this application, we show that the best performance is achieved by the Dempster-Shafer theory.

Keywords: ATR, information fusion, fuzzy integral, possibility theory, Dempster-Shafer theory.

1 Introduction

The high resolution radar range profiles are often used for the automatic target recognition (ATR) [1,2,3]. A range profile of a target can be regarded as its one dimensional signature, generated by the electromagnetic reflection of a high-frequency broadband signal. It is obtained as the distribution of the target reflectivity along the line of sight.

The more information is available, the more effective is the radar target recognition process. The ideal solution would be the fusion of information coming from several sensors. Nevertheless, this solution is not acceptable for an autonomous system, of average complexity, as it is the case for the most current radar systems. Another possibility is to extract the information confined by the dataset using different types of classifiers, which are supposed to be partially complementary. The available information being more complete, we can expect better performances in classification by the fusion of the decisions at the outputs of the classifiers.

Fusion makes the system not only more powerful but also more robust because different classifiers have specific discriminating capabilities and robustness with respect to noise according to the angle of sight.

Several methods of fusion are presented in the literature. Thus, [4] proposes an empirical fusion method based on the majority vote rule and a distance criterion. However, the problem is to know how to combine the answers given by various classifiers in order to obtain the best result. According to this idea, other approaches, much

more systematic, are based on the Bayes method, possibility theory, Dempster-Shafer theory, fuzzy logic or fuzzy integral.

The paper is organized as follows. Section 2 describes three classifiers which are aimed to produce the partial decisions required by the fusion system. The basic principles of the fusion techniques considered in this paper are given in Section 3, while Section 4 is dedicated to the range profile database and to the simulation results issued from the application of the fusion algorithms previously presented. Finally, some conclusions are drawn together with the future works we plan to do.

2 Classifiers

Three classifiers based on three different principles have been used for classifying the radar target signatures. They are briefly described in this section.

2.1 Multilayer perceptron classifier

The multilayer perceptron (MLP) is a feedforward fully connected neural network [5,6] having the structure shown on Fig 1.

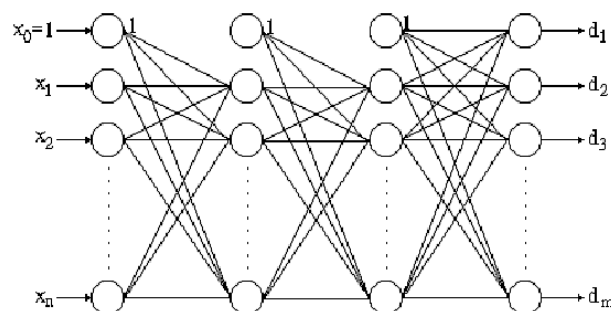


Fig. 1 The multilayer perceptron structure

The data x is described by n coefficients (x_1, \dots, x_n) . Each unit of the network is an artificial neuron (perceptron) with the structure shown on Fig 2.

All the unit outputs of every layer are connected to all the unit inputs of the next layer weighted by the values w_{lj} , where l is the source unit and j is the target unit. These weights are initialized with small random values and they

reach stable values after the learning process if it converges.

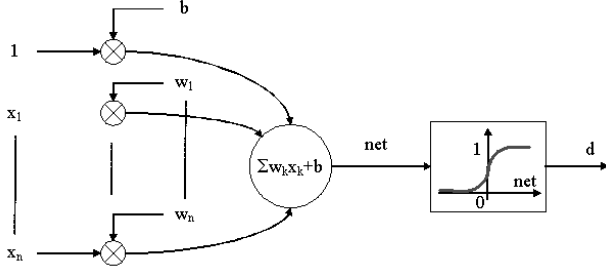


Fig. 2. Artificial neuron structure.

The learning process consists in repeated presentations to the network of the training vector and the corresponding desired output vector. The objective of the learning process is to minimize the quadratic error:

$$\varepsilon = \frac{1}{2} \sum_{j=1}^m (d_j - o_j)^2 \quad (1)$$

where o_j are the real outputs of the multilayer perceptron units.

If the sigmoid function shown on figure 2 is used then the following learning algorithm is obtained:

$$\begin{cases} w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j(t) o_i(t) \\ \delta_j = c o_j [1 - o_j] [d_j - o_j], \text{ for the output layer} \\ \delta_j = c o_j [1 - o_j] \sum_l \delta_l w_{lj}, \text{ elsewhere} \end{cases} \quad (2)$$

The constant c controls the slope of the sigmoid function and η stands for the learning rate.

This rule is known as the back propagation algorithm or the generalized delta rule. Its convergence can be improved if a momentum term is added and the learning rate is varied in an appropriate manner [6].

2.2 Fuzzy K Nearest Neighbor Classifier

The decision rule for the KNN (K Nearest Neighbor) method [7] is very simple and can be easily generalized to an arbitrary number of classes. Thus, if $V_K(\mathbf{x})$ stands for the K^{th} order neighborhood of the vector \mathbf{x} and if:

$$K_j(\mathbf{x}) = \text{card} \{ \mathbf{x}_n \mid \mathbf{x}_n \in C_j, \mathbf{x}_n \in V_K(\mathbf{x}) \} \quad (3)$$

then:

$$K_k(\mathbf{x}) = \max_{j=1, m} [K_j(\mathbf{x})] \Rightarrow \mathbf{x} \in C_k \quad (4)$$

The decision rule consists therefore in classifying the unknown vector \mathbf{x} in the same class as the most of its neighbors belong to.

The integration of the fuzzy logic with the classical KNN method principle results in powerful classifier, which performs better in terms of recognition rate and is more robust to outliers [8]. A central concept for the fuzzy

KNN method is the membership coefficient of a vector to a class. Its value varies between 0 and 1 and can be seen a measure of the vector capability to be more or less representative for a class. The sum of the membership coefficients of a vector to all the classes has to be equal to 1. Defining the membership in this way is less constraining and better matches the physical reality, because the transition from one pattern to another is very often continuous.

The fuzzy KNN classifier needs a training or fuzzyfication stage, when the membership coefficients to each class are calculated for all the training vectors. The following relationship has been used to calculate the membership coefficient of the training vector \mathbf{x}_t to the class C_j :

$$u_{jt} = \frac{K_j^{(t)}}{K_F}, \quad K_j^{(t)} = \text{card} \{ \mathbf{x}_n^{(j)} \mid \mathbf{x}_n^{(j)} \in V_{K_F}(\mathbf{x}_t) \} \quad (5)$$

K_F defines the neighborhood value in the training stage, while $K_j^{(t)}$ represent the number of the nearest neighbors of the vector \mathbf{x}_t belonging to the class C_j . For a test vector \mathbf{x} its membership coefficients to all the classes are firstly calculated using the following relationship:

$$o_j(\mathbf{x}) = \sum_{i=1}^K (u_{jt} / \|\mathbf{x} - \mathbf{x}_i\|^{2/\delta}) / \sum_{i=1}^K (1 / \|\mathbf{x} - \mathbf{x}_i\|^{2/\delta}) \quad (6)$$

where K stands now for the neighborhood value in the test phase.

Finally, a decision is made based on these coefficients:

$$o_k(\mathbf{x}) = \max_j \{ o_j(\mathbf{x}) \} \Rightarrow \mathbf{x} \in C_k \quad (7)$$

The membership coefficients for each test vector can be seen as confidence measures for the classification results. They are saved and then used just as the multilayer perceptron outputs by the different fusion methods, which are described in section 3.

2.3 SART classifier

The SART (Supervised ART) classifier [9] uses the principle of prototype generation like the ART neural network, but unlike this one, the prototypes are generated in a supervised manner. It has the capability to learn fast using local approximations of the class pdf and its operation does not depend on any chosen parameter.

The basic idea is to create a new prototype for a class whenever the actual set of prototypes is not capable anymore to classify the training data set satisfactorily. The prototypes are updated using the mean of the samples which are correctly classified by each of them. The updating process is repeated as long as there are classification errors on the training samples and as long as it dynamically changes the location of the prototypes.

Speaking in terms of probability distribution, the prototypes become the centroids of the modes of the multimodal class probability distribution.

The classifier structure is similar to that of an RBF or LVQ neural network. The algorithm described above is then used to train the hidden layer, while a MADALINE network implements the output layer (Fig. 3).

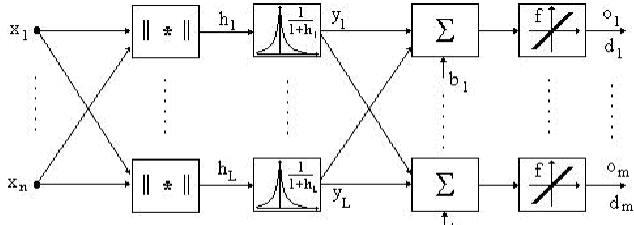


Fig. 3. SART classifier structure.

An important property of the algorithm is that it needs no initial system parameter specifications and no prespecified number of codebook vectors.

3 Fusion models

Three fusion models have been used for the information fusion of the three previous radar target classifiers.

3.1 Sugeno's fuzzy integral

Sugeno's fuzzy integral is a non-linear functional, similar to a Lebesgue's integral, defined with respect to a fuzzy measure [10].

Let us consider a set Q and $h: Q \rightarrow [0, 1]$ a fuzzy subset of this one. Then, the fuzzy integral of the function h on Q , with respect to the fuzzy measure g , is expressed by the relationship:

$$\begin{aligned} \int_Q h(q) \circ g(\cdot) &= \max_{A \subseteq Q} \left[\min \left[\min_{q \in A} h(q), g(A) \right] \right] \\ &= \max_{\alpha \in [0,1]} \left[\min(\alpha, g(h_\alpha)) \right] \end{aligned} \quad (8)$$

where $h_\alpha = \{q \mid h(q) > \alpha\}$.

$h(q)$ quantifies the decision taken by the classifier q concerning the membership of the unknown target to some class. In other words, this value measures the degree with which the concept h is satisfied by q [11].

The term $\min_{q \in A} h(q)$ measures then the degree with which the concept h is satisfied by all the elements of the subset A . $g(A)$ represent the importance of the group of classifiers constituting the set A for the final decision or, in an equivalent way, the degree with which they satisfy the concept g .

Consequently, the value obtained by the comparison of the two quantities through the operator "min" will indicate the degree with which the classifier set A satisfies the two criteria. One can then conclude that the fuzzy integral looks for the maximum degree of agreement between the real possibilities and expectations, measured by the functions h and g respectively [11].

In order to calculate the fuzzy integral, the values $h(q_i)$ are supposed sorted in descending order: $h(q_1) \geq h(q_2) \geq \dots \geq h(q_N)$; if they are not naturally, one can always change the order of q_i , so that this condition is met. Hence, the fuzzy integral can be calculated with the following equation:

$$\chi = \int_Q h(q) \circ g(\cdot) = \max_{i=1, N} \left[\min(h(q_i), g(A_i)) \right] \quad (9)$$

where: $A_i = \{q_1, q_2, \dots, q_i\}$.

$g(A_i)$ can be calculated in a recursive way using the following property of any fuzzy measure g_λ :

$$\begin{cases} g(A_1) = g(\{q_1\}) = g^1 \\ g(A_i) = g^i + g(A_{i-1}) + \lambda \cdot g^i \cdot g(A_{i-1}), i = \overline{2, N} \end{cases} \quad (10)$$

The data fusion of the N outputs of the N classifiers is then carried out by the algorithm given below.

A. Training stage

1) Measure the performances of the classifiers and obtain their fuzzy density function for each class.

$g_j^i = g_j(\{q_i\})$ is considered to be the recognition rate of the classifier q_i for the class C_j .

2) Calculate, for each class C_j , the corresponding value λ_j using the equation:

$$g(Q) = 1 \Rightarrow \lambda + 1 = \prod_{i=1}^N (1 + \lambda g_j^i) \quad (11)$$

B. Classification stage

1) Calculate the outputs $o_j^i = h_j(q_i)_{j=1, m, i=1, N}$ of the N classifiers corresponding to the m classes.

2) Calculate the fuzzy integral for each class with $g_j(A_i)$ determined with the equation (8):

$$\chi_j = \max_{i=1, N} \left[\min(o_j^i, g_j(A_i)) \right] \quad (12)$$

3) Decide on the membership of the target according to the rule:

- a) if $\max_{j=1, m} \chi_j \geq \chi_0$ and $k = \arg \left[\max_{j=1, m} \chi_j \right] \Rightarrow \mathbf{x} \in C_k$
- b) if $\max_{j=1, m} \chi_j < \chi_0 \Rightarrow \mathbf{x}$ - unknown target

where χ_0 indicates a confidence threshold below which the target is declared unknown. It can be considered as the lowest value of the fuzzy integral χ_k obtained during the training process. The block diagram describing the fusion procedure is represented on figure 4.

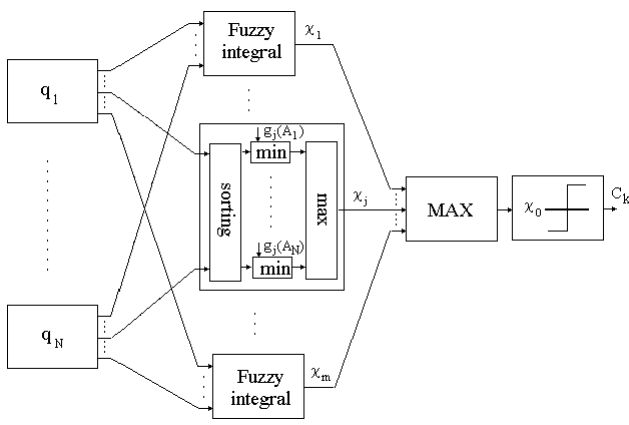


Fig. 4. Fuzzy integral based decision fusion

3.2 Possibility theory

The possibility theory has been initially developed by L.A. Zadeh [12], D. Dubois and H. Prade [13]. It has the capability to handle both the imprecision and the uncertainty by means of a distribution of possibility and two functions used for the event characterization: the possibility and the necessity.

The central concept for the possibility theory is the possibility distribution, which is defined over the definition domain $D=\{C_1, \dots, C_m\}$ of the given variable x as:

$$\pi : D \rightarrow [0,1], \quad \sup_{x \in D} \pi(x) = 1 \quad (13)$$

It describes the degree of membership of each value of x to the domain D , so that it is basically a fuzzy operator. Nevertheless, it is considered here in a very specific framework, namely to handle the imprecision associated to the exact value of the data.

A possibility measure can be then obtained from the distribution above with the following relationships:

$$\forall A \in 2^D, \quad \Pi(A) = \sup_{x \in A} \pi(x) \quad (14)$$

In order to be able to represent both the imprecision and the uncertainty a necessity measure is also introduced. It is related to the possibility measure by the equation below:

$$\forall A \subseteq D, \quad N(A) = 1 - \Pi(A^c) \quad (15)$$

In the framework of the possibility theory the information fusion is carried out using several combination operator families: t-normes, t-conormes, averages, symmetric sums etc. They use as inputs the values provided by the possibility distribution which is supposed to be known in this processing stage.

The choice of the combination operator is the central problem for the possibility theory because it depends on the considered application and the objectives to be reached. The main criteria which can support the choice procedure are the operator general behavior (conjunctive, disjunctive or trade-off), desired properties, behavior in

conflicting situations, capability to discriminate different cases etc. Several combination operator classes can be used in different applications, e.g. max (t-norme operator), min (t-conorme operator), mean and median (average operators).

The final decision is made using the following rule:

$$x \in C_k \text{ if } \mu_k(x) = \max_{i=1,N} \mu_i(x) \quad (16)$$

where $\mu_i(x)$ is the membership coefficient of x to the class C_i .

3.3 Dempster-Shafer theory

As the possibility theory, the DS theory or evidence theory allows for a representation of both imprecision and uncertainty through two functions: plausibility and belief [14, 15]. Both functions are derived from a mass function defined on each subset of the space of discernment $D=\{C_1, \dots, C_m\}$ (i.e. 2^D) onto $[0,1]$, such that:

$$\sum_{A \subseteq D} m(A) = 1 \quad (17)$$

The first difficulty is the choice of a mass function. There are two types of approaches: one based on distance transformation [16] and another one based on a probabilistic model [15]. Appriou [15] proposes two equivalent models based on 3 axioms. The first one that we will use in this article is given by:

$$\begin{cases} m_j^i(\{C_i\})(x) = \alpha_{ij} R_j p(q_j / C_i) / (1 + R_j p(q_j / C_i)) \\ m_j^i(\{C_i\}^c)(x) = \alpha_{ij} / (1 + R_j p(q_j / C_i)) \\ m_j^i(D)(x) = 1 - \alpha_{ij} \end{cases} \quad (18)$$

where q_i is the i^{th} classifier (supposed cognitively independent), $i=1, N$, α_{ij} are reliability coefficients on each classifier i for each class $j = \overline{1, m}$ (in our application $\alpha_{ij}=1$), and $R_j = (\max_{q_j, i} (p(q_j / C_i)))^{-1}$. Hence a mass

function is defined for each source and each class. The second step of fusion: the combination is based on the orthogonal Dempster-Shafer's rule:

$$m(\cdot) = \oplus_{i=1, N} (\oplus_{j=1, m} m_j^i(\cdot)) \quad (19)$$

Other conjunctive rules are proposed: the Dempster-Shafer's rule normalized by a conflict measure given by:

$$K = \sum_{B_1 \cap B_2 \cap \dots \cap B_N = \emptyset} \prod_{i=1}^N m_i(B_i) < 1 \quad (20)$$

The Yager's rule [17] redefines $m(D)$ adding K . In this article, we chose the Smets's rule [18] that suppose an "open world", for each A of 2^D :

$$m(A) = \sum_{B_1 \cap B_2 \cap \dots \cap B_N = A \neq \emptyset} \prod_{i=1}^N m_i(B_i) \quad (21)$$

$$m(\emptyset) = K$$

These rules give similar results on our data. The last step of fusion is the decision. In the DS theory, we can use the maximum of plausibility, maximum of belief or maximum of pignistic probability [18]. We use the maximum of plausibility in this article; the three previous criteria give the same results on our data.

4 Database and simulation results

The fusion models previously described have been used to classify 10 scale reduced (1:48) targets (Mirage, F14, Rafale, Tornado, Harrier, Apache, DC3, F16, Jaguar and F117). The real data were obtained in the anechoic chamber of ENSIETA (Brest, France) using the experimental setup shown on Fig 5.

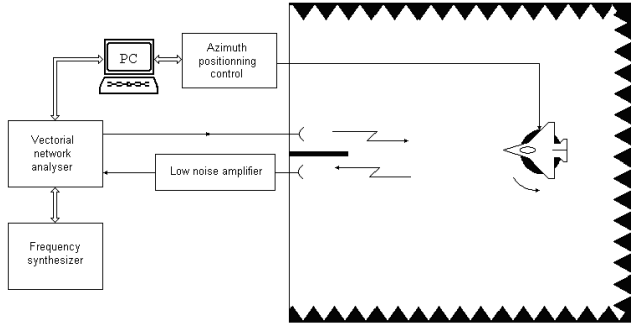


Fig. 5. Experimental setup

Each target is illuminated in the acquisition phase with a frequency stepped signal. The data snapshot contains 32 frequency steps, uniformly distributed over the band $B=[11650,17850]$ MHz, which results in a frequency increment of $\Delta f = 200$ MHz. Consequently, the slant range resolution and ambiguity window are given by:

$$\Delta R_s = c/(2B) \cong 2.4m, W_s = c/(2\Delta f) = 0.75m \quad (22)$$

The complex signature obtained from a backscattered snapshot is coherently integrated via FFT in order to achieve the slant range profile corresponding to a given aspect of a given target. For each of the 10 targets 150 range profiles are thus generated corresponding to 150 angular positions, from -5^0 to 69.5^0 , with an angular increment of 0.5^0 .

The procedure of evaluation is based on a Knowledge Discovery in Database (KDD) process (see Fig 6) developed in [19].

The first step is the partition of the database in a training set (for the three supervised classifiers) and test set (for the evaluation). When all the range profiles are available, the training set is formed by randomly selecting 2/3 of them, the others being considered as the test set.

The four combination operators of the possibility theory presented in section 3 have been tested. We present only the best operator on our data: the mean operator. Unless

the Sugeno's integral and the possibility theory, the DS theory is here applied on the decision of the classifiers and not on the numeric output (o_k) of the classifiers.

The fusion of the three fusion models is based on the Dempster-Shafer theory related to as FF.

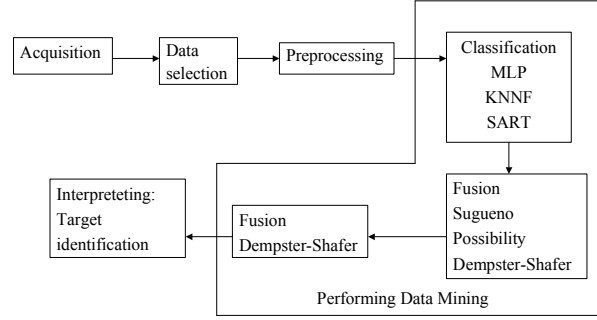


Fig. 6. KDD process for fusion of classifiers.

We repeat the training 15th times in order to achieve a good estimator of the classification rate, and take the mean of the classification rate (Table 1).

Table 1: Classification rate (%).

MLP	KNNF	SART	Sug	Pos	DS	FF
75.76	87.32	88.45	89.48	90.55	93.76	94.12

Note that the three fusion models give better results in mean than every individual classifier. The improvement of the DS theory is statistically significant. The final fusion achieves the best classification rate of 94.12%. The improvement is not the same for all the 15th tests, but for each test the DS theory gives the best classification rate compare to the two other fusion models (see Fig. 7).

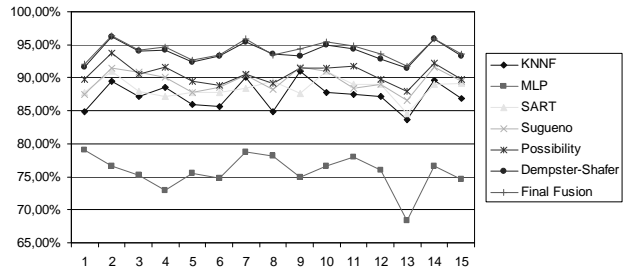


Fig. 7. Classification rate means calculated for each test.

The classification rate means are calculated for each target. We can see that the FF does not give the best classification rate for each test.

The problematic targets are not the same for each classifier (see Fig 8: on this figure the classification rate means are calculated on the 15th tests). For example the F14 is better recognized by the MLP classifier than the KNNF classifier, and it is the contrary for the Rafale. Hence the fusion of the models can exploit well the differences to obtain better results.

We can also note that the problematic targets are not the same for each fusion approach. Hence we can see that the final fusion achieves the best results. However the

improvement is not significant compared to the DS model. We have also tested the fusion of the six classifiers (MLP, KNNF, SART, Sugeno classifier, possibility classifier, and DS classifier). The results show a very small improvement for a high computational cost.

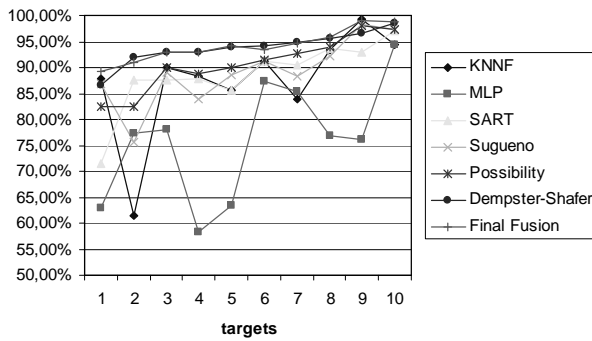


Fig. 8. Classification rate mean calculated for each target.

5 Conclusions and future works

We have used three classifiers based on three different principles: a Multilayer perceptron, a fuzzy K Nearest Neighbor, and a Supervised ART classifier. These three classifiers allow good performances. However, the performances are not similar according to the target. So, we have used three fusion models: Sugeno's integral, possibility, and Dempster-Shafer theory to fusion the outputs or decisions of the classifiers. These three approaches outperform the results of the classifiers. The Dempster-Shafer approach achieves the best classification rate. The fusion of the three fusion models gives the best classification rate. However the improvement is low for a high computational cost.

As future work, we would like consider a fusion model as an unique classifier. This means a very good radar data modeling, especially the imperfections of the information. The difficulty is this modeling, especially for the uncertainty and imprecision numeric data and the modeling of symbolic data.

References

- [1] R. Wehner, *High Resolution Radar*, Artech House, Boston, 1987.
- [2] V.G. Nebabin, *Methods and Techniques of Radar Recognition*, Artech House, London, 1994.
- [3] H.J. Li, S.H. Yang, Using Range Profiles as Feature Vectors to Identify Aerospace Objects, *IEEE Trans. on Antennas and Propagation*, 41(3): 261-268, 1993.
- [4] H.J. Li, R.Y. Lane, Utilisation of Multiple Polarisation Data for Aerospace Target Identification, *IEEE Trans. on Antennas and Propagation*, 43(12): 1436-1440, 1995.
- [5] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, *Parallel Distributed Processing*, vol. 1, 318-362, MIT Press, Cambridge, 1986.
- [6] B. Kosko, *Neural Networks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, 1992.

- [7] R.O. Duda, P.O. Hart, D.G. Stork, *Pattern Classification*, Second Edition, John Wiley & Sons, New York, 2001.
- [8] J.M. Keller, D.J. Kunt, Incorporating fuzzy membership function, *IEEE Trans. on Systems, Man and Cybernetics*, 15: 580-585, 1985.
- [9] E. Radoi, *Contribution à la reconnaissance des objets 3D à partir de leur signature électromagnétique*, Ph.D. Thesis, Brest, 1999.
- [10] M. Sugeno, Fuzzy Measures and Fuzzy Integrals – A Survey, *Fuzzy Automata and Decision Processes*, Amsterdam, North Holland, pages 89-102, 1977.
- [11] S.B. Cho, J.H. Kim, Combining Multiple Neural Networks by Fuzzy Integral for Robust Classification, *IEEE Trans. on Systems, Man and Cybernetics*, 25(2): 380-384, 1995.
- [12] L.A. Zadeh, *Fuzzy Sets as a Basis For a Theory of Possibility*, *Fuzzy Sets and Systems*, 1: pages 3-28, 1978.
- [13] D. Dubois, H. Prade, *Possibility Theory*, Plenum Press, New York, 1988.
- [14] I. Bloch, Some Aspects of Dempster-Shafer Evidence Theory for Classification of multi-modality Medical Images Taking Partial Volume Effect into Account, *Pattern Recognition Letters*, 17: 905-919, 1996.
- [15] A. Appriou, Situation Assessment Based on Spatially Ambiguous Multisensor Measurements, *International Journal of Intelligent Systems*, 16(10): 1135-1166, 2001.
- [16] T. Denoeux, A k -nearest neighbor classification rule based on Dempster-Shafer Theory, *IEEE Trans. on Systems, Man and Cybernetics*, 25(5): 804-813, 1995.
- [17] R.R. Yager, On the Dempster-Shafer Framework and New Combination Rules, *Information Science*, 41: 93-137, 1987.
- [18] P. Smets, The Combination of the Evidence in the Transferable Belief Model, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(5): 447-458, 1990.
- [19] B. Hoeltzner, F. Pellen, A. Khenchaf, Analysis of Complexity, Different Levels of Exploitation: Application to Radar Automatic Target Recognition, Vol 2, pages 261-266, *Proceedings of the sixteenth International Conference on Systems Engineering (ICSE2003)*, Coventry, UK, 9-11 September 2003.