Available online at www.sciencedirect.com

SCIENCE DIRECT®

Information Fusion 7 (2006) 197–206

INFORMATION FUSION

www.elsevier.com/locate/inffus

Comparative implementation of two fusion schemes for multiple complementary FLIR imagery classifiers

Pierre Valin [a,c], Francois Rhéaume [b,c], Claude Tremblay [a], Dominic Grenier [b,*], Anne-Laure Jousselme [b,c], Éloi Bossé [b,c]

[a] Lockheed Martin Canada, 6111 Royalmount avenue, Montréal, QC, Canada H4P 1K6

[b] Dept. de Génie Electrique et Génie Informatique, Université Laval, Pavilion Pouliot, Quebec Que., Canada G1K 7P4

[c] Defence R&D Canada Valcartier, 2459 Pie-XI Blvd North, Val-Bélair, Que., Canada G3J 1X5

Received 11 March 2004; received in revised form 24 September 2004; accepted 30 September 2004
Available online 11 November 2004

Abstract

Several classifiers for forward looking infra-red imagery are designed and implemented, and their relative performance is benchmarked on 2545 images belonging to 8 different ship classes, from which 11 attributes are extracted. These are a Bayes classifier, a Dempster–Shafer classifier ensemble in which specialized classifiers are optimized to return a single ship class, a *k*-nearest neighbor classifier, and an optimized neural net classifier. Two different methods are then studied to fuse the results of selected subsets of these classifiers. The first method consists of using the outputs of various classifiers as inputs to a second neural net fuser. The second method consists of converting the outputs of these classifiers into masses for use in a Dempster–Shafer fuser. In both approaches, the fused classifier achieves better results than the best classifier for any given class.
Crown Copyright © 2004 Published by Elsevier B.V. All rights reserved.

Keywords: FLIR imagery; Fusion of classifiers; Bayes; Dempster–Shafer; *k*-Nearest neighbor classifier; Neural net; Meta-fuser

1. Introduction

In the last decades many classification methods and fusers have been developed. Considerable gains have been achieved in the classification performance by fusing and combining different classifiers. In this paper we consider two different methods for fusing classifiers for ship recognition using infra-red imagery.

Multiple classifiers, if combined appropriately, can provide a solution to get higher performance in terms of recognition rate and reliability. In general, the advantage of using them is that they make decisions by considering information coming from classifiers that do not behave in the same way and that can complement each other. Some classifiers may perform well in cases where others perform poorly, and then it is obvious that there are more chances to find the correct answer among several classifiers than only one. Now this raises the issue of evaluating the quality of the answers returned by each classifier in order to make the most accurate decision.

Fusion is a promising technique [1] to efficiently utilize the complementarity of separate classifiers. We experiment a new method [2] based on a fusion of a set of classifiers. Recently, Rao has demonstrated that individual results can be fused in order to obtain a more reliable decision and he has also demonstrated that the performance of a fuser can be guaranteed to perform at least as good as the best classifier under certain conditions [3].

The objective of a good fuser is to perform at least as good as the best classifier in any situation. To this end,

[*] Corresponding author. Tel.: +1 418 6562806; fax: +1 418 6563159.
E-mail address: dgrenier@gel.ulaval.ca (D. Grenier).

1566-2535/$ - see front matter Crown Copyright © 2004 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2004.09.001

we consider two different fusers (neural net and Dempster–Shafer) and four different classifiers (Bayes, *k*-nearest neighbors, neural net and Dempster–Shafer).

In the first fusion method, we consider three classifiers: DS classifier, Bayes classifier and *k*-nearest neighbors classifier and a feed-forward neural net fuser. We compare the results of the best classifier with the results of the fusion of a combination of classifiers. Our work is closely related to Rao's work [2], but in a more practical way, since we compare the results of the classifiers with the results of the fusion of two or three classifiers using a neural net fuser. The neural net fuser, under the assumption of a very large sample data set of images, obeys Rao's conditions for guaranteed improved performance. In practice however, we were far from having enough images, so we could only hope to demonstrate an improvement by experimentation. The fuser is indeed found to give a performance equal or superior to the best classifier in all cases.

To optimize the results for each ship class, we specialized classifiers using the Dempster–Shafer (DS) method for each class, i.e. a specialized classifier returns whether the ship belongs to the corresponding class or not. We compare the result of the generic DS classifier with the results of those specialized DS classifiers. The improvement in recognition varies between 3% and 20% for a class.

In the second fusion method, a different set of classifiers (*k*-nearest neighbors, Bayes and neural net) are combined using DS theory by appropriately defining the weights that best represent individual classifier evidences. This is called the measure-based method because it relies on the internal information of each rather than statistics.

2. Common data set and feature selection

Park and Sklansky [4] developed an automated design of linear tree classifiers for ship recognition. The data set that they studied was composed of 2545 forward looking infra-red (FLIR) ship images, belonging to eight possible output classes, but which are not equally represented in the sample. The same data set is used throughout this work.

The whole data set will be separated randomly into training and test/validation subsets for each classifier. This will remove any trends and/or correlations in the data, which are commonly present when acquiring FLIR image sequences. The quoted performance of each classifier will be the average performance over Monte-Carlo runs. In addition the relative size of the training and test/validation data sets will be varied to estimate the relative importance of this ratio.

2.1. Eight desired output classes

Let us recall that a classifier is a method for assigning a class to an object according to its attribute values. Each extracted attribute (or feature) will be considered as independent, and be classified through different methods. The desired output classes are shown below, together with the actual number of ship images for each class:

1. Destroyer (D) with 340 images
2. Container (CO) with 455 images
3. Civilian Freighter (CF) with 186 images
4. Auxiliary Oil Replenishment (AOR) with 490 images
5. Landing Assault Tanker (LAT) with 348 images
6. Frigate (F) with 279 images
7. Cruiser (CR) with 239 images
8. Destroyer with Guided Missile (DGM) with 208 images

Typical silhouettes for the best imagery from the eight classes are shown in Fig. 1 below, with the eight classes being from left to right and top (1–4) to bottom (5–8).

2.2. Eleven attributes extracted

Features are an abstraction of the raw data intended to represent the original information. Feature extraction attempts to find characteristics of the data that aid in the identification process. In passive image recognition, it is a difficult task to find features, which are discriminating for every class or give an accurate identification for all classes. For example, the FLIR images are taken from different acquisition angles (close to, but not always

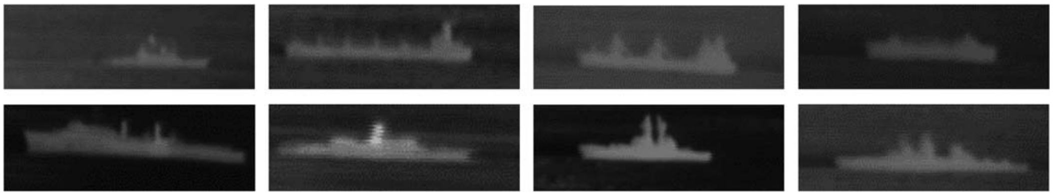

Fig. 1. Typical imagery for the eight classes (from left to right, top row of D, CO, CF, AOR, bottom row of LAT, F, CR, DGM).

broadside), from different zoom factors, etc. Therefore the chosen features have to be invariant under three transformations: translation, rotation and scaling.

We used the same features that Park and Sklansky [4] used, namely 11 attributes consisting of 7 invariant moments under scaling (to account for different zooms), rotation (because of different ship headings) and translation (since the ship image is not necessarily centered) that account for general features of the ships, and 4 auto-regressive parameters that provide more detailed target information.

The 7 invariant moments m_i were originally given by Hu [5], and are built from the second and the third order moments given by

$$\mu_{nm} = \sum_{(x,y)\in S} (x-\bar{x})^n (y-\bar{y})^m,$$

where the order of the moment is $(n + m)$, and $x(y)$ denote the horizontal (vertical) coordinates, in the silhouette S, and where $\bar{x}$ and $\bar{y}$ are the coordinates of the centroid of S.

The weakness of choosing invariant features is that these contain only information on the global shape of a given ship and represent poorly the details of the object. To overcome this disadvantage, a second set of features was extracted by fitting an auto-regressive (AR) model to the one-dimensional sequence of the projected image along the horizontal axis.

The projection of a ship image onto the horizontal axis usually preserves its shape, provided that the major axis of the ship is (nearly) parallel to the horizontal axis. If $r(i)$, $i = 1,2,\ldots,N$ denote the sequence of the projection sampled at N equally spaced points along the horizontal axis, an AR model can be constructed that expresses $r(i)$ as a linear combination of the previous projections $r(i-j)$, $j = 1,\ldots,m$, plus a bias α, and the error $\varepsilon(i)$ associated with the model, according to the equation

$$r(i) = \sum_{j=1}^{m} \theta_j r(i-j) + \alpha + \sqrt{\beta}\varepsilon(i).$$

The parameters are estimated by a least squares fit of the model to the one-dimensional sequence $r(1),\ldots,r(N)$. The least square estimates approximate the maximum likelihood estimates. Let θ, α and β now denote the least squares estimates of θ, α and β respectively. Thus the complete feature vector of an image consists of 7 invariant moments m_i and 4 AR parameters (for $m = 3$):

$$f_i = m_i, \quad i = 1,\ldots,7,$$

$$f_{i+7} = \theta_i, \quad i = 1,2,3,$$

$$f_{11} = \frac{\alpha}{\sqrt{\beta}}.$$

All these parameters have also been shown to be invariant to rotation, translation and scaling, so that these can indeed be used as a feature vector for the purpose of classification [4].

2.3. Frequency distribution of classes according to attributes

Each attribute will discriminate different classes to varying degrees. The classifiers are expected to have a performance that depends slightly on the discreteness of the binning scheme used for the attributes. Very precise values of each attribute are neither desired, nor easily measured. The width of the bins is a compromise between the expected extraction accuracy of the given attribute from the imagery, and having a representative number of classes in each bin of the attribute. Given that a certain image provides values (within a bin) for each attribute, each classifier will use that value in a different manner: Bayesian probability, Dempster–Shafer basic probability assignment (BPA or mass), etc.

The most convenient way of showing the discriminatory power of an attribute for every class is by means of one frequency graph per attribute. Frequency graphs have thus to be made for each of the 11 attributes [6]. Such a frequency graph is shown in Fig. 2 for attribute 1 (first entry of the feature vector m_1) for the entire data set.

In Fig. 2, the vertical axis represents the number of times that images of each type were found to the have the attribute values shown on the horizontal axis. The relative lengths of each bar corresponding to each class (or type) indicates how relatively often the attribute value is obtained for imagery of that class. For example a value between 1600 and 1800 strongly favors class 4 while being very rare for class 3, while values below 600 can only be representative of classes 1, 4 and 6.

When a classifier uses all the attributes for all classes, this will be referred to as a "generic" classifier throughout the rest of the paper. When only a selected set of attributes are used for a given class, this will define a "specialized" classifier for that class.

Several comments are in order at this time:

- The class (type) varies a lot from bin to bin, making fitting smooth curves difficult. Therefore this will not be attempted.
- It can happen, due to the limited statistics in the data, that a given class is not represented for an attribute bin sandwiched between bins that are populated by that class. Such a zero probability would seem accidental, thus a Bayesian classifier implementation will have to account for this fact.
- For certain classes, it may occur that a given attribute is not discriminatory (that class can have relatively uniform distribution across all bins of the attribute),

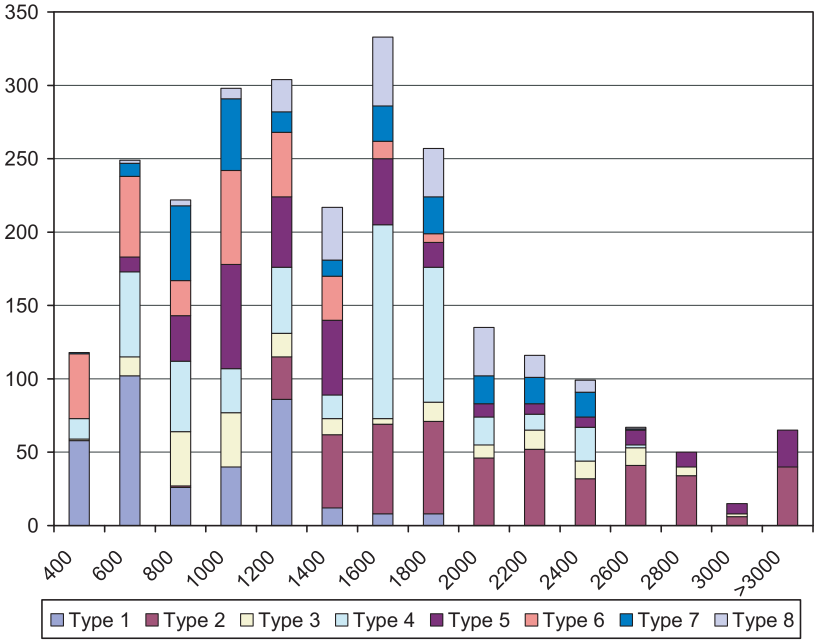

Fig. 2. Frequency graph for attribute 1 binned in increments of 200.

hence it may be beneficial to leave that attribute out in certain classifiers. Leaving certain attributes out will thus generate more efficient specialized classifiers for a given class.

- In all classifier examples below, the data sample is distributed randomly into training and test sets (varying between 1000 and 1500 examples), and the two extreme choices of 1000 and 1500 respectively will be shown to provide consistent results.
- The original sample distribution is not uniform across all classes, with class 3 least represented and class 4 the most. It should therefore be expected that classifiers which can exploit this a priori information (e.g. Bayes) perform better. Of course, if the distribution changes in future field exercises, such Bayesian classifiers would introduce a bias.

3. Classifiers

3.1. Generic DS classifier

The DS theory of evidence [7,8] is a good means of reasoning under uncertainty. A key aspect of this theory is its ability to combine evidences by using the technique of orthogonal summation.

The DS theory requires that the hypotheses pertain to the set of all possible propositions that can be output. This set is called the frame of discernment denoted by θ whose elements are mutually exclusive. The power set of θ is $P(\theta)$, which is the set of all the 2^θ subsets of θ. Let A be an element of $P(\theta)$.

A basic probability assignment (BPA) is a function m from $P(\theta)$ to $[0,1]$ that must satisfy the following constraints:

$$m(\Phi) = 0,$$

$$\sum_{A \in P(\theta)} m(A) = 1,$$

where Φ denotes the empty set, and the first equation shows that we are dealing with a closed universe. The BPA is often called mass function and can be used to estimate the probability distribution of $P(\theta)$. The precise probability distribution of $P(\theta)$ may not be known exactly so bounds of probability distributions are defined. The lower probability and upper probability of a subset A of $P(\theta)$ is denoted as belief measure $\mathrm{Bel}(A)$ and plausibility measure $\mathrm{Pls}(A)$, respectively. They can be determined from the mass function as follows:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A \subset P(\theta)} m(B),$$

$$\mathrm{Pls}(A) = \sum_{A \cap B \neq \Phi} m(B).$$

Generally, $\mathrm{Bel}(A) \neq \mathrm{Pls}(A)$, and the true Bayesian probability of A is between $\mathrm{Bel}(A)$ and $\mathrm{Pls}(A)$.

The combined mass functions of two independent mass functions m_1 and m_2 is calculated by using Dempster's rule of combination, denoted by $m_1 \oplus m_2$:

Table 1
Confusion matrix for the DS classifier with AIR = 74.5%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.871	0.006	0.000	0.009	0.006	0.079	0.026	0.003
Class 2	0.000	0.864	0.000	0.095	0.042	0.000	0.000	0.000
Class 3	0.038	0.081	0.070	0.640	0.172	0.000	0.000	0.000
Class 4	0.000	0.037	0.000	0.957	0.006	0.000	0.000	0.000
Class 5	0.000	0.182	0.000	0.190	0.629	0.000	0.000	0.000
Class 6	0.151	0.007	0.000	0.008	0.047	0.735	0.004	0.047
Class 7	0.326	0.113	0.000	0.008	0.050	0.000	0.490	0.013
Class 8	0.005	0.010	0.000	0.000	0.010	0.090	0.000	0.885

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K},$$

$$K = \sum_{B \cap C = \Phi} m_1(B) m_2(C),$$

where K is a normalization constant, called the conflict because it measures the degree of conflict between B and C.

We fused sequentially the 11 attributes with this method, by using for the mass the normalized frequency distributions (such as the one of Fig. 1 for attribute 1) for each attribute extracted from the given FLIR image. After the combination, we return the class with the highest mass. Table 1 shows the DS classifier confusion matrix, with average identification rate (AIR) of 74.5%. This AIR is simply the number of correct ship classifications, divided by the number of items to classify.

3.2. Results of specialized DS classifiers

To improve the results of every class of ship, we implemented specialized classifiers using the DS method for each class, i.e. a specialized classifier returns whether the ship belongs to the class or not [6]. For every specialized classifier, a subset of features was chosen which optimized the performance of the class. The choice of attributes was made by testing a large random sample of the 11! possibilities, and choosing the best one in terms of recognition rate. The recognition rate of the DS method increases, sometimes substantially, for each specialized DS classifier:

1. For class 1 (D), from 87.1% to 93.2% (only attributes 1, 3, 6, 9, 10, 11 are used)
2. For class 2 (CO), from 86.4% to 95.8% (only attributes 1, 6, 9 are used)
3. For class 3 (CF), from 7% to 24.2% (only attributes 4, 5, 11 are used)
4. For class 4 (AOR), from 95.7% to 98.4% (only attributes 1, 2, 7, 8, 9, 10, 11 are used)
5. For class 5 (LAT), from 62.9% to 74.7% (only attributes 2, 3, 7, 11 are used)
6. For class 6 (F), from 73.5% to 80.3% (only attributes 2, 4, 7, 11 are used)
7. For class 7 (CR), from 49% to 68.6% (only attributes 1, 3, 5, 11 are used)
8. For class 8 (DGM), from 88.5% to 92.8% (only attributes 2, 4, 5, 10 are used)

We see that each specialized DS classifier gives better results for all the classes than the generic DS classifier.

3.3. Modified additive Bayes classifier

Bayes classifiers use a probabilistic approach to assign a class. They compute the conditional probabilities of different classes given the values of the attributes and then predict the class with the highest conditional probability. The frequency graphs (such as shown in Fig. 1 for attribute 1) are used to assign the probabilities.

The equation below represents the probability of an object belonging to in the ith class (C_i) knowing the value of the jth attribute (A_j), where i represents the number of classes $i = \{1, 2, \ldots, m\}$ and j the number of attributes = $\{1, 2, \ldots, n\}$.

$$P(C_i \mid A_j) = \frac{P(A_j \mid C_i) P(C_i)}{\sum_{i=1}^{m} P(A_j \mid C_i) P(C_i)}.$$

We then compute the probability of an object to be in the class i, knowing the value of jth attribute (for each attribute), and sum them:

$$P(C_i) = \sum_{j=1}^{N} P(C_i \mid A_j).$$

Finally, we identify the class of the object X. We choose the class with the highest probability.

$$X = \mathrm{Arg}\{\max_{i \leqslant i \leqslant m} [P(C_i)]\}.$$

The confusion matrix for this modified additive Bayes classifier results in an AIR of 77.7% and is shown in Table 2.

It should be noted that the classical Bayes classifier would give the same results as the generic DS classifier in the absence of ignorance, since both use the same statistical inputs (frequency graphs) to determine either the probability or mass, and that both methods multiply these quantities, and renormalize at each fusion step.

Table 2
Confusion matrix for the modified Bayes classifier with AIR = 77.7%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.767	0.030	0.000	0.000	0.000	0.045	0.124	0.035
Class 2	0.000	0.784	0.026	0.101	0.086	0.000	0.004	0.000
Class 3	0.018	0.072	0.514	0.288	0.072	0.018	0.018	0.000
Class 4	0.003	0.017	0.030	0.922	0.017	0.000	0.000	0.010
Class 5	0.010	0.107	0.049	0.039	0.741	0.015	0.010	0.029
Class 6	0.164	0.030	0.000	0.000	0.000	0.691	0.018	0.097
Class 7	0.110	0.110	0.000	0.000	0.007	0.000	0.706	0.066
Class 8	0.009	0.009	0.000	0.000	0.000	0.051	0.000	0.932

3.4. *k*-Nearest neighbors classifier

The *k*-nearest neighbor (*k*-NN) classifier finds the *k* nearest neighbors based on a metric distance and returns the class with the greatest frequency (majority vote).

We used a distance weighted by the inverse of the inter-classes covariance matrix:

$$d_{\Gamma}^2(x_1, x_2) = (x_1 - x_2)^{\mathrm{T}} \Gamma^{-1} (x_1 - x_2).$$

The results of the *k* = 3 nearest neighbors classifier are shown in Table 3. Its high AIR of 94.8% is indicative that the 11-dimensional distribution of vectors is well separated for this small set of FLIR images.

The traditional criticism of the *k*-nearest neighbor rule, namely the large storage space required for the entire training set, and the necessity to search for the *k*-nearest neighbors in the entire training set in order to make a single object classification, is minimized here by choosing a small value of *k*. Nevertheless various values of *k* were tried, with *k* = 3 providing representative results while limiting computational complexity. This should be enough for the purpose of this paper, since the emphasis will be to show improved performance by the fusion of multiple classifiers, neither of which needs to be optimized.

3.5. Neural net classifier

We used the following parameters for the feed-forward backpropagation neural net classifier: two hidden layers, 50 neurons on the first layer, 30 neurons on the second layer, momentum = 0.5, maximal error = 0.001, epsilon = 0.1, number of maximal iterations = 100. The results are presented in Table 4 and show an AIR of 92.7%.

Table 3
Confusion matrix for the 3-NN classifier with AIR = 94.8%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.907	0.000	0.000	0.000	0.000	0.043	0.049	0.000
Class 2	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Class 3	0.000	0.100	0.800	0.060	0.030	0.000	0.000	0.010
Class 4	0.004	0.013	0.008	0.971	0.014	0.000	0.000	0.000
Class 5	0.000	0.006	0.006	0.011	0.949	0.023	0.006	0.000
Class 6	0.000	0.000	0.000	0.000	0.022	0.877	0.007	0.094
Class 7	0.062	0.035	0.000	0.000	0.009	0.062	0.832	0.000
Class 8	0.010	0.000	0.000	0.000	0.000	0.020	0.000	0.980

Table 4
Confusion matrix for the neural network classifier with AIR = 92.7%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.858	0.000	0.000	0.000	0.000	0.080	0.056	0.006
Class 2	0.000	0.987	0.000	0.013	0.000	0.000	0.000	0.000
Class 3	0.000	0.000	0.840	0.150	0.000	0.010	0.000	0.000
Class 4	0.000	0.000	0.008	0.992	0.000	0.000	0.000	0.000
Class 5	0.000	0.000	0.012	0.046	0.943	0.000	0.000	0.000
Class 6	0.029	0.000	0.000	0.000	0.029	0.906	0.007	0.029
Class 7	0.106	0.000	0.000	0.009	0.000	0.000	0.885	0.000
Class 8	0.010	0.000	0.000	0.000	0.000	0.030	0.020	0.940

4. Fusion of selected classifiers

To quote Roli [10], measurement-level fusion rules can be used when each classifier outputs class "confidence" levels for each input pattern. The soft outputs of the N individual classifiers can be considered as features of a new classification problem. In other words, classifiers can be regarded as feature extractors, such that another classifier can be used as fuser. For this reason, sets of classifiers will be selected, and another classifier will be used as a fuser.

Having at our disposal, four distinct classifiers, we will fuse them in two different ways:

1. Using a neural net fuser for Bayes, DS and k-NN classifiers, with $N = 2$ or 3
2. Using a DS fuser for Bayes, neural net and k-NN classifiers, with $N = 3$

By deliberately not using a neural net classifier with a neural net fuser, nor DS classifiers with a DS fuser, the hope is to preserve the best features of each approach, and take full advantage of the variety of techniques employed.

4.1. Results of the first fuser approach with a neural net

Let us recall typical classification results of each method:

1. for the modified Bayes, an AIR of 77.5%
2. for the generic DS, an AIR of 74.5%
3. for the three nearest neighbor, an AIR of 94.8%

Table 5 below shows that the AIR performance of the classifiers depends only slightly on the relative sizes of the training vs. test data sets, such that the above statements are meant as averages.

Table 5
Single classifier AIR results

Classification method	Tested on 1000 images	Tested on 1500 images
Modified Bayes	0.777	0.773
DS	0.745	0.746
3-Nearest neighbor	0.946	0.950

The results of any two or all three classification methods were then fused with a feed-forward neural network fuser [6], for which a detailed analysis can be found in [9]. Our neural network fuser has thus 16 or 24 inputs (these inputs are the results of selected subsets of two or three classifiers) and has 8 outputs, one for each class, with all other parameters fixed to the ones of the neural net classifier. For more details, consult the thesis of [11].

Thus 3 experiments were performed with the NN fuser, with the result that the fuser gives performance equal or superior to the best classifier in all cases (any combination of classifiers and any training set size). The following results are shown in Tables 6–8, only for the case of training on 1500 images, and testing on 1000 images, and only for one Monte-Carlo run.

1. First, the results of modified Bayes classifier and DS classifier were fused, resulting in the confusion matrix of Table 6, with a considerable improvement over the 77.7% of the best classifier of Table 5. This is a clear indication that fusing relatively poor classifiers can result in a big improvement.
2. Second, the results of the modified Bayes classifier and 3-NN classifier were fused, resulting in Table 7, showing a small improvement (95.5%) over the 94.6% of the best classifier in Table 5. This is probably an indication that fusing good classifiers results in only a moderate improvement.
3. Third, the results of modified Bayes, DS, and 3-NN classifiers were fused resulting in Table 8. In this case again, a small improvement (95.1%) is obtained over the 94.6% of the best classifier in Table 5. This is probably a confirmation that fusing good classifiers can only result in a moderate improvement.

We also interchanged the training/test data set sizes for the fuser to check the variability of the performance improvement. From Table 9, we can see that the fuser gives performance equal or superior than the best classifier, and provides the best improvements when the fused classifiers are not very efficient, such as would be

Table 6
Fusion of Modified Bayes and DS Confusion Matrix—AIR = 84.4%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.940	0.007	0.000	0.015	0.007	0.022	0.000	0.007
Class 2	0.000	0.910	0.017	0.034	0.039	0.000	0.000	0.000
Class 3	0.027	0.082	0.384	0.370	0.137	0.000	0.000	0.000
Class 4	0.000	0.032	0.079	0.863	0.026	0.000	0.000	0.000
Class 5	0.000	0.072	0.014	0.022	0.884	0.000	0.000	0.007
Class 6	0.101	0.000	0.000	0.000	0.037	0.844	0.000	0.018
Class 7	0.170	0.011	0.000	0.021	0.000	0.011	0.777	0.011
Class 8	0.000	0.012	0.000	0.000	0.012	0.060	0.000	0.917

Table 7
Fusion Modified Bayes and 3-NN Confusion Matrix—AIR = 95.5%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.970	0.000	0.000	0.000	0.000	0.007	0.015	0.007
Class 2	0.000	0.994	0.000	0.000	0.000	0.000	0.006	0.000
Class 3	0.000	0.055	0.890	0.055	0.000	0.000	0.000	0.000
Class 4	0.000	0.011	0.005	0.963	0.021	0.000	0.000	0.000
Class 5	0.000	0.000	0.000	0.036	0.949	0.007	0.007	0.000
Class 6	0.055	0.000	0.000	0.000	0.009	0.890	0.018	0.028
Class 7	0.021	0.000	0.000	0.000	0.000	0.011	0.968	0.000
Class 8	0.000	0.000	0.000	0.000	0.000	0.036	0.000	0.964

Table 8
Fusion of Modified Bayes, DS, and 3-NN Confusion Matrix—AIR = 95.1%

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.931	0.005	0.000	0.010	0.000	0.035	0.015	0.005
Class 2	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Class 3	0.000	0.009	0.928	0.045	0.009	0.000	0.000	0.009
Class 4	0.000	0.000	0.003	0.993	0.003	0.000	0.000	0.000
Class 5	0.000	0.000	0.010	0.029	0.951	0.010	0.000	0.000
Class 6	0.036	0.000	0.000	0.006	0.012	0.873	0.006	0.067
Class 7	0.066	0.015	0.000	0.000	0.007	0.007	0.904	0.000
Class 8	0.017	0.000	0.000	0.000	0.000	0.026	0.000	0.957

Table 9
Fusion results of classifiers with feed-forward neural networks

Training size	Testing size	Bayes and DS (%)	Bayes and *k*-NN (%)	Bayes, *k*-NN and DS (%)
1000	1500	81.4	95	95.1
Best single classifier		77.3	95	95
1500	1000	85.3	95.5	95.6
Best single classifier		77.7	94.6	94.6

expected if more complementarity was present across poor classifiers. Table 9 is an average over Monte-Carlo runs (for test size 1000), while Tables 6–8 show typical results for just one run. Table 9 shows that adding more classifiers results in better performance on the average. Comparing Table 9 results (for test size 1000) with the AIR of Tables 6–8 shows the extent of the variability in the results expected from fusion.

4.2. Results of the second fuser approach by a measure-based method

In this case, the DS classifier has been replaced by a neural network classifier using the input 11 attributes with the same number of hidden layers and 8 output classes as for the neural network fuser mentioned above, but one performs instead the fusion of classifiers using DS theory [11].

The most important feature lies in the choice for the masses of the propositions for the class from the outputs of each classifier.

For the Bayes classifier, the masses are identified to the a posteriori probabilities of occurrence of each class and hence each proposition is a singleton, and there are 8 such propositions.

For the neural net classifier, the masses are the outputs for each class and hence each proposition is also a singleton, and there are 8 such propositions.

For the *k*-NN classifier, the situation is more complicated, since two complex propositions are selected, and the assignment of normalized masses to these propositions is more complicated [12]. Thus, if d_1 denotes the distance to the nearest neighbor, all classes represented in the hypershell $[d_1, Cd_1]$ make up a proposition having mass m_1,

$$m_1 = \frac{1}{T}\frac{k_1}{\sum_{i=1}^{k_1} d_i},$$

where k_1 is the number of neighbors in the hypershell. The other proposition is made up of the classes of the *k*-nearest neighbor and has a similar expression for its mass m_2,

$$m_2 = \frac{1}{T}\frac{k}{\sum_{i=1}^{k} d_i},$$

with T a normalization constant ensuring that $m_1 + m_2 = 1$. An example of results for such a DS fuser is given by the confusion matrix for the 8 classes as shown in Table 10.

Naturally the results turn out to depend on the hypershell thickness C that contains the various classes close to the selected nearest neighbor hyperpoint, and to the value of k, which thus become parameters affecting somewhat the performance of the fusion. The meas-

Table 10
Measure-based method confusion matrix

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Class 1	0.963	0.000	0.000	0.000	0.000	0.037	0.000	0.000
Class 2	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
Class 3	0.000	0.010	0.960	0.030	0.000	0.000	0.000	0.000
Class 4	0.000	0.004	0.000	0.996	0.000	0.000	0.000	0.000
Class 5	0.000	0.000	0.000	0.006	0.989	0.006	0.000	0.000
Class 6	0.015	0.000	0.000	0.000	0.000	0.942	0.000	0.043
Class 7	0.018	0.009	0.000	0.000	0.000	0.000	0.973	0.000
Class 8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

ure-based method can have an AIR as high as 98.1% when the k-NN classifier is properly selected ($k = 15$ and $C = 1.2$). For more details, see the thesis of [13].

4.3. Comparing other approaches on the same real FLIR data

The same FLIR data can be treated through distributed learning for classification. A system in which an agent network processes observational data and outputs beliefs (in the DS sense) to a fusion center module (the fuser) is considered [14]. The agents are modeled using evidential neural networks, whose weights reflect the state of learning of the agents. One agent processes the 7 moments, while another agent processes the AR parameters. Training of the network is guided by reinforcements received from the environment as decisions are made. Two different sequential decision making mechanisms were attempted: the first one is based on a "pignistic ratio test" and the second one is based on "the value of information criterion", providing for learning utilities (for more details, see [14]). The results for the class recognition rate (the AIR) that can be obtained oscillate between 54.1% and 58.7%, far lower than what can be achieved by fusing several classifiers. These low results may be due to the common evidential neural net classifier design for the two agents, or to the fusion center functionality, which receives beliefs from the agents, but makes decisions using pignistic probabilities. It could also be that distributed learning by only two agents may not be sufficient for this problem.

Different attributes can also be extracted from the same FLIR data [15]. Image segmentation can also be performed by a more complicated biologically-motivated algorithm, namely the visual perception segmentation process, rather than simple thresholding. Two classifiers can then be built and fused by two fusers.

1. For the first classifier, the segmented ship is partitioned into 7 equal sections along the x-axis and into two sections along the y-axis delimited by a centroid, and two sets of moments are calculated, one being structural, the other intensity based, resulting in 14 attributes. The structural moments are computed on the part of the segmented ship that is above the centroid, i.e. on the discriminating part which is above the hull. Again DS theory is used to combine the 14 expert opinions, and the output is a belief score for each ship's class. This is very similar to the "generic" DS classifier discussed previously, but over different attributes.
2. For the second classifier, a template-based method is used, attribute extraction in this case consists of computing shape descriptors, which will be used for template matching. Since the algorithm is quite complex, the reader is referred to [15]. A transformation must be made for the output to be compatible with the results of the DS classifier, namely "masses" summing up to 1.

Results [15] show that the overall accuracy of the template-based classifier is slightly lower (73.1%) than the moment-based DS one (75.5%). Note that the latter DS result using 14 different attributes calculated over seven ship sections is quite close to the 74.5% for the 11 attributes described previously. This can be interpreted to show that attribute determination (in number and by procedure) is not very important and that an AIR of 75% is typical of DS classifiers.

Two fusers are then tested [15], one using the product rule, resulting in an AIR of 80.8%, and one using DS, resulting in an AIR of 80.5%. Again, any fuser increases the AIR substantially, because both classifiers have rather poor individual AIRs, and are complementary. In this case the improvement was slightly over 6%, when compared to average classifier results (74.3%).

These results should be compared to the results from the neural net fuser when fusing its two worst fusers (Bayes and DS, for an average AIR of 76%). The result obtained in this case has a much better AIR of 84.4% (from Table 6), for an improvement of over 8%, rather than 6%. This could be interpreted to mean that fuser design is indeed important when fusing poor classifiers, or alternatively that the choice of the type of classifiers to be fused is important, since classifiers may exhibit varying degrees of complementarity.

5. Conclusions

The results indicate that specialized classifiers can be a good choice for identification of ship classes of FLIR images. In our particular case, the specialized DS classifiers consistently perform better. We also showed that a feed-forward neural network fusing Bayes, DS and *k*-NN classifiers results in better performance for this kind of identification. We have also used a DS fuser on Bayes, neural net and *k*-NN classifiers with a resulting improvement in performance of about 3% compared to the first fuser. This fact may be due to the extra parameters in this approach (*k* and *C*) which can be optimized. In all of our experiments, the performance of any fuser was always at least as good as the best classifier.

Acknowledgements

FLIR images are a courtesy of the United States Naval Air Warfare Center (NAWC), at China Lake (California), and attributes were provided by Dr. Jack Sklansky of the University of California at Irvine.

References

[1] J. Kittler, F. Roli, Multiple Classifiers Systems, vol. 1857, Springer-Verlag, Berlin, 2000.

[2] N.S.V. Rao, Multisensor fusion under unknown distributions: finite sample performance guarantees, in: A.K. Hyder, E. Shahbazian, E. Waltz (Eds.), Multisensor Fusion, Kluwer Academic Publishers, 2002.

[3] N.S.V. Rao, On design and performance of metafusers, in: Proceedings of the Workshop on Estimation, Tracking and Fusion: A Tribute to Yaakov Bar-Shalom, Monterey, CA, May 2001, pp. 259–268.

[4] Y. Park, J. Sklansky, Automated design of linear tree classifiers, Pattern Recognition 23 (12) (1990) 1393–1412.

[5] M.K. Hu, Visual pattern recognition by moment invariant, IEE Trans. Inform. Theory IT-8 (1962) 179–187.

[6] C. Tremblay, P. Valin, Experiments on individual classifiers and on fusion of a set of classifiers, in: FUSION 2002, Annapolis, MD, 7–11 July 2002, pp. 272–277.

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

[8] A. Dempster, Upper and lower probabilities induced by multivalued mapping, Ann. Math. Statist. 38 (1967) 325–339.

[9] N.S.V. Rao, Fusion Methods in multiple sensor systems using feedforward neural networks, Intell. Automat. Soft Comput. 5 (1) (1999) 21–30.

[10] F. Roli, A gentle introduction to fusion of multiple pattern classifiers, in: E. Shahbazian, G. Rogova, P. Valin (Eds.), Proceedings of the NATO ASI held in Armenia, 18–29 August 2003, Kluwer, in press.

[11] C. Tremblay, Classification d'images infrarouges de bateaux à l'aide du raisonnement évidentiel de Dempster–Shafer et étude de la fusion de plusieurs classifieurs, M.Sc. Thesis, Université de Montréal, Département de mathématiques et de statistique, September 2002.

[12] F. Rhéaume, A.-L Jousselme, D. Grenier, E. Bossé, P. Valin, New initial basic probability assignments for multiple classifiers, signal processing, sensor fusion, and target recognition XI, in: SPIE Aerosense 2002, Orlando, USA, 1–5 April 2002, vol. 4729, pp. 319–328.

[13] F. Rhéaume, Combinaison de classificateurs supervisés à l'aide de la théorie de l'évidence—Application à la reconnaissance d'objets issus d'images infrarouges, M.Sc. thesis, Université Laval, November 2002.

[14] G.L. Rogova, P. Scott, C. Lollett, Distributed fusion: learning in multi-agent systems for time critical decision making, in: E. Shahbazian, G. Rogova, P. Valin (Eds.), Proceedings of the NATO ASI, on Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management, held in Armenia, 18–29 August 2003, Kluwer, in press.

[15] H. Demers, Fusion of two imagery classifiers: a case study, in: E. Shahbazian, G. Rogova, P. Valin (Eds.), Proceedings of the NATO ASI, on Data Fusion for Situation Monitoring, Incident Detection, Alert and Response Management, held in Armenia, 18–29 August 2003, Kluwer, in press.