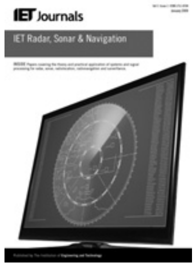
Published in IET Radar, Sonar and Navigation
Received on 6th November 2008
Revised on 22nd January 2009
doi: 10.1049/iet-rsn.2008.0182



ISSN 1751-8784

Classification of primary radar tracks using Gaussian mixture models

L.P. Espindle   M.J. Kochenderfer
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02420, USA
E-mail: mykelk@ll.mit.edu

Abstract: Classification of primary surveillance radar tracks as either aircraft or non-aircraft is critical to a number of emerging applications, including airspace situational awareness and collision avoidance. Substantial research has focused on target classification of pre-processed radar surveillance data. Unfortunately, many non-aircraft tracks still pass through the clutter-reduction processing built into the aviation surveillance radars used by the Federal Aviation Administration. This paper demonstrates an approach to radar track classification that uses only post-processed position reports and does not require features that are typically only available during the pre-processing stage. Gaussian mixture models learned from recorded data are shown to perform well without the use of features that have been traditionally used for target classification, such as radar cross-section measurements.

1 Introduction

Radar target classification based solely on primary surveillance radar (PSR) returns is known to be very difficult. In fact, the difficulty of distinguishing aircraft from clutter, such as birds, insects and weather, was a major incentive for the Federal Aviation Administration (FAA) to mandate the installation of transponders on many classes of aircraft to aid air traffic control [1]. Now, all commercial aircraft and almost 85% of general aviation aircraft are equipped with transponders [2]. Air traffic control uses secondary surveillance radar (SSR) to track transponder-equipped aircraft and provide separation services. Aircraft without transponders, often referred to as 'non-cooperative' aircraft, must rely on pilot visual acquisition to avoid collision and cannot expect air traffic control separation services because of the difficulty in accurately identifying aircraft tracks using only primary radar. A system that accurately filters out clutter tracks caused by birds and other non-aircraft would greatly enhance situational awareness and collision avoidance systems that require automated surveillance of non-cooperative aircraft.

A significant amount of literature has accumulated on the topic of radar target classification. Some research has focused on suppressing or filtering out PSR clutter returns, including birds, within the radar data processing unit itself. This effort has come from the air traffic and weather surveillance communities as well as the ornithological community [3]. Methods for identifying birds on various radar systems can be found in the radar literature [1, 4–11]. A separate, more recent, set of studies has focused on using techniques including support vector machines (SVMs) to classify the targets as aircraft or clutter [12, 13]. These studies, however, still used information such as radar cross-section that is not recorded by the FAA.

The contribution of this work is the development of a classifier that can discriminate radar tracks as aircraft from clutter using only target position reports. Such a classifier is necessary for automated situation awareness systems and collision avoidance. Building a classifier by hand would be very difficult, even for someone with operational experience with aviation surveillance radar. Instead, we identify a collection of properties of radar tracks that may be useful for target classification, e.g. airspeed, and apply automated supervised and semi-supervised learning techniques to infer the best classifier from a large collection of data.

Supervised learning involves learning from both positive and negative examples, in this case tracks known to be birds and tracks known to be aircraft. One may easily obtain a large collection of known bird tracks by collecting the tracks that penetrate Class B airspace, the airspace surrounding major airports where a transponder is required. Obtaining a large collection of primary tracks that are known to be aircraft is more difficult. However, we can use SSR tracks from aircraft with transponders as a surrogate for aircraft without transponders. We limit ourselves to aircraft whose beacon code is 1200, indicating that they are flying under visual flight rules (VFR), because they better resemble the kinds of aircraft that are not equipped with transponders. We also investigate semi-supervised learning methods, which involves learning from positive and unlabelled examples – in this case tracks known to be birds and tracks of unknown classification. Semi-supervised learning has not been previously used for track classification, but we chose to investigate this type of learning because we were concerned that the traditional supervised learning methods trained using beacon tracks would misclassify some types of unconventional aircraft that do not behave like traditional, transponder-equipped aircraft.

There are many different supervised and semi-supervised learning algorithms that have been well-studied in the pattern classification and machine learning literature. We compare several of these methods in terms of their classification accuracy, but we focus primarily on Bayesian classifiers that use Gaussian mixture models (GMMs) to represent the class-conditional distributions over target features [14]. Other probabilistic models could be used to represent the class-conditional distributions, but GMMs have several desirable properties. They have been well-studied in the literature and efficient algorithms have been developed for learning the parameters of the model from data. Because GMMs can have an arbitrary number of components, they can approximate multi-modal densities very well. They work with continuous, multi-dimensional feature spaces without requiring discretisation. Although GMMs have been used in a variety of classification tasks, they have not been used for radar target classification to the best of our knowledge. The purpose of this paper is not to argue that GMM-based Bayesian classification provides the best classification accuracy, although we did find that GMM-based classifiers performed better than the collection of Bayesian and non-Bayesian classification algorithms we tried. Instead, we wish to show how one may go about applying techniques from the machine learning community to the difficult task of target classification using only position reports.

Section 2 describes the data used to train our classifiers. Section 3 describes the features we extract from the data in order to perform classification. Section 4 describes the methods of classification, and Section 5 discusses results. Section 6 presents our conclusions and discusses future work. In this paper, we use standard measures for aviation. Distance is in nautical miles (NM), altitude is in feet (ft) and azimuth is in degrees from true north.

2 Data

The radar data used to train our classifiers streams from FAA and Department of Defense sensors distributed throughout the United States, including long-range ARSR-3 and ARSR-4 radars and short-range ASR-8, ASR-9 and ASR-11 radars. The data stream includes both primary and secondary reports as received directly from the radars, unaffected by any automation systems operating on the data. For this investigation, we used reports from eight ASR-9 sensors based at major airports with non-overlapping coverage during the month of June 2008 (see Fig. 1).

As Figs. 2*a* and *b* illustrate, the internal clutter rejection filters on the ASR-9 are highly sophisticated, with only a small percentage of primary reports actually forming tracks. Our classifiers are trained and evaluated on track position reports. Prior to feature extraction, the raw tracks undergo outlier removal, smoothing and interpolation to 1 Hz using a process described in prior work [15].

3 Features

Feature extraction involves converting an interpolated track into a vector of variables describing various aspects of the track relevant for classification. Features we considered included

- velocity: The median of the true airspeed. We assume most aircraft travel faster than most birds, but recognise there is overlap, especially towards the bottom of the aircraft performance spectrum which includes many non-cooperative aircraft.

- turnrate: The mean turn rate. Aircraft tend to turn more than birds in a single direction.

- stdev-turnrate: The standard deviation of turn rate. We observed that bird tracks tend to 'zig-zag' more than aircraft tracks, while still maintaining a relatively constant heading.

Figure 1 *Coverage of ASR-9 radar sensors used to construct the database of primary-only radar tracks*

doi: 10.1049/iet-rsn.2008.0182

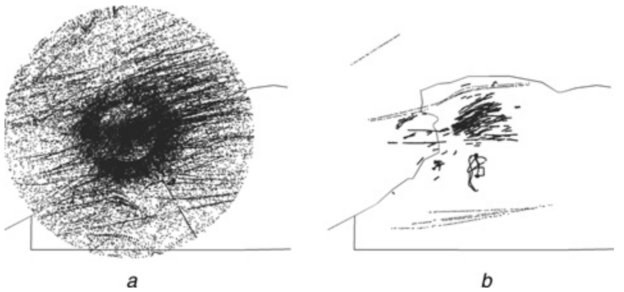


Figure 2 *Untracked and tracked PSR reports for the Buffalo, New York sensor from 9 June 2008*

a Untracked reports (3 h)
b Tracked reports (24 h)

- stdev-turnrate-adj: The standard deviation of turnrate divided by the average range of the track from the radar. We divide by range to correct for the fact that tracks far from the radar may experience fluctuations in calculated turn rate due to measurement error unrelated to target flight characteristics.

- turnrate-fluc: The average of the absolute value of the difference in turn rate from second to second. This is included as a feature for similar reasons as the standard deviation of turn rate.

- turnrate-fluc-adj: Turn rate fluctuation divided by range from the radar; the adjustment is for measurement error.

- duration: The duration of the track in seconds. It was hypothesised that aircraft tracks last longer than clutter.

- distance: The distance of the track measured by the maximum distance the track travelled from the starting point of the track. It was hypothesised that bird tracks may not be tracked for a long distance because their tracks are more easily dropped by the radar, or because they do not fly as far as aircraft.

- wind: This feature is calculated by first determining the average heading of all non-cooperative tracks in the sensor coverage, plus or minus 6 h from the average time of the track. The actual statistic is the percentage of time the track has a heading within 30° of the average heading of all other non-cooperative tracks for that time period. We observed that many presumed bird tracks around the same time and place tend to have a similar heading, presumably due to the wind.

- skipped: The fraction of times the radar failed to record a scan on the track during a sweep. Skipped scans can be due to low returned energy, typical of most bird tracks.

- timeofday: The average time of day for the track represented as a fraction of a total day. Time of day is local mean time. Migrating birds tend to fly at night, and recreational flying tends to be during daylight.

- stdev-heading: The square root of the circular variance of heading [16]. This value grows as tracks perform manoeuvres. Migrating birds tend to maintain their heading, so they have a low standard deviation of heading as compared to manoeuvreing aircraft.

Our feature set shares some similarities with features used in other projects. Dizaji and Ghadaki use various measures of track variation from a smoothed heading, as well as speed, skipped scans and distance travelled [13]. They also, however, include a handful of features using measurements such as returned amplitude from the target, which is generally unavailable after internal radar processing. Our approach using GMMs may be generalised to include this additional data. The feature set we chose attempts to extract the most relevant information from target position reports only.

4 Methods

A tremendous amount of research in machine learning has focused on the problem of learning classifiers from a set of training data [17]. In our application, we use primary-only tracks that occur in Class B airspace as examples of birds to train our classifiers. Aircraft without transponders are prohibited from entering Class B airspace without a waiver, and so we expect to see primary-only radar returns from aircraft in Class B airspace only very rarely, such as in formation flight [18]. Not only are transponders typically required in Class B airspace, but they are also required within a Mode C veil that extends beyond the borders of Class B airspace. To allow some buffer for blunders into a Mode C veil, we only use tracks that occur within the lateral borders of Class B airspace as examples of birds. For examples of aircraft to train our classifier, we use tracks of transponder-equipped aircraft squawking 1200, the code used by aircraft flying under VFR without Air Traffic Control services.

One way to approach the problem of classification is to first learn a set of class-conditional probability density functions, expressed as $p(x|\omega)$, where $\boldsymbol{x}$ is a continuous (potentially multidimensional) random variable and $\boldsymbol{\omega}$ is a class. In our application, $\boldsymbol{x}$ is a vector of feature values extracted from radar tracks and $\boldsymbol{\omega}$ is either bird or aircraft. (A radar track may, in fact, be generated by phenomena unrelated to either birds or aircraft. In this paper, we use the category 'bird' for all tracks that are not aircraft.) If the components of $\boldsymbol{x}$ are real-valued, as opposed to categorical, then an appropriate representation of a class-conditional probability density is as a GMM [14]. A GMM is a density function that can be defined as a weighted sum of multivariate Gaussian density functions

$$p(\boldsymbol{x}) = \sum_{c=1}^{C} \alpha_c f(\boldsymbol{x}; \boldsymbol{\mu}_c, \Sigma_c) \tag{1}$$

where $f(\boldsymbol{x}; \boldsymbol{\mu}_c, \Sigma_c)$ is the multivariate Gaussian density with mean $\boldsymbol{\mu}_c$ and covariance matrix Σ_c. The value α_c is the weight of the cth component of the mixture with the constraints that $0 \leq \alpha_c \leq 1$ and $\sum_{c=1}^{C} \alpha_c = 1$.

One attractive feature of GMMs is that they can approximate a wide variety of densities very well with a suitable selection of Gaussian density components. Several approaches have been suggested for learning GMMs from data. If we have a collection of bird examples, we can use expectation maximisation to find $\boldsymbol{\mu}_1, \ldots \boldsymbol{\mu}_C$ and $\Sigma_1, \ldots, \Sigma_C$ that represent $p(\boldsymbol{x}|\text{bird})$, so long as we know exactly how many components to use [19]. If we do not know the number of components in advance, we can use the Figueiredo–Jain algorithm [20] or a greedy expectation maximisation algorithm [21] to infer the number of components from the data. The GMMBAYES Matlab Toolbox contains implementations of these algorithms and was used for the experiments described in this paper [22].

Once we have learned $p(\boldsymbol{x}|\text{bird})$ and $p(\boldsymbol{x}|\text{aircraft})$ from the training data, we can use the Bayes decision rule to classify tracks whose class label is unknown. The Bayes decision rule minimises the probability of misclassification by deciding $\boldsymbol{x}$ corresponds to a bird if $P(\text{bird}|\boldsymbol{x}) > P(\text{aircraft}|\boldsymbol{x})$, and $\boldsymbol{x}$ corresponds to an aircraft otherwise. Applying Bayes' rule, we arrive at an equivalent decision rule where we decide $\boldsymbol{x}$ corresponds to a bird if $P(\text{bird})p(\boldsymbol{x}|\text{bird}) > P(\text{aircraft})p(\boldsymbol{x}|\text{aircraft})$ and an aircraft otherwise. The prior probabilities $P(\text{bird})$ and $P(\text{aircraft})$ represent the fraction of tracks that are birds and aircraft, respectively. How we estimated the priors is discussed later.

There are many other algorithms for learning classifiers. In addition to those based on GMMs, we also experimented with the following algorithms:

- parzen-pnn: Learns a Parzen probabilistic neural network [23].
- svm: Learns a SVM [24].
- tree-gdi: Learns a decision tree [25] using Gini's diversity index as a splitting rule.
- tree-twoing: Learns a decision tree using the twoing rule for splitting.
- tree-deviance: Learns a decision tree using maximum deviance reduction for splitting.
- linear: Fits a Gaussian distribution to each class with a pooled estimate of covariance.
- diaglinear: Fits a Gaussian distribution to each class with a diagonal covariance matrix estimate.
- quadratic: Fits a Gaussian distribution to each class with covariance estimates stratified by class.
- diagquadratic: Fits a Gaussian distribution to each class with a diagonal covariance matrix estimate stratified by class.
- mahalanobis: Uses Mahalanobis distance with stratified covariance estimates.
- vote: Learns all of the classifiers above and then assigns a class label based on which class receives the most votes [26].

We can also learn to discriminate birds based only on positive and unlabelled examples. Such an approach is known as semi-supervised learning [27]. Although this approach is less common than regular supervised learning that uses both positive and negative examples, it has been applied to a variety of domains. In our application, we learn the density $p(\boldsymbol{x}|\text{bird})$ for the known bird tracks and for the unknown tracks $p(\boldsymbol{x})$ using the Figueiredo–Jain algorithm. Then by Bayes' rule,

$$P(\text{bird}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\text{bird})P(\text{bird})}{p(\boldsymbol{x})} \tag{2}$$

Because $p(\boldsymbol{x}|\text{bird})$ and $p(\boldsymbol{x})$ are only estimated from the data and do not actually represent the true distributions, it is possible that the estimated $P(\text{bird}|\boldsymbol{x})$ is not a valid probability value, i.e. not a value between 0 and 1. We label an instance $\boldsymbol{x}$ as a bird if $P(\text{bird}|\boldsymbol{x}) > 0.5$ and as an aircraft otherwise.

The first step in estimating the prior probability of a track being a bird, $P(\text{bird})$, is to estimate the density of bird tracks in class B airspace, ρ_{bird}, by dividing the number of PSR tracks observed in Class B by the surface area of Class B in our dataset. Assuming the density of birds is independent of airspace class, the number of bird tracks we expect to observe in our sensor coverage, which we denote n_{bird}, is then ρ_{bird} times the surface coverage of all of the sensors. We can estimate the number of non-cooperative aircraft in our dataset, n_{aircraft}, by $\alpha \cdot n_{\text{vfr}}$, where α is the ratio of tracks from non-cooperative aircraft to tracks from aircraft with transponders and n_{vfr} is the number of VFR tracks in our dataset. Although statistics on flight hours by equipage is unavailable, we may approximate α by dividing the number of registered aircraft without transponders by the number of registered aircraft with transponders, as provided by the FAA [2], to get $\alpha = 0.195$. This estimate for α is only an approximation; the ratio is a nationwide average and may vary from region to region, and the rate at which aircraft generate tracks may not be independent of transponder equipage. Finally, we can estimate $P(\text{bird})$ as

$$P(\text{bird}) = \frac{n_{\text{bird}}}{n_{\text{bird}} + n_{\text{aircraft}}} \tag{3}$$

and $P(\text{aircraft})$ is $1 - P(\text{bird})$. We estimate $P(\text{bird}) = 0.8$

doi: 10.1049/iet-rsn.2008.0182

and $P(\text{aircraft}) = 0.2$ for our dataset. Depending on the region or time period, $P(\text{bird})$ may range, conceivably, from about 0.5 to about 0.9. The effect of an incorrect prior on the accuracy of our classifier is discussed in Section 5.1.

5 Results

This section describes how we choose which subset of the features to use in our GMM-based classifier. It then explains how our GMM-based classifier compares to other supervised learning algorithms. It discusses the differences in classification when using supervised and semi-supervised learning. Finally, it discusses the potential for using the classifier in a real time system. The data from June 2008 include 95,863 VFR beacon tracks and 14,861 PSR tracks, of which 3688 are assumed to be birds because they are in the lateral borders of Class B airspace.

5.1 Feature selection

We identified nine track features, listed in Section 3, that help differentiate aircraft from birds. We derived this feature set based on engineering judgement and prior work. Each feature individually, it was hypothesised, had predictive value in terms of classification, but the optimal set of features, in terms of predictive classification ability as a whole, was unknown. As studied in the machine learning literature, the optimal set of features will often be a subset of the possible features presented to the classifier and will depend on the classifier [28, 29].

In order to identify the best combination of features for the GMM-based classifier, we evaluated each of the $2^9 - 1 = 2047$ possible feature combinations. We selected a random half of the bird tracks, or 1844 tracks, and 500 random VFR tracks to form our training set. We limited the number of VFR beacon tracks to 500 so that our training population roughly reflects the 80%/20% mix of birds to aircraft in our unknown PSR track population, although our particular algorithm is insensitive to this ratio. The other half of the known bird tracks and a different set of 500 VFR tracks represented our validation set. We then trained a classifier on each combination of features using the training set, tested each classifier with the validation set, and ranked them by expected accuracy.

Accuracy is an unbiased scalar metric that summarises the confusion matrix achieved from our sample validation population. The confusion matrix is a matrix where the rows represent different instances of predicted classes and the columns represent different instances of actual classes. Each cell in the matrix represents a count for a particular predicted and actual class. For our investigation, the confusion matrix was a two-by-two matrix, where a track could either be classified as an aircraft or bird. Table 1 shows the confusion matrix from a classifier trained on velocity, turnrate-fluc, turnrate-fluc-adj and stdev-heading.

Accuracy is calculated by

$$\frac{P(\text{bird}) \cdot C_{11}}{C_{11} + C_{21}} + \frac{(1 - P(\text{bird})) \cdot C_{22}}{C_{12} + C_{22}} \tag{4}$$

where C_{ij} is the element in the ith row and jth column of the confusion matrix. Accuracy weights the true positive and true negative rates on the validation set by the expected percentage of the respective classes in the true population. The metric

Table 1 GMM confusion matrix

Predicted	Actual		
	Bird	Aircraft	Total
bird	1752	107	1859
aircraft	92	393	485
total	1844	500	2344

Table 2 The top 10 feature sets using GMM classification, all providing 92% accuracy. Outlined is our chosen feature set

velocity	stdev-turnrate	turnrate-fluc	stdev-turnrate-adj	turnrate-fluc-adj	skipped	stdev-heading	timeofday
		×		×			
	×	×	×				
	×	×	×			×	×
	×		×	×			
	×			×	×		
×		×		×			
		×	×	×	×		
×		×		×		×	
		×	×		×	×	×
		×		×	×		

doi: 10.1049/iet-rsn.2008.0182

reduces the bias which may exist due to an unrealistic mix of aircraft and birds in the validation set. For example, if the validation set contained ten times as many aircraft tracks than bird tracks, we would not want to overstate the accuracy of a classifier that was only strong at correctly identifying birds. The confusion matrix in Table 1 represents an accuracy of 92%, when $P(\text{bird}) = 0.8$. Of course, we do not know $P(\text{bird})$ exactly, as mentioned in Section 4. If $P(\text{bird})$ is really 0.5, the accuracy for Table 1 is 87%. If $P(\text{bird}) = 0.9$, the accuracy is 93%.

The feature set used to produce the confusion matrix in Table 1 is one of highest scoring in our tests. However, other feature combinations may be just as valid, and produce similar levels of accuracy, as shown in Table 2. The highlighted feature set in Table 2 was chosen because it included features that are the most independent of each other and capture most of the variation between birds and aircraft. The full distribution of scores over all possible feature sets is shown in Fig. 3.

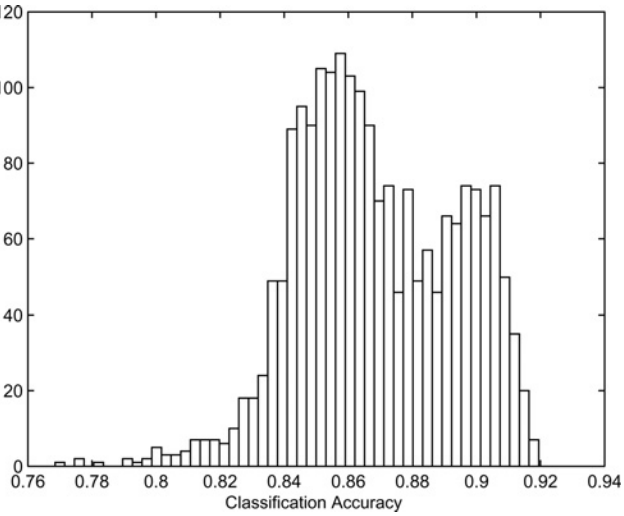


Figure 3 *Distribution of GMM classification accuracy from the 2047 possible feature combinations*

5.2 Classifier comparison

We tested our GMM classifier against all of the supervised learning classifiers listed in Section 4. GMM performed the best in terms of accuracy, but other methods produced comparable results using the same feature, training and testing sets (Table 3).

It required 1.76 s to train our GMM classifier on the training set of 1844 bird tracks and 500 aircraft tracks using a 3 GHz Intel Pentium 4 with 4-GB memory (Table 3). Although training the GMM required more time than other methods, the computation required is acceptable, especially since training is done offline. Classification performance is more important. Classifying the 2344 tracks in the evaluation set required 0.01 s, which is significantly faster than the SVM, neural network and decision-tree methods.

We further tested two other methods to see whether the results would be sensitive to the feature set selected. We investigated tree-deviance because it had the next highest score to GMM and SVM because it was used in the most recent work on this topic [12]. We found that the optimal results for those two classifiers, shown in Table 4, still fall short of GMM, but they are close enough to be considered viable alternatives.

Table 3 Classifier results

Classifier	Accuracy		Computation	
	Training	Validation	Training (s)	Validation (s)
gmm	0.92	0.92	1.76	0.01
vote	0.88	0.91	3.81	2.83
tree-deviance	0.98	0.91	0.22	0.18
tree-gdi	0.97	0.90	0.21	0.18
tree-twoing	0.97	0.90	0.19	0.18
diaglinear	0.87	0.87	<0.01	0.02
diagquadratic	0.87	0.87	<0.01	0.02
linear	0.87	0.86	<0.01	0.02
quadratic	0.86	0.86	<0.01	0.01
mahalanobis	0.86	0.85	<0.01	0.02
svm	0.96	0.84	1.38	1.35
parzen-pnn	0.84	0.83	0.01	0.83

doi: 10.1049/iet-rsn.2008.0182

Table 4 Peak accuracy and optimal feature set of selected classifiers

Classifier	velocity	stdev-turnrate	turnrate-fluc	stdev-turnrate-adj	turnrate-fluc-adj	skipped	stdev-heading	timeofday	duration	Accuracy
gmm	×		×		×		×			0.92
svm				×	×	×			×	0.91
tree-deviance	×	×		×	×	×		×		0.91

5.3 Supervised and semi-supervised learning

Fig. 4 shows the distribution of stddev-heading against velocity for the training set and the categorised test set. (We could have included similar figures for comparing the joint distributions of other pairs of features, but we found the variation for the joint distribution of velocity and standard deviation of heading to be the most illustrative.) The supervised approach generally classifies slow straight PSR tracks as birds, and other tracks as aircraft. The classification results correlate to similar patterns observed in the training set. One notable exception is the correct classification of high-velocity tracks as aircraft, despite the fact that high velocity aircraft (above 250 kt) are not observed at all in the VFR training data set. High-velocity PSR tracks are believed to be from military aircraft flying in formation with their transponder in stand-by mode, according to military operating procedures [18].

There are some aircraft that fly neither like birds nor like transponder-equipped aircraft. By excluding positive examples of transponder-equipped VFR aircraft, it was hypothesised that a semi-supervised classifier could detect unconventional aircraft such as ultra-lights and gliders among others. Indeed, more tracks were classified as aircraft using semi-supervised classification than using supervised classification because the track had to simply behave as a 'non-bird', rather than also 'like an aircraft'. Thus, PSR tracks that are classified differently using the two approaches are tracks that exhibit behaviour unlike that of birds or aircraft, and so may be unconventional aircraft.

Figs. 5*a* and *b* show how the tracks from the radar picture shown in Fig. 2*b* were classified using supervised and semi-supervised learning, respectively, and Fig. 5*c* shows the unusual tracks resisting consistent classification. Fig. 5*c* shows that, in some cases, semi-supervised learning was able to detect tracks that may be the result of unconventional aircraft. For instance, semi-supervised learning classified as 'non-bird' some flight activity near a remote control airplane club that supervised learning classified as birds.

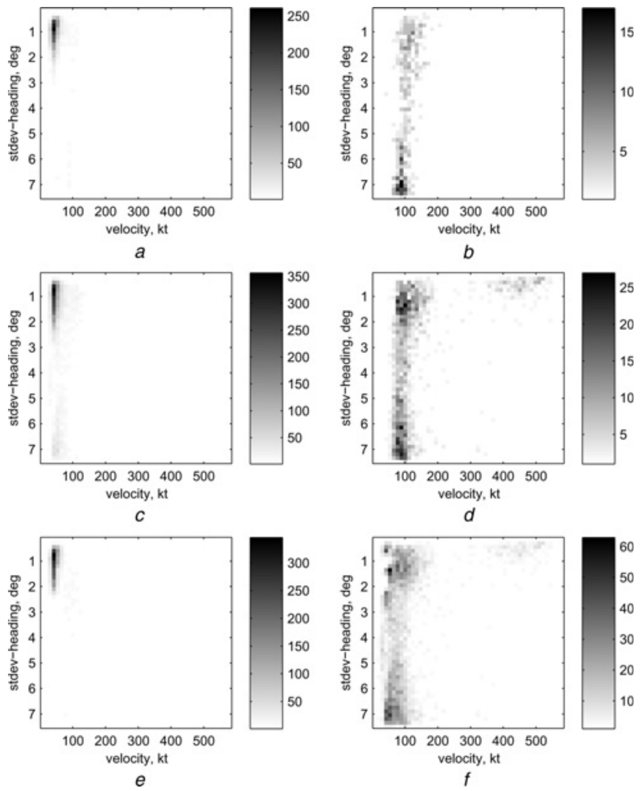


Figure 4 *Distribution of velocity and stdev-heading for bird and non-bird tracks*

a Birds training
b VFR training
c Birds supervised
d VFR supervised
e Birds semi-supervised
f VFR semi-supervised

5.4 Real-time classification

For many applications, one is only interested in classifying tracks after observing the entire extent of the track. This section describes how the GMM-based classifier was used for real-time classification.

We tested our GMM-based, supervised-learning classifier against increasing amounts of exposure time to the tracks in the test set. By performing this experiment, we hoped to evaluate how well the classifier would

Figure 5 *Example classifications for tracks for the Buffalo, New York sensor from 9 June 2008*

Airport locations are denoted with stars
a Supervised track classification (birds: grey; aircraft: black)
b Semi-supervised track classification (birds: grey; non-birds: black)
c Non-bird, non-aircraft

perform against a challenging set of tracks in a simulated real-time environment. Fig. 6 shows the results of our experiment.

If we simply assume every track is due to birds, we expect to achieve an accuracy of around 80% because we estimate 80% of the PSR tracks to be birds. Classification performance does not significantly increase until we have at least a minute of track information, corresponding to about 12 radar scans on an ASR-9. After approximately 3 minutes of tracking, we achieve an accuracy near our limit of 92% achieved with the full track history.

The results in Fig. 6 indicate promise for fairly accurate real-time target classification based only on position reports. Any real-time system, for example a situational awareness system, would need to be tailored for the local environment of the sensor. Tailoring the system, including both the prior and the feature set, on local data would improve classification accuracy. In terms of the prior, it may be the case that bird density varies geographically. The prior belief that a track is due to a bird in a given geographic region may be closer to 0.5 than 0.8, for example. In terms of the feature set, it may be the case that, for example, flocks of birds on the east coast tend to fly more with the wind due to migration patterns than they do in Texas, where migration patterns are not as pronounced. In this case, the feature 'wind' may be included in a classifier for the Boston sensor, but not for the Ft. Hood sensor. With an appropriate amount of data for a given geographic region, optimal classifiers specific to that region can be devised using the same methods and steps described.

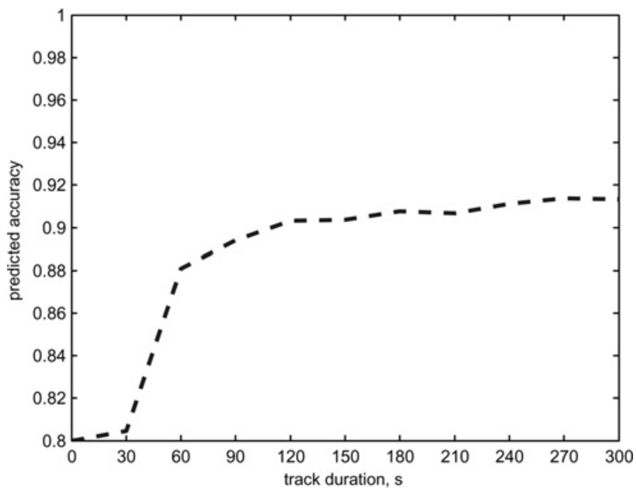


Figure 6 *Simulated real-time classification results*

6 Conclusions and further work

Previous research on target classification exploited radar features, such as cross-section measurements, not typically recorded by the FAA. This study demonstrated that target classification based on target position alone is sufficient for discriminating aircraft from birds. We found that classifiers based on Gaussian mixture models performed the best, but some of the other classifiers trained using supervised learning performed nearly as well.

Prior work on primary target classification focused on supervised learning methods where aircraft beacon tracks were used as examples of true aircraft tracks. The problem with relying upon beacon tracks for training is that some categories of non-cooperative aircraft do not have the same characteristics as transponder-equipped aircraft. We demonstrated that GMMs can be used in a semi-supervised learning context where only known bird tracks are used for training. Using semi-supervised learning, we were able to identify instances of unconventional aircraft that resist proper classification using conventional supervised learning methods.

The real-time application of our classifiers showed relatively fast convergence in classification accuracy. Ways

doi: 10.1049/iet-rsn.2008.0182

to improve accuracy include training a classifier based on local data to capture regional differences in behaviour and density. The real-time application of target classification for situational awareness would also require further analysis of misclassification cost.

Future work will include using models of different categories of non-cooperative aircraft, such as gliders and ultra-lights, to further classify targets. Probabilistic dynamic models have been developed for a wide variety of different kinds of non-cooperative aircraft [30]. After we use the GMM-based classifier to determine whether a track is indeed an aircraft, we compute the likelihood the altitude, airspeed, turn rate and vertical rate associated with the track were generated given the various models. By multiplying the likelihood by a prior over the aircraft categories and normalising, we can obtain a posterior distribution over the aircraft categories. The track is then classified as being generated by the aircraft with the maximum posterior probability.

The parameters of the GMMs and the data used to train and test the classifiers are available in electronic form from the authors.

7 Acknowledgments

The authors greatly appreciate the support provided by John Appleby from the Department of Homeland Security and comments from James K. Kuchar, Melvin L. Stone, James P. Chryssanthacopoulos and Robert E. Boisvert from MIT Lincoln Laboratory. The feedback of the anonymous reviewers is also appreciated.

This work is sponsored by the Air Force under Air Force Contract no. FA8721-05-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

8 References

[1] HAYKIN S., STEHWIEN W., DENG C., WEBER P., MANN R.: 'Classification of radar clutter in an air traffic control environment', *Proc. IEEE*, 1991, **79**, (6), pp. 742–772

[2] Federal Aviation Administration, 'General aviation and Part 135 activity surveys', http://www.faa.gov/data_statistics/aviation_data_statistics/general_aviation/CY2006/, accessed December 2008

[3] EASTWOOD E.: 'Radar ornithology' (Methuen & Company, London, England, 1967)

[4] VAUGHN C.R.: 'Birds and insects as radar targets: a review', *Proc. IEEE*, 1985, **73**, (2), pp. 205–227

[5] DENG C., HAYKIN S.: 'A multi-layer neural network classifier for radar clutter'. Proc. Int. Joint Conf. Neural Networks, 1990, vol. 1, pp. 241–246

[6] HAYKIN S., DENG C.: 'Classification of radar clutter using neural networks', *IEEE Trans. Neural Netw.*, 1991, **2**, (6), pp. 589–600

[7] JOUNY I.: 'Description of radar targets using the bispectrum', *IEE Proc. Radar, Sonar Navig.*, 1994, **141**, (3), pp. 159–163

[8] KUMAR P., PRABHU K.: 'Classification of radar returns using Wigner-Ville distribution'. Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1996, vol. 6, pp. 3105–3108

[9] NECHVAL N.A.: 'Algorithm for statistical classification of radar clutter into one of several categories', *Data Struct. Target Classif.*, 1991, **1470**, (1), pp. 282–293

[10] STEHWIEN W., HAYKIN S.: 'A statistical radar clutter classifier'. Proc. IEEE National Radar Conference, 1989

[11] KARP D., ANDERSON J.R.: 'Moving target detector (Mod II) summary report'. Project Report ATC-95, MIT Lincoln Laboratory, 1981

[12] GHADAKI H., DIZAJI R.: 'Target track classification for airport surveillance radar (ASR)'. Proc. IEEE Conf. Radar, 2006, pp. 136–139

[13] DIZAJI R., GHADAKI H.: 'A classification system for radar and sonar applications'. United States Patent Application 20070024494, February 2007

[14] MCLACHLAN G., PEEL D.: 'Finite mixture models' (Wiley, New York, 2000)

[15] KOCHENDERFER M., KUCHAR J., ESPINDLE L., GRIFFITH J.: 'Uncorrelated encounter model of the national airspace system version 1.0'. Project Report ATC-345, MIT Lincoln Laboratory, 2008

[16] BERENS P.: 'Circular statistics toolbox', http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=10676, accessed December 2008

[17] DUDA R.O., HART P.E., STORK D.G.: 'Pattern classification' (Wiley, New York, 2000, 2nd edn.)

[18] Air Force Instruction 11-202: 'General Flight Rules', April 2006

[19] BILMES J.: 'A gentle tutorial on the EM algorithm including Gaussian mixtures and Baum-Welch'. Technical Report 97-021, International Computer Science Institute, 1997

[20] FIGUEIREDO M., JAIN A.: 'Unsupervised learning of finite mixture models', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (3), pp. 381–396

[21] VERBEEK J.J., VLASSIS N., KRÖSE B.: 'Efficient greedy learning of Gaussian mixture models', *Neural Comput.*, 2003, **15**, (2), pp. 469–485

[22] PAALANEN P., KAMARAINEN J.-K., ILONEN J., KÄLVIÄINEN H.: 'Feature representation and discrimination based on Gaussian mixture model probability densities – Practices and algorithms', *Pattern Recognit.*, 2006, **39**, (7), pp. 1346–1358

[23] LOMBARDI G.: 'Parzen PNN', http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=11880, 2006, accessed December 2008

[24] CHANG C.-C., LIN C.-J.: 'LIBSVM: a library for support vector machines', http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001, accessed December 2008

[25] BREIMAN L., FRIEDMAN J.H., OLSHEN R.A., STONE C.J.: 'Classification and regression trees' (Wadsworth, Belmont, California, 1984)

[26] KITTLER J., ALKOOT F.M.: 'Sum versus vote fusion in multiple classifier systems', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, pp. 110–115

[27] CHAPELLE O., SCHOÖLKOPF B., ZIEN A.: 'Semi-supervised learning' (MIT Press, Cambridge, Massachusetts, 2006)

[28] KOHAVI R., JOHN G.H.: 'Wrappers for feature subset selection', *Artif. Intell.*, 1997, **97**, pp. 273–324

[29] GUYON I., ELISSEEFF A.: 'An introduction to variable and feature selection', *J. Mach. Learn. Res.*, 2003, **3**, pp. 1157–1182

[30] EDWARDS M., KOCHENDERFER M., KUCHAR J., ESPINDLE L.: 'Encounter models for unconventional aircraft version 1.0'. Project Report ATC-348, MIT Lincoln Laboratory, 2008

doi: 10.1049/iet-rsn.2008.0182