

# A feature extraction method for automatic identification of flying targets from radar data

M.T. Nguyen\*, A.T. Nguyen\*, V.D. Hoang\*\*, D.D. Lam\*

\*C4I Center, VHT, Viettel Group, Hanoi 10000, Vietnam

\*\*Faculty of Information Technology, Ho Chi Minh City University of Technology and Education, Ho Chi Minh, Vietnam

tiemnm@viettel.com.vn, tuanna@viettel.com.vn, dunghv@hcmute.edu.vn, duongld11@viettel.com.vn

**Abstract**— Flying target identification is an important and challenging problem in airborne surveillance systems. In this report, we propose a method to automatically identify the flying targets by physical information (coordinates, heading, speed), time, and identification information (3/A code). This method includes two steps: features extraction and building a machine learning model. In the features extraction step, features that are extracted include cell indexes corresponding to coordinates, information of flight path, time in day/night format, heading, speed, and 3/A code, constructing n-dimensional vector. This vector is used as input for training a Random Forest model, to automatically identify class labels of flying targets.

**Keywords**— Machine learning, flying object identification, features extraction, random forest

## I. INTRODUCTION

In military, target identification is often conducted manually and based on personal experience. The goal is mainly focused to clarify the target allegiance and its class like ally or enemy, bomber, fighter or commercial jet, etc. There are some proposed methods to identify the flying objects in the world, however, there remained certain drawbacks. Those methods require specifying some basic attributes such as the flight path of the target. Some early works determine the target class based on kinematic features such as the maximum speed or estimated acceleration. The most popular way to identify the flight path of a flying target is to calculate the distance from the target to all available flight paths and choose the nearest one.

Flying target identification is a more complex and challenging problem because flying objects may have different types of motion (straight, turns, decelerations, etc.) on each segment of trajectory and their flight parameters change rapidly in a short period time.

Information for target classification can be obtained from high-resolution radar, electronic support measure. S. Challa *et al* reported a unified framework for the joint treatment of target tracking and target classification with two sources of data from radar and electronic support measure. Machine learning techniques have been focusing on prediction at an aggregate level rather than on the detail level of trajectories. There are many approaches to classify trajectory data in the literature through 3 steps: trajectory preparation, feature extraction, and

classification. Representing trajectories in a unified length may cause some loss of information. After several works have to use some distance measurements like Euclidean Distance, Hausdorff Distance, or Bhattacharyya Distance to evaluate how close a trajectory is to the pre-defined ones[7].

However, this method is expensive in computation due to orthographic projection calculations. It also requires searching for the nearest flight path whenever the coordinates of the flying target are updated. Moreover, when the target is near the airport or the intersection point of flight paths, it is not certain that the nearest flight path is the one that the target follows.

To overcome these difficulties, many works have been focusing on finding sophisticated machine learning models that can improve the prediction performance. More recently, several deep learning models have been used to link GPS sub-trajectories to classified activities[8-11]. However, the major limitation of this approach is the multidimensional data problem.

In this report, we propose a new method of feature extraction that could automatically identify flying targets with a simple Random Forest machine learning model. We process the pre-preparation step based on the features of motion information, time, and 3/A code. This proposed method will optimize the process of target identification, reduce human errors, and improve the accuracy of results.

## II. TECHNICAL METHOD

Our purpose is to propose a method to automatically identify a flying target based on its information of motion, time, and 3/A code. The 3/A code is the identification information that contains 4-digit code which is assigned for each flying target based on the convention of aviation to distinguish among the flying targets. This method is used to classify and label the targets corresponding with their activities including military/domestic/transit/international. We apply Random Forest technique with the flying target's feature information of longitude, latitude, heading, speed, time, and 3/A code. From the feature space of the flying target built from physical information (longitude, latitude, heading, and speed), time, and identification information (3/A code), we will construct a random forest containing multiple decision trees.

Identification labels include military (flying targets being planes which belong to Ministry of Defense to perform

mission), domestic (flying targets being planes with domestic flight route), transit (flying targets being foreign planes transiting in current's country airspace), international (flying targets being planes with flight route between the domestic airport and the foreign airport).

### A. Step 1: Features extraction

This step is performed to construct feature spaces from input data. The original input data of the flying target include coordinates, heading, speed, time, and 3/A code (that information is only at an exact time). This information is too detailed, hence not suitable for being the feature of the data. As such, we need to generalize this information through the feature extraction process.

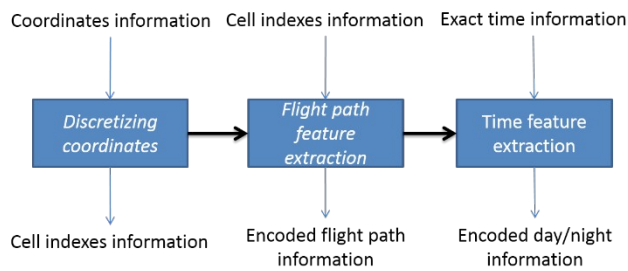


**Figure 1.** The block diagram explaining the feature extraction process of a flying target.

With the reference to figure 1, which describes the diagram of extracting features. Features extraction process includes minor steps below:

- Discretizing the coordinates information: coordinates of flying target are too accurate, hence they will cause over-fitting problems when being applied to the model. To handle this problem, we proposed to discretize that information in the longitude-latitude grid. The discretization method is explained as below:

- Constructing the discretized longitude-latitude grid including the cell with longitude width  $\Delta\rho$  and latitude width  $\Delta\theta$ . Those values of width are adjustable. In our implementation, the value of latitude width  $\Delta\theta=0,5^\circ$  and longitude width  $\Delta\rho=0,5^\circ$  give the best result. The coordinates of the flying target will be replaced with cell indexes that the plane belongs to, which means we map the plane's coordinates to the cell indexes. For instance, if the coordinates of the flying targets are  $102^\circ12'50''$  ( $102,2138^\circ$ ) and  $21^\circ48'34''$  ( $21,8094^\circ$ ), the flying target will locate in cell (204, 43).

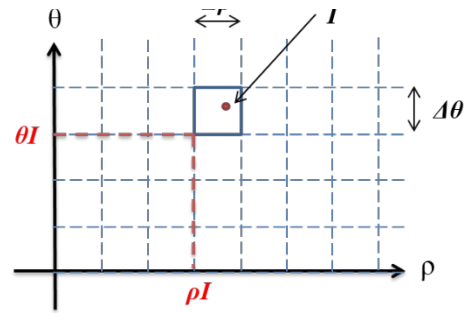


**Figure 2.** The diagram describing coordinates discretization method in longitude-latitude grid

With reference to figure 2 describing the discretized latitude-longitude grid, the result when mapping the flying target's coordinates to cell indexes is:

$$102,2138 / 0,5 = 204,4276$$

$$21,8094 / 0,5 = 43,6188$$

Extracting flight path information: flight-path is also a good feature reflecting the character of the flying target. The identification labels of flying targets correlate with the characteristics of those flight paths. For instance: domestic flights must operate along the flight paths between domestic airports; international flights must operate along the flight paths which cross the country; transit flights must operate along the flight paths which connect domestic airports and foreign airports.

However, as mentioned above, the method determining the optimal flight path from the set of many available flight paths has a drawback regarding the expensive cost of computation due to the orthographic projection calculations. To solve this problem, we suggest the method to extract information from the flight path by supplementing the information of the flight path crossing each cell in the discretized latitude-longitude grid. For example, if cell (0, 0) has 3 flight paths A, B, C cross it, the flight path information of the cell (0, 0) is (A, B, C)

Determining the flight path of the flying target as the flight path of the cell where the coordinates of the flying targets belong to. This proposed method reduces the cost of computation significantly, as we have already constructed the flight path information of the cells, thus there is no need to recalculate the flight path when the flying target's coordinates change. Additionally, each cell contains information on multiple flight paths, which allows assigning flying targets to a set of flight paths instead of only a single one.

The information of flight paths are stored in text format, thus we need to convert them into the numeric format as below:

Each flight path is assigned to a dimension of feature space, it is encoded by binary method with two values 0 and 1. To be specific: if the flying target does not follow a particular flight path, this one is encoded as 0; if the flying target follows a particular flight path, this one is encoded as 1. For example, if

we have totally three available flight paths: A, B, and C, hence the feature space will have three dimensions. If the flying target is a plane following the A path, its feature will be encoded as (1,0,0), similarly, the B and C one will be encoded as (0,1,0) and (0,0,1) and (1,1,1) for the whole three paths. In our work that applied in Vietnam, there are 78 flight paths, hence flight path information for each plane flying target will be encoded in 78 dimensions of the feature space.

Extracting time information is similar to the location information. The time information of the flying target in each coordinate point is too detailed, which is not suitable for the training phase of the model. We hence propose converting the time format from hour/minute/second format to day/night one, in which, day is assigned for time range from 6 am to 6 pm, and night is assigned for time range from 6pm to 6am. Day value is then be encoded as 1, while night value is encoded as 0.

This method is extremely good at reflecting the characteristics of Military and Domestic flights, which tend to operate in daytime. The remaining information (heading, speed, 3/A code) is used directly as input features.

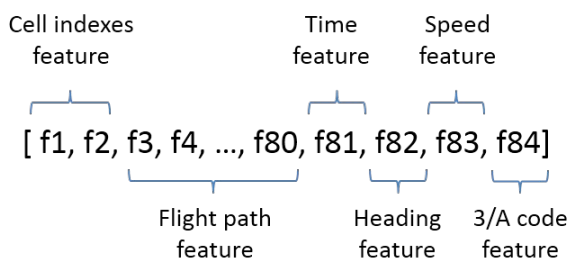


**Figure 3.** An illustration of feature vector representing dimensions of feature space

In an aspect of the implementation of this work, with reference to figure 3, after the feature extraction step, features extracted include cell indexes, flight path, day/ night flight, heading, speed, and 3/A code. We combine them to construct an 84-dimensional vector, which contains: 2 dimensions for discretized longitude-latitude cell indexes, 78 dimensions for encoded flight paths, 1 dimension for encoded timing day/night, 2 dimensions for heading and speed, and the last one for 3/A code.

### B. Step 2: Voting for identification

After extracting features, the next step is to build a machine learning model to automatically identify the class label of the flying targets. The set of class labels include military, domestic, transit, and international. The main technique of our method is Random Forest, which is a machine learning-based method, building ensemble decision trees. Classification is performed by traversing through output decision corresponding to attribute value from internal nodes of the tree until we reach a leaf node to decide the final class label. Random Forest technique is proposed in this research to reduce the variances for the decision tree technique. Random Forest uses a combination of decision trees instead of only single one to

improve the stability and accuracy of the classification result. figure 4 shows the overview of Random Forest method.

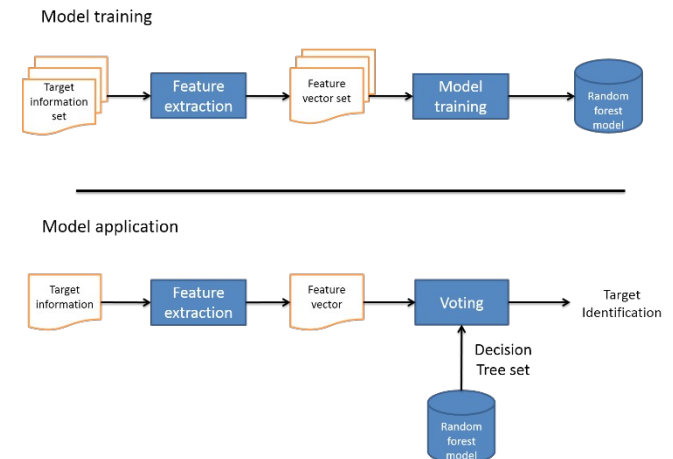


**Figure 4.** Diagram demonstrating the training and application phases of the model

The 84-dimension vectors which were extracted from step 1 will be the input for training the Random Forest model. Given the training dataset  $S$  of size  $n$ , the training phase of Random Forest is conducted as below:

Split the dataset  $S$  into  $m$  subsets  $S_i$ ,  $i = 1, \dots, m$ , each subset has the size of  $n'$ , by sampling from  $S$  uniformly and with replacement. Sampling with replacement is a technique allowing some examples from original set  $S$  to be repeated in subsets  $S_i$ . If  $n = n'$ , then for large  $n$  the set  $S_i$  is expected to have the fraction  $(1 - 1/e) \approx 63.2\%$  of the unique examples of  $S$ , the rest will be duplicated. The detail of the training phase is explained as below: Generating randomly subsets  $S_i$  from the original dataset  $S$  by sampling with replacement; Training those subsets with decision tree model; Calculating the average prediction result for each new subset.

When determining the best features for splitting data at the nodes of decision trees, we only choose  $m$  features among the total of  $M$  features, regularly  $m = \sqrt{M}$ , a typical value for classification problems. In our implementation,  $m = \sqrt{84} \approx 9$ . After training the Random Forest model, to identify the class labels of the flying targets, we conduct extracting features of the flying targets and traversing them through each decision tree of the Random Forest. Each decision tree will give a vote for a class label with the input data are 84-dimensional vectors. Aggregating the vote on each decision tree, we finalize the class label of the flying target based on majority voting. The class label which received the most votes is the class label of the flying target. In case there are class labels that received the same number of votes, we choose one label randomly.

## III. RESULTS AND DISCUSSION

After using this model for a series of input data, the results are recorded in table 1.

TABLE I. PERFORMANCE AND THE ACCURACY OF THE MODEL

| Train    | Test    | By Gini  |              | By Entropy |              |
|----------|---------|----------|--------------|------------|--------------|
|          |         | Time (s) | Accuracy (%) | Time (s)   | Accuracy (%) |
| 4367478  | 1441    | 0.01404  | 95.84        | 0.01365    | 95.24        |
| 4367478  | 2251452 | 13.7462  | 96.8         | 13.58762   | 96.60        |
| 6455685  | 2251452 | 8.8391   | 96.88        |            |              |
| 10835288 | 4514529 | 20.6388  | 97.37        |            |              |

There are some remarkable observations. Calculation using Gini index is faster than that using entropy. With the small number of objects (1441 targets), the difference in computation time is not significant, but with a large number of objects (2251452 targets), we can see a profound difference (5s). The accuracy of the classification results in these two methods reaches the same level.

Similar to the other supervised learning methods, the larger the training data set is, the more accurate the model is. From our experiment with 4306373 data in the training set, we obtained the result with the accuracy of 93.13 % after 0.0011 seconds for the Gini index and with the accuracy of 91.46 % after 0.0012 seconds for entropy criteria. When we increase the size of the training data set to 1093628, the accuracy reaches 97.36 %, which is much higher than that of the Random Forest method in the work of C.S. Bosson *et al*[12]. This demonstrates that with proper feature extraction, Random Forest can provide a reliable model for flying target classification.

#### IV. CONCLUSIONS

We have reported a comprehensive method of a machine learning model to classify flying objects with high accuracy and quick response. Our system is in an online learning regime with increasing labeled data that are accumulated over time. We have implemented a system with fully automatic processes from collecting raw information of the objects, discretizing the geodesic positions, extracting the features, training, and testing. The accuracy of more than 97 % is applicable for our problem and is improving during daily operations.

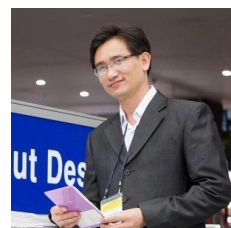
#### REFERENCES

- [1] B. Ristic, N. Gordon, and A. Bessell, "On target classification using kinematic data," *Inf. Fusion*, vol. 5, no. 1, pp. 15–21, 2004, doi: 10.1016/j.inffus.2003.08.002.
- [2] W. A. B. J.G. Tetti, R.P. Gorman, "A Multifeature Decision Space Approach to Radar Target Identification," vol. 32, no. 1, pp. 0–7, 1996.
- [3] H. Leung and J. Wu, "Bayesian and Dempster-Shafer target identification for radar surveillance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 432–447, 2000, doi: 10.1109/7.845221.
- [4] J. Bian, D. Tian, Y. Tang, and D. Tao, "Trajectory data classification: a review," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 4, 2019, doi: 10.1145/3330138.
- [5] R. Fung, "Target identification with Bayesian networks in a multiple hypothesis tracking system," *Opt. Eng.*, vol. 36, no. 3, p. 684, 1997, doi: 10.1117/1.601266.
- [6] S. Challa and G. W. Pulford, "Joint target tracking and classification using radar and ESM sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 3, pp. 1039–1055, 2001, doi: 10.1109/7.953266.
- [7] Q. Gao, F. Zhou, K. Zhang, G. Trajcevski, X. Luo, and F. Zhang, "Identifying human mobility via trajectory embeddings," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 0, pp. 1689–1695, 2017, doi: 10.24963/ijcai.2017/234.
- [8] J. Bian, D. Tian, Y. Tang and D. Tao, "Trajectory Data Classification: A review," *ACM Trans. Intell. Syst. Technol.* 10, 4, Article 33, 2019, doi: 10.1145/3330138.
- [9] F. Zhou, Q. Gao, G. Trajcevski, K. Zhang, T. Zhong, and F. Zhang, "Trajectory-user linking via variational autoencoder," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2018-July, pp. 3212–3218, 2018, doi: 10.24963/ijcai.2018/446.
- [10] F. Zhou, R. Yin, G. Trajcevski, K. Zhang, J. Wu, and A. Khokhar, "Improving human mobility identification with trajectory augmentation," *Geoinformatica*, vol. 25, no. 3, pp. 453–483, 2021, doi: 10.1007/s10707-019-00378-7.
- [11] N. C. A. de Freitas, T. L. Coelho da Silva, J. A. F. de Macêdo, L. M. Junior, and M. G. Cordeiro, "Using deep learning for trajectory classification," *ICAART 2021 - Proc. 13th Int. Conf. Agents Artif. Intell.*, vol. 2, no. Icaart, pp. 664–671, 2021, doi: 10.5220/0010227906640671.
- [12] C. S. Bosson and T. Nikoleris, "Supervised learning applied to air traffic trajectory classification," *AIAA Inf. Syst. Infotech Aerospace*, 2018, no. 209989, 2018, doi: 10.2514/6.2018-1637.



**Nguyen Manh Tiem** received his B.Sc in 2012 and PhD degrees from Irkutsk National Research Technical University (INRTU), Russia in 2017. He is currently an algorithm research engineer at Viettel High Technology and Industries Corporation (VHT) – Viettel Group. His main areas of research include artificial intelligence, machine learning.

**Nguyen Anh Tuan** received the BE. and MSc. in Information Technology from Hanoi University of Science and Technology, Vietnam, in 2011 and 2018, respectively. He is working as the Chief Engineer at C4I Center, Viettel High Technology Industries Corporation (VHT) – Viettel Group. Mr. Tuan's main area of research is target tracking for military surveillance and reconnaissance system.



**Hoang Van Dung** received the Ph.D. degree from the University of Ulsan, South Korea, in 2015. He has been serving as a Professor in computer science with AI Laboratory, Faculty of Information Technology, Ho Chi Minh City University of Technology and Education, Vietnam. He has published numerous research articles in ISI, Scopus indexed, and high-impact factor journals. His research interests include a wide area, which focuses on pattern recognition, machine learning, medical image processing, computer vision application, vision-based robotics and ambient intelligence, and communication networks.



**Lam Duc Duong** received the B.Sc from Vietnam National University in Hanoi with Engineering Physics and Nanotechnology. After he obtained M.Sc and PhD in Engineering Science from Osaka University Japan in 2011 and 2014, respectively. From 2015 to 2017, Dr. Duong worked in AIST (Japan) and University of Paris 11 (France) as a postdoc. Currently, he is working at C4I Center, Viettel High Technology Industries Corporation (VHT) – Viettel Group. His research interests include machine learning, data mining and operation research.