Research on Auxiliary Target Extraction and Recognition Technology in Big Data Environment

Gui Liu(✉), Yongyong Dai, Haijiang Xu, Wei Xia, and Wei Zhou

Jiangnan Institute of Computing Technology, Wuxi, Jiangsu, China

Abstract. With the rapid development of global aviation industry, the flight volume is rising rapidly, and the aviation industry is also rapidly entering the era of big data. In this paper, the digital extraction of image information assisted target recognition technology is used to extract the route and waypoint information on the picture and make it into vector civil aviation route. Combined with the massive flight plan and target trajectory on the Internet, track feature matching is realized to assist users to quickly discover targets.

Keywords: Track feature matching · Track deviation

1 Introduction

With the rapid development of global aviation industry, the flight volume shows a rapid upward trend, and the aviation industry has rapidly entered the era of big data [1, 2]. In this paper, based on the massive vector civil aviation routes, combined with flight plan and target trajectory, track feature matching is realized, which provides a new way for users to quickly capture key targets and find targets.

2 Research Status at Home and Abroad

Using electronic map information query function, computer high-speed calculation and decision-making function to automatically generate planned route is a hot research topic in civil aviation field. There are many methods for automatic route generation, such as Dijkstra algorithm, heuristic search algorithm, fuzzy algorithm, neural network, and relatively new ant colony algorithm and genetic algorithm control [3–7].

Ant colony algorithm is easy to search, but it has a long time and slow convergence, and it is easy to stagnate and fall into local optimum; genetic algorithm has fast convergence and good optimization effect, but it is difficult to obtain the initial population and the coding is complex. With the development of digital information, a large number of map images with route and waypoint information can be obtained from the Internet. If the information of route and waypoint on the picture can be extracted digitally and made into an electronic map with vector civil aviation route, the track feature matching can be realized by combining the trajectory of aircraft flight target, which can assist users to quickly discover air targets.

X. Sun et al. (Eds.): ICAIS 2021, LNCS 12736, pp. 659–665, 2021.
https://doi.org/10.1007/978-3-030-78609-0_55

3 Digital Extraction of Image Information to Assist Target Recognition

3.1 Registration and Digital Extraction of Civil Aviation Routes

Traditional maps are passive for users, and they cannot choose the expression form of maps according to their requirements. Digital maps allow users to choose scale, projection data, symbols, colors, schemata, etc. according to their own design. The introduction of digital extraction technology of map data can solve the digital problems such as poor scaling effect and unclear of civil aviation route map, which is of great significance to improve user experience.

The registration of civil aviation route mainly includes two processes: selecting control point and obtaining coordinate information of control point. The representative aviation control points, such as large airports and route intersections, are selected as control points, and it is better to have more than three evenly distributed control points. The more accurate the coordinate is, the more accurate the vector information is obtained by reference map.

With the registration control points and coordinate information, a new ShapeFile file is established by using the map digitization extraction tool, and the ShapeFile file type is selected according to the point, line and surface attributes to be digitized. Setting appropriate projection and coordinate system parameters, the coordinate of civil aviation route map is CGCS2000 coordinate system. A new layer is created in the extraction tool, and the registration control point information and layer information are input, and the features to be digitized are selected, and then the digital extraction can be realized directly. In order to facilitate the use and management, different types of information need to be classified and digitized. For example, after the digital extraction of civil aviation route map, it is divided into three shapefiles: point element, point element buffer and route.

3.2 Target Recognition Based on Target Trajectory Feature Matching

The target trajectory can be used as an important basis for the research and judgment of the identity of the target. If the user can quickly judge the attribute of the target through the trajectory, he can quickly locate the important target, and further analyze and mine the deep-seated motion law. Generally, civil aviation must strictly follow the established route. When a certain target deviates from the center line of the route by a certain distance or obviously does not fly along the route, it can be regarded as an aircraft with unknown intention. Although some aircraft will generally follow the route when actual fly, they will not strictly follow the route when some waypoints and routes are changed. In this paper, based on vector civil aviation routes, combined with the flight plan of Internet traffic volume and target trajectory, track feature matching is realized, as shown in the figure below, which provides a new way for users to quickly capture key targets and find targets (Fig. 1).

Resampling of Track Pattern Based on Proximity Tracking. For an original target track sample, the initial feature points are sampled intensively, which is difficult to meet the requirements of subsequent track feature matching. Therefore, it is necessary to

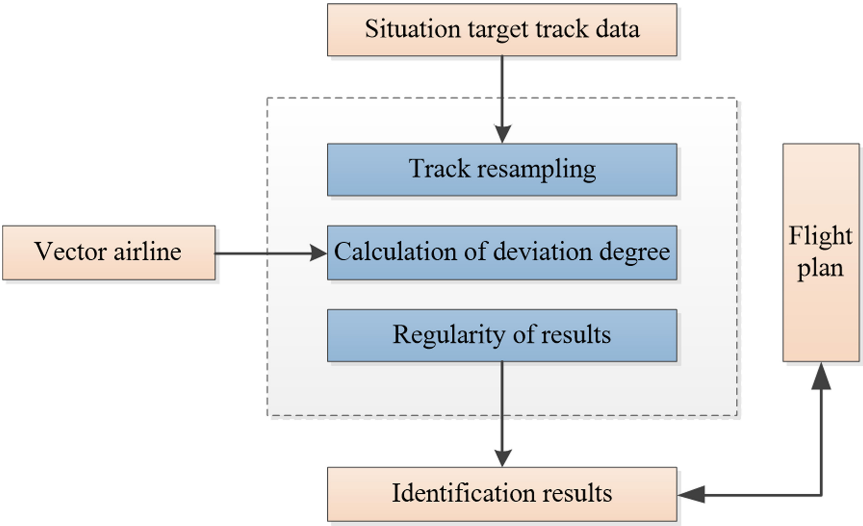

Fig. 1. Sketch map of digital extraction effect

resample the initial feature points. The resampling of track pattern based on proximity tracking not only improves the smoothness of the track, but also ensures the shape of the track and eliminates the undesirable twists and turns, thus ensuring a better track model quality.

The basic idea of neighborhood tracking resampling is: firstly, the spatial partition strategy is used to divide the initial graph data points to improve the efficiency of neighborhood search. Secondly, the optimal tracking idea is used to determine the initial point of the tracking point column by calculating the mean value of the data points contained in each partition area, and then the tracking direction and tracking step size are determined by the neighborhood search. The specific steps are as follows:

Firstly, the initial graph data points are obtained, and the minimum and maximum values of X, Y, Z coordinates are obtained, so that a square area parallel to the coordinate axis surrounding all feature points can be formed. According to the number and distribution of the original feature points, the square area is further divided into small cube grids.

Then, the initial points of the tracking point column are determined and the local coordinate system is established to determine the stereo grid with the most data points Q_0. If the mean value of the data points in the grid is taken as the initial point of the tracking point column, the tracking points can be listed as $\{Q_0, Q_1, \ldots\ldots Q_k\}$. For the current tracking point Q_k, it is assumed that they are two known adjacent points in the tracking point column Q_{k-1}, Q_k, as shown in the following Fig. 2:

Take the vector

$$U = (Q_k - Q_{k-1}) / \| Q_k - Q_{k-1} \|$$

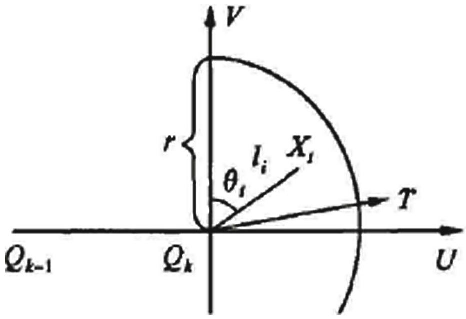

Fig. 2. Network structure

V is the vector *U* rotates in a counter clockwise direction, namely.

$$V = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} U$$

The local coordinate system is established with the tracking point Q_k as the origin and *V* and *U* as the coordinate axes.

Finally, the tracking step size and tracking direction are determined based on neighbor search. Assuming that $X = \{X_1, X_2, \ldots, X_n\}$ is the set of sampling points on the curve to be reconstructed, k nearest points are found in the stereo grid where the detection point X_i is located and among the $27(3 * 3 * 3)$ small three-dimensional grids, i.e. up, down, left and right, front and back, which are recorded as S. Take point Q_k as the center of the circle and take the distance r of the farthest adjacent point as the radius to make a circle. All data points X_i in the circle with an angle less than 90° with *U* are investigated. Suppose θ_i is the angle between the vector $Q_k X_i$ and *V*, and $\overline{\theta_k}$ is the angle between the tracking direction *T* and *V*, and the tracking step length $l_i^2 = \| X_i - Q_i \|^2$. In order to make the constructed graph curve reflect the correct shape, the objective function is constructed.

$$I_1 = \sum_{X_i \in S} \left(\theta_i - \overline{\theta_k}\right)^2 l_i^2$$

Find the I_1 corresponding to the minimum of $\overline{\theta_k}$, and then determine the tracking party *T*.

$$\overline{\theta_k} = \frac{\sum_{X_i \in S} l_i^2 \theta_i}{\sum_{X_i \in S} l_i^2}$$

After determining the tracking direction *T*, we need to find Q_{k-1} along this direction to make it located in the center of *S*.

Constructor is:

$$I_2 = \sum_{X_i \in S} \left[l_i^2 + \overline{l_k^2} - 2 l_i \overline{l_k} \cos\left(\theta_i - \overline{\theta_k}\right) \right] \cos\left(\theta_i - \overline{\theta_k}\right)$$

Find the $\overline{l_k}$ that makes I_2 the smallest:

$$\overline{l_k} = \sum_{X_i \in S} \left[l_i \cos^2\left(\theta_i - \overline{\theta_k}\right) \right] \Big/ \sum_{X_i \in S} \cos\left(\theta_i - \overline{\theta_k}\right)$$

Thus, a new tracking point is obtained:

$$Q_{k-1} = Q_k + \overline{l_k} \cdot T$$

The resampling of track feature points is completed.

Calculation of Track Deviation Degree Based on Maximum Distance. The system takes the digitized T_i of each given civil aviation route as the sample model, resamples the situation target track C to be matched, and then matches the target track C to be matched with all the established sample models T_i one by one, and calculates the distance d_i between the corresponding points of the graph C to be matched and the sample model T_i.

$$d_i = \frac{\sum_{k=1}^{N} \sqrt{(C[k]_x - T_i[k]_x)^2 + \left(C[k]_y - T_i[k]_y\right)^2}}{N}$$

When d_i is greater than a certain value set by the user, it means that the target obviously deviates from the center line of the route by a certain distance, or obviously does not fly along the established civil aviation route. It can be regarded as an aircraft with unknown intention. Using the maximum deviation distance, the unconventional changes of target track can be expressed intuitively and quickly.

Regularization of Recognition Results Combined with the Civil Aviation Flight Plan. For similar track shapes, there are great differences between the track points. We use seven track shape similarity indicators to measure from different angles, such as roundness, complexity, eccentricity, distance between center of mass and long axis, distance between center of mass and short axis, average distance between center of mass and distance between key points. Seven indexes of trajectory shape similarity are used to comprehensively evaluate the similarity between track shape and template track as the recognition result.

The integrity of the data obtained on the Internet is not necessarily the same, which also causes difficulties in the decision-making of trajectory shape similarity. However, the accuracy of the verification results can be greatly improved by combining with the flight plan. Using the information provided by the flight plan, we can know the target's estimated take-off time, flight number, aircraft type and take-off and landing airport, target updated estimated take-off time, flight number, aircraft type and take-off and landing airport information, accurate take-off time, flight number and transponder number of the target.Using this information, we can find out the correct target aircraft type according to the flight number Association and identify the target aircraft type, so as to achieve the purpose of aircraft identification.

Optimization Model Considering Data Noise. When building the model, it is assumed that the data is accurate, but the actual situation is not so, and the data noise needs to be considered. In the case of considering noise, the optimization of the model is mainly to solve the optimization problem by using the sequential minimum optimization algorithm, so as to provide effective parameter values for the aircraft identification model and obtain the optimal attribute classification recognizer.

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K\left(x_i, x_j\right)$$

$$s.t. \quad 0 \le \alpha_i \le C,\; i = 1, \cdots, n, \sum_{i=1}^{n}\alpha_i y_i = 0$$

Where n is the number of flying targets in the training set, a_i is the i-th Lagrange operator, x_i is the eigenvector of the i-th flying target, y_i is the identity attribute of the i-th flying target, C is a preset constant, $K\ (x_i,\ x_j)$ is a kernel function, which can be any one of linear kernel function, polynomial kernel function, Gaussian kernel function and sigmoid function. The Lagrange operator $a_i(i = 1,\ldots)$ can be obtained by using the sequential minimal optimization (SMO) algorithm to solve the optimization problem Then, the optimal classifier is obtained.

Algorithm Flow of Aircraft Identification Based on SVM Model. The aircraft identification model based on SVM model is actually a process of information fusion on the input and output nodes. The model transforms the fuzzy nonlinear relationship between feature parameters and related attributes into the mapping relationship in the model.The algorithm flow is as follows:

Step 1: according to the requirements of the actual system, determine the number of input target characteristic parameters x, and establish the output and input relationship of model nodes;

Step 2: before learning and training, it is necessary to extract and normalize the aircraft feature parameter data, in order to provide SVM with appropriate recognition input and training samples; According to the original and indirect characteristics of the flying target, the input parameters can be generated:

$$x_{ij} = \mu_{ij} + \sigma_{ij} \times randn$$

Among them, randn is a random number which obeys the standard normal distribution.

Step 3: select the appropriate SVM model, and train a certain number of training samples with one to many classification (OAA) and one to one classification (OAO) to get the expected SVM;

Step 4: test the test data with the SVM model. If the accuracy of the system is satisfied, the model will be used as the target identification classifier model. Otherwise, go back to the third step again until the accuracy of the system is satisfied.

The process of aircraft identification attribute recognition based on SVM is the process of computing using SVM. With the increase of feature parameters, when the parameters reach a certain degree, SVM can achieve higher classification ability.

4 Concluding Remarks

In this paper, using the amount of civil aviation information on the Internet, track matching recognition and flight plan recognition are combined to regularize the recognition results, and consider how to optimize the model in the case of data noise, so as to improve the recognition accuracy to a great extent.

References

1. Chen, S., Huang, Y.H., Huang, W.Q.: Big data analytics on avaiation social media. In: The Case of China Southern Airlines on Sina Weibo IEEE Second International Conference on Big Data Computing Service and Applications, pp.152–155 (2016)
2. Sangaiah, A.K., Gaol, F.L., Mishra, K.K.: Guest editorial special section on big data & analytics architecture. Intell. Autom. Soft Comput. **26**(3), 515–517 (2020)
3. Alhroob, A., Alzyadat, W., Imam, A.T., Jaradat, G.M.: The genetic algorithm and binary search technique in the program path coverage for improving software testing using big data. Intell. Autom. Soft Comput. **26**(4), 725–733 (2020)
4. Zhou, S., Wang, J., Jin Y.: Route planning for unmanned aircraft based on ant colony optimization and voronoi diagram. In: Second International Conference on Intelligent System Design and Engineering Application, pp. 732–735. IEEE (2012)
5. Yakovlev, K.S., Makarov, D.A., Baskin, E.S.: Automatic path planning for an unmanned drone with constrained flight dynamics. Sci. Tech. Inf. Process. **42**(5), 347–358 (2015). https://doi.org/10.3103/S0147688215050093
6. Li, J., Meng, X., Dai, X.: Collision-free scheduling of multi-bridge machining systems: a colored traveling salesman problem-based approach. IEEE/CAA J. Autom. Sinica 1–9 (2018)
7. Liu, G., Xu, H., Yang, S.: Multi-stage replica consistency algorithm based on activity. In: Shen, J., Chang, Y.-C., Su, Y.-S., Ogata, H. (eds.) IC3 2019. CCIS, vol. 1227, pp. 96–104. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-6113-9_11