

Airplane Recognition and Location On the Airport Based On Computer Vision

JianFeng Mu^{1, a}, YanYang Wang^{2, b}

¹ School of Aeronautic Science and Engineering of Beihang University, Beijing, 100191, China

² School of Aeronautic Science and Engineering of Beihang University, Beijing, 100191, China

^aemail: mjfbuaa@buaa.edu.cn, ^bemail: wangyanyang@buaa.edu.cn

Keywords: Airplane; Computer Vision; Recognition; Location; Security;

Abstract. In order to improve safety for airliner ground movement at the airport, a positioning method which can identify airliner and find the location of the airliner when the airliner taxiing at the airport based on computer vision is presented. Deep learning method is applied to recognize an airliner from complicated background of an image. After that, image segmentation based on graph is used to acquire the outline of the airliner. Then, after obtaining the nose undercarriage tyre, the location of airliner is obtained by means of coordinate transformation. The simulation results illustrate the validity of the proposed method.

1. Introduction

At present, world's air transportation industry is already booming. However, the lack of infrastructure of most airports, especially the dragging of field surveillance, seriously hindered the development of civil aviation. During the daily airline operations at airport, the case of collision between aircraft -aircraft, aircraft-vehicle occurs, which leads to not only significant safety risk, but also personal and property damage [1].

Using advanced technology to increase the ability of field monitoring and regulation is an essential measure to insure the normal operation, reduce the damage and property losses. There are some normal field surveillance means, such as field surveillance radar, automatic dependent surveillance – broadcast (ADS-B) [2][3], etc. For some small airports, tower controllers still use their eyes and voice as the main communication way. But these ways still have some drawbacks.

Field surveillance radar is expensive and the costs of installation and maintenance is very high, which prevent it from being widely used. As for ADS-B, operating costs are high, and technology incompatibility problems need to be solved, as well as hack problem [4]. In china, a lot of airplane and airport are not equipped with ADS-B.

From the beginning of the 21th century, computer vision has been used in scene surveillance, such as AVITRACK system [5], vision based Safe berth system and INTERVUSE system [6]. In this paper, we propose a method that use computer vision with deep learning model to recognize airplane in images. Then, under the application of image segment and structure analysis, we can find the nose undercarriage tyre. Finally, a method that measuring planar objects with a calibrated camera has been used to solve the position of an airplane on the airport.

2. Airplane Recognition in Images

The chosen method of airplane recognition is combining the deep architecture model from deep learning with sliding window. Adjusting the sliding window's size and step, the position of airplane can be precisely determined in the image.

Deep learning (DL) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using a deep graph with multi-layer neural network [7].

Artificial neural network is the beginning of deep learning and some of the most successful deep learning methods involve artificial neural networks. An artificial neural network is an interconnected group of nodes, akin to the complex network of neurons in a brain.

Artificial neural network(ANN) is a network connected by a series of artificial neurons that consist of three parts each: Input, Artificial Neuron and Output, as shown in Figure.1.

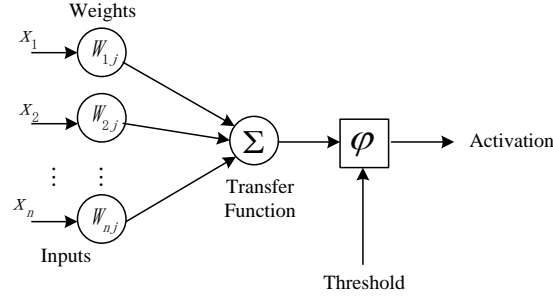


Figure.1 Artificial Neurons

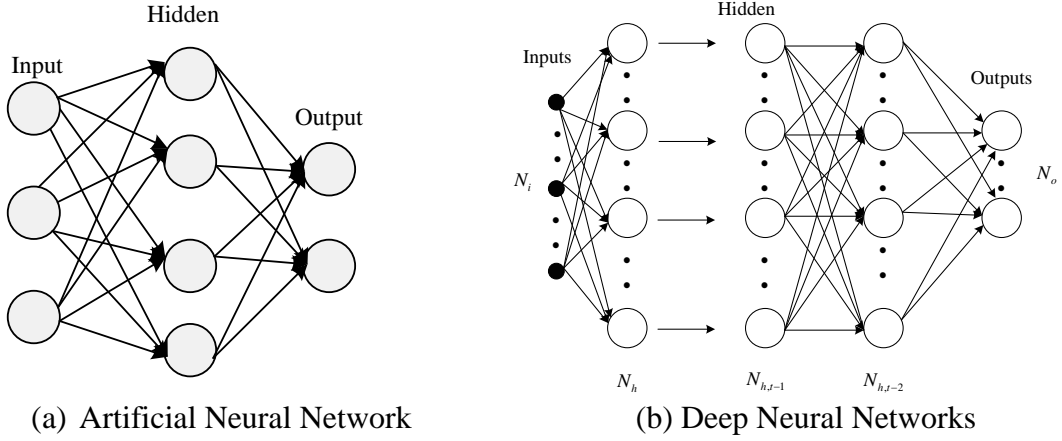


Figure.2 Artificial Neural Network

According to Figure 1, the Artificial Neurons can be mathematically stated as follows:

$$net = \sum_{i=1}^n w_{ij} \cdot x_i = W \cdot X \quad (1)$$

Where $X = (x_1, \dots, x_n)$, $W = (w_{1j}, \dots, w_{nj})$. Threshold φ can be defined, for example, as,

$$\varphi(X) = \begin{cases} 1 & \sum_{i=1}^n w_i x_i > 0 \\ -1 & \text{others} \end{cases} \quad (2)$$

ANN is composed by the joint of many artificial neurons. Figure. 2(a) shows a typical simple three layer ANN with 3 inputs and 2 outputs: Input, Hidden, Output. It is also a shallow network.

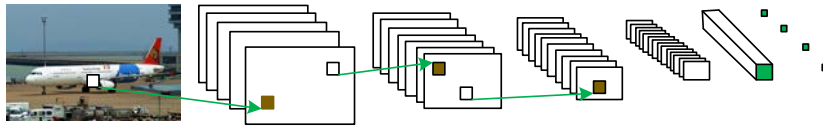
A deep neural network (DNN) is an artificial neural network (ANN) with multiple hidden layers of units between the input and output layers, such as Figure.2(b). The extra layers enable composition of features from lower layers, giving the potential of modeling complex data with fewer units than a similarly performing shallow network. In fact, deep learning contains two part: deep architecture and learning algorithms.

Deep learning presents mainly the features as followings.,

- (1) Learn features automatically from big data;
- (2) Combining feature selection and classifier ensemble
- (3) Deep learning can reuse the unit of the middle layer, reducing the parameters.
- (4) Deep learning is a kind of distributed learning method.
- (5) Extract information from the raw pixel data into the abstraction semantic layer by layer.

These features of deep learning make it is extremely suitable for image recognition and it has been widely used in image recognition filed.

Since 2006, when Geoffrey Hinton published his papers about deep learning [7,8], deep learning has been a great success in many areas. At the same time, there is a growing interest for deep learning methods used in image recognition and computer vision . In 2012, Hinton Group won the first prize in ILSVRC2012, making breakthroughs in image recognition and computer vision.



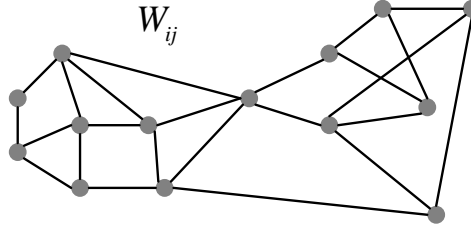


Figure.6 A drawing of a graph

A graph is an ordered pair comprising a set of nodes together with a set of edges. Figure.6 is a simple structure of a graph, which can be stated as

$$G = (E, V) \quad (3)$$

Where $V(G) = \{v_1, v_2, \dots, v_i, \dots, v_n\}$ stands for a set of nodes or vertices and let edge connection node i, j are indicated by $e_{ij} = \{v_i, v_j\}$, so $E(G) = \{e_1, e_2, \dots, e_i, \dots, e_m\}$ stands for a set of edges.

Some graph maybe has weight and it can present the pairwise correlations or similarity among variables. The weight between i, j can be indicated by

$$w_{ij} = w(v_i, v_j) \quad (4)$$

In the case of graph-based image segmentation, each node $v_i \in V$ corresponds to a pixel in the image, and the edges in E connect certain pairs of neighboring pixels. A weight is associated with each edge based on some property of the pixels that it connects, such as their pixels' dissimilarity. So, image segmentation problem can be translated into graph partition problem. Then, graph algorithms, like Minimal Spanning Tree, Shortest path, Max. graph flow, Min.cut, can be used.

Assume that each image pixel p_i has a corresponding vertex $v_i \in V$, we use an edge weight function based on the absolute intensity difference between the pixels connected by an edge,

$$w(v_i, v_j) = |I(p_i) - I(p_j)| \quad (5)$$

$I(p_i)$ is the intensity of the pixel p_i , such as its RGB.

Internal difference of a component $C \subseteq V$ can be defined as the largest weight in the minimum spanning tree of the component,

$$\text{Int}(C) = \max_{e \in \text{MST}(C, E)} w(e) \quad (6)$$

Difference between two components $C_1, C_2 \subseteq V$ can be defined as the minimum weight edge connecting the two components,

$$\text{Dif}(C_1, C_2) = \min_{\substack{v_i \in C_1, v_j \in C_2 \\ (v_i, v_j) \in E}} w(v_i, v_j) \quad (7)$$

Pairwise comparison predicate can be defined as

$$D(C_1, C_2) = \begin{cases} \text{true} & \text{Dif}(C_1, C_2) > \text{Mint}(C_1, C_2) \\ \text{false} & \text{otherwise} \end{cases} \quad (8)$$

Where Mint is the minimum internal difference and is defined as,

$$\text{Mint}(C_1, C_2) = \min(\text{Int}(C_1) + \tau(C_1), \text{Int}(C_2) + \tau(C_2)) \quad (9)$$

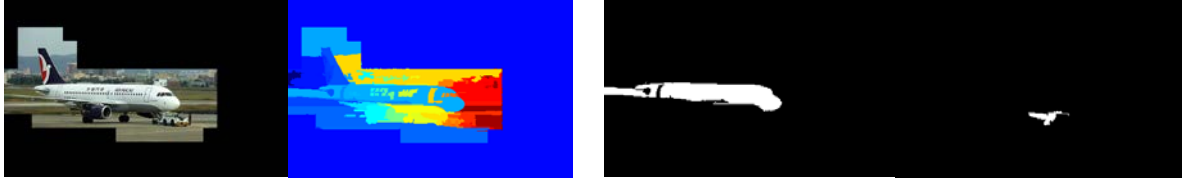
$\tau(C)$ is a threshold function based on the size of the component C and $\tau(C) = K/|C|$, where $|C|$ stands the size of C and K is given manually. Segmentation algorithm can be described as,

- (1) Compute the dissimilarity between pixel p_i and its 8 neighbors.
- (2) Sort E by non-decreasing edge weight.
- (3) If $\text{Id}(v_i) \neq \text{Id}(v_j)$ and $w_{ij} < \text{Min}(C_i, C_j)$, then merge the two components otherwise do step 4
- (4) Set both $\text{Id}(v_i), \text{Id}(v_j)$ into $\text{Id}(v_i)$ and $w_{ij} = w_{ij} + K/(|C_i| + |C_j|)$
- (5) If $n < N$, repeat step (4), otherwise end.

Figure.7 (a) shows the result of graph-based image segmentation with an airplane.

After image segmentation, we can use deep architecture model from deep learning again for each

component, then we can find airframe, as shown in Figure.7 (b) left, and Gear. Because we have find the tail, we can easily get the nose gear as shown in Figure. 7 (b) right.

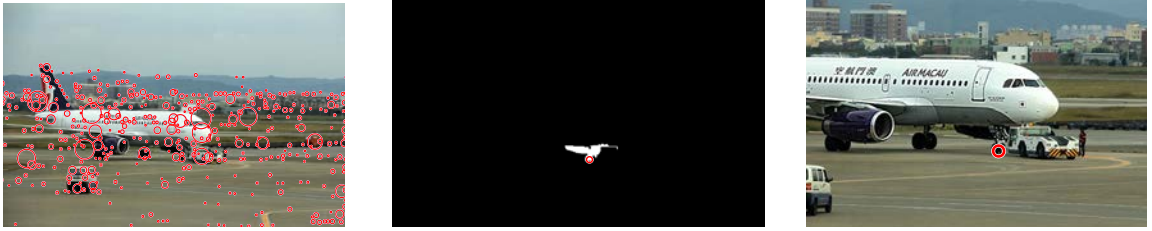


(a) Result of graph-based image segmentation

(b) Airframe and Nose Gear

Figure.7 Results of airplane structure analysis

Through the analysis of the structure of airplane, we can find the nose gear. But we need to find the point closet to the ground. So, we have to find the Nose Undercarriage tyre. In this paper, The HOUGH transform circle detection method has been used to find the candidate area of tyre [10], as shown in Figure.8 (a). The condition that Circle must contain the lowest point of nose gear and the proportion of overlap is the largest should be met to determine the circle of tyre..



(a) Candidate area of tyre

(b) Circle of Nose Undercarriage tyre

Figure.8 Results of Nose Undercarriage tyre

Figure.8 (a) and Figure.8 (b) show the circle we found which contains the nose undercarriage tyre. Because the method of localization need the point closet to the ground, assume that the center of the circle is (x_c, y_c) , and the radius is r , the feature points is

$$C = (x_c + r, y_c) \quad (10)$$

4. Airplane Location

In this paper, Zhang Zhengyou's plane calibration method of camera calibration algorithm has been used [11], through which we can obtain the transfer matrix we need.

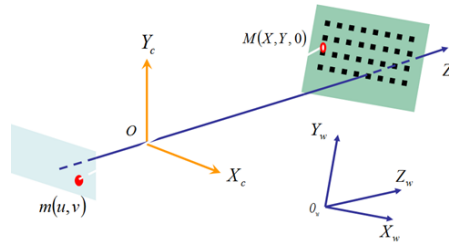


Figure.9 Camera Projection Model

Suppose that $m = (u, v)$ is a 2D point and $M = (X, Y, Z)$ is a 3D point, Figure 9 [11]. For the convenience of calculations, we use \tilde{x} to denote the augmented vector: $\tilde{m} = (u, v, 1)$ and $\tilde{M} = (X, Y, Z, 1)$. A camera is modeled by the usual pinhole as shown in Figure.9 and the transfer function is given by

$$s\tilde{m} = A \begin{bmatrix} R & t \end{bmatrix} \tilde{M} \quad (11)$$

where s is an arbitrary scale factor. (R, t) is the extrinsic parameters and A is camera intrinsic matrix which is given by

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

With (u_0, v_0) the coordinates of the principal point, α and β the scale factor in image u and v axes,

and γ the parameter describing the skewness of an image. Set $Z=0$, we get

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & r_3 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (13)$$

Where H is a holography matrix which illustrates the mapping relationship between the two chess board. So

$$\widetilde{sm} = H\widetilde{M} \quad (14)$$

Using the knowledge that r_1 and r_2 are orthonormal, we have

$$h_1^T A^{-T} A^{-1} h_2 = 0, \quad h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2 \quad (15)$$

There are 5 parameters we need to know but there are only 2 equations. we could use multi-image to solve these parameters of A . Thus

$$r_1 = \lambda A^{-1} h_1, \quad r_2 = \lambda A^{-1} h_2, \quad r_3 = r_1 \times r_2, \quad t = \lambda A^{-1} h_3 \quad (16)$$

Where $\lambda = 1/\|A^{-1}\|$.

If n images are used and there are m points on each model plane, maximum likelihood estimation is got by minimizing the following function,

$$\sum_{i=1}^n \sum_{j=1}^m \|m_{ij} - \hat{m}(A, R_i, t_i, M_j)\|^2 \quad (17)$$

Where $\hat{m}(A, R_i, t_i, M_j)$ is the projection of point M_j in image i .

A desktop camera unusually exhibits significant lens distortion, especially radial distortion. Let (\tilde{u}, \tilde{v}) be the real normalized image coordinates. We have

$$\begin{bmatrix} (u - u_0)(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2) \\ (v - v_0)(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2) \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \tilde{u} - u \\ \tilde{v} - v \end{bmatrix} \quad (18)$$

in matrix form: $DK=d$, and the solution is given by

$$k = (D^T D)^{-1} D^T d \quad (19)$$

Thus, we can get

$$\begin{bmatrix} X & Y & 1 \end{bmatrix}^T = (A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix})^{-1} \cdot s \begin{bmatrix} u & v & 1 \end{bmatrix}^T \quad (20)$$

Let (x_r, y_r) be the coordinates in the ground coordinate system, so

$$\begin{aligned} x_r &= X + x_e \\ y_r &= Y + y_e \end{aligned} \quad (21)$$

Because, at the airport, the airplane's landing gear is always on the ground, the undercarriage tyre is used to pinpoint airplane. In this paper, we use iPhone5s, C919 model (1:100) and the Guangzhou Baiyun Airport model (1:40) to test this method, as shown in Figure.10.

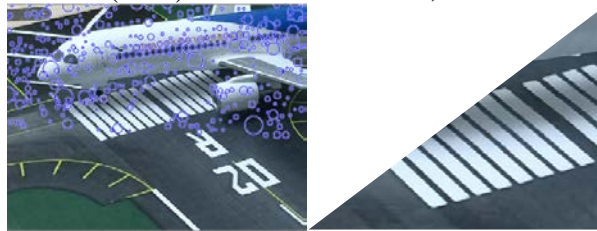


Figure.10 Undercarriage Tyre of C919

Intrinsic and extrinsic parameter matrix are solved through camera calibration,

$$A = \begin{bmatrix} 2834.6 & 1310.9 & 1595.1 \\ 0 & 2805.1 & 1310.9 \\ 0 & 0 & 1 \end{bmatrix}, R = \begin{bmatrix} -0.5560 & -0.4258 & 0.7139 \\ 0.8292 & -0.2247 & 0.5118 \\ -0.0575 & 0.8765 & 0.4700 \end{bmatrix}, t = \begin{bmatrix} 66.7690 \\ 96.8692 \\ 843.2909 \end{bmatrix}$$

The calculated results of the position of C919 are shown in table.1.

Table. 1 Airplane location and errors

No	Gauge value/cm	Actual values/cm	absolute error/cm	Relative error
1	(80. 7716, 37. 7113)	(83. 5000, 39. 0000)	(-2.7282, -1.2886)	(3.26%,3.30%)
2	(65. 0606, 28. 7720)	(67. 5000, 28. 0000)	(-2.4393, +0.7720)	(3.61%,2.76%)
3	(79. 5203, 30. 7839)	(81. 5000, 30. 0000)	(-1.9797, +0.7839)	(2.40%,2.61%)

5. Conclusion

By combining the deep architecture model from deep learning and the method of sliding window, we recognized the airplane from complex background from the image. Then, through graph-based image segmentation and circle detection, the nose undercarriage tyre is found to locate the airplane. Finally, airplane position has been determined by using transfer matrix which computed through camera calibration. The results show that the proposed method based on computer vision is effective. Due to its low cost and easy to use, the proposed method will be widely used in airport surveillance.

References

- [1] Li Min, Wang Bang Feng, Ding Meng Application of ADS-B for Airport Surface Surveillance[J]. Journal of Civil Aviation Flight University of China.Jan.2014, 1(25) 11-14.
- [2] Kakubari Y, Koga T, Miyazaki H, et al. Enhancement of Passive Surveillance System for Airport Surface Movement[J]. Electronics and Communications in Japan, 2014, 97(3): 24-30
- [3] Powell J D, Jennings C, Holforty W. Use of ADS-B and perspective displays to enhance airport capacity[C].24th Digital Avionics Systems Conference. IEEE, 2005, 1: 4. D. 4-4.1-9 Vol. 1
- [4] McCallie D, Butts J, Mills R. Security analysis of the ADS-B implementation in the next generation air transportation system[J]. International Journal of Critical Infrastructure Protection, 2011, 4(2): 78-87
- [5] Thirde D, Borg M, Ferryman J. A real-time scene understanding system for airport apron monitoring[C]. Proc 4th IEEE Int. Conf. on Computer Vision Systems, 2006.
- [6] Pavlidou N, Grammalidis N, Dimitropoulos K, and et al. Using Intelligent Digital Cameras to Monitor Aerodrome Surface Traffic[J]. IEEE Intelligent Systems. 2005, 20(3):76-81
- [7] Li Deng, Dong Yu. Deep Learning: Methods and Applications, First ed. Now, Beijing,2013.
- [8] Hinton, G. E., Osindero, S. and Teh, Y. A fast learning algorithm for deep belief nets. Neural Computation,18(2006): 1527-1554
- [9] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation [J]. International Journal of Computer Vision, 2004, 59(2): 167-181.
- [10] Xu L, Oja E, Kultanen P. A new curve detection method: randomized Hough transform (RHT)[J]. Pattern recognition letters, 1990, 11(5): 331-338
- [11] Zhang Z. A flexible new technique for camera calibration[J]. IEEE Transactions on pattern analysis and machine intelligence, 2000, 22(11): 1330-1334