A distributed photoelectric sensing and simulation training method for the UAV swarm

Rang Liu¹, Jian Wu¹, Yanjie Zhao¹*, Changqing Liu¹

¹ China Academy of Electronics and Information Technology, Beijing, 100041, China

*Corresponding author’s e-mail: zhaoyanjie1@cetc.com.cn

Abstract—Aiming at the shortage of the training samples in the process of target recognition of the UAV swarm distributed photoelectric sensing system, the distributed photoelectric sensing and training technology based on the loop simulation is studied. This paper discussed the hardware architecture and software architecture of distributed photoelectric sensing and training system, as well as the design of distributed video splicing and target recognition algorithm. Finally, a simple distributed photoelectric sensing hardware in the loop simulation training system applied to UAV cluster was built in the laboratory, and the feasibility of this technology was verified by experiments.

1. Introduction

UAV has become an important force of the battlefield intelligence because of its small shape, strong penetration ability and no life cost. A single UAV usually uses a photoelectric payload to collect the image of a fixed field of view, whose field of view is relatively small. In recent years, the UAV swarm is coming to the field of vision, which can carry out the wide area reconnaissance through the distributed payloads. Sometimes, in order to avoid the collision, a single UAV of the swarm can carry the distributed photoelectric payload for all-round visual environment perception. Omnidirectional perception refers to the superposition of multiple videos and images with overlapping areas as reference to form a 360 panoramic image with greater information and more intuitive display effect. It adopts global brightness balance optimization to form an image more suitable for human eyes, and realizes the detection and recognition of specific targets based on the panoramic image. Through the research on omni-directional sensing technology, we can realize the comprehensive monitoring of a wide range of scenes and the intelligent identification of high-value targets, and greatly improve the efficiency and reliability of reconnaissance and surveillance [1-3].

The panoramic image has a large amount of data, including target and background scene, and the traditional target recognition algorithm is difficult to meet the timeliness requirements. As a new data processing technology in recent years, artificial intelligence technology, on the one hand, has the ability of feature extraction and data fitting unmatched by traditional methods in dealing with large amount of data and complex structure data, which can effectively solve the above problems; On the other hand, artificial intelligence technology has the generalization ability to simulate the learning of human brain neurons[4]. Before 2012, target detection mainly used artificially designed features to train shallow classifiers to complete the classification task. In 2012, the image classification algorithm based on deep convolution neural network based on deep learning theory is proposed, which greatly improved the accuracy of image classification. Since 2014, great breakthrough has been made in target detection. A large number of target detection algorithms based on convolutional neural networks and candidate regions have emerged, such as R-CNN, SPP-NET, Faster R-CNN and so on[5,6], which
have been greatly improved in speed and accuracy. Although the Faster R-CNN has made great progress in computing speed, it can hardly meet the requirements of real-time detection. Therefore, some people propose a force based regression method to directly return the position and type of the target object from the picture. The two representative methods are Yolo and SSD.

The development of target detection algorithm based on deep learning is very rapid, and the detection accuracy is becoming higher and higher. However, there are often few data sets of military targets, and the problem of the insufficient samples makes the training model inaccurate, which brings severe challenges to the target detection.

2. System architecture design
Aiming at the training problem of UAV swarm omni-directional photoelectric sensing system, this paper designed a set of hardware in the loop simulation sensing training system to simulate and realize the distributed photoelectric payload sensing and identification simulation training of the UAV swarm. The system is mainly divided into target scene simulation platform, projection equipment, photoelectric payload and perception processing module, training platform, effect display platform, etc. The system scheme is shown in the figure below. The projection equipment adopts high-definition laser engineering projector, which is used to play images or videos with specific targets during function demonstration. The effect display platform mainly displays the target image results of the UAV swarm photoelectric perception and recognition.

The photoelectric sensing and simulation training system is based on the private container cloud distributed architecture, which fully considered the decoupling of algorithm and computing power, and provided open capabilities at all levels of infrastructure, data, algorithm and intelligent application. For example, the signal data analysis application based on deep learning continuously accumulates the sample data of air targets, iteratively trains and upgrades the algorithm model, and continuously improves the indicators of the algorithm model. The training system provides the computational support for the training tasks in the above application scenarios, and provides a powerful parallel execution environment for algorithm model reasoning execution.

In this scheme, multi-channel video images are panoramically stitched and the target categories specified in the image are recognized. The design idea can be divided into single image recognition and then image stitching, or stitching to get the stitched image, and then the whole stitched image for target recognition. When the target is small, the proportion of the target in the spliced image is too low, which will easily lead to target detection and recognition. Therefore, the method of identifying a single
image first and then stitching the image has better recognition effect, but the time will increase a lot which does not meet the real-time performance. The method of which stitching the image first, and then the whole stitched image for target recognition can design a better target recognition algorithm to improve the difficulty of small target detection and recognition. This method meets the requirements of real-time. Therefore, in this project, the stitched image is obtained first, and then the target detection and recognition are carried out for the stitched image.

3. Software algorithm design

3.1. Overall software architecture
The software logical architecture level of perception training system is shown in the figure below.

![Fig.2 The software logical architecture level of perception training system](image)

The logic level of the software logical architecture level of perception training system is divided into three layers: physical layer, foundation layer and application layer.

The physical layer provides the distributed optoelectronic payload hardware environment, including embedded artificial intelligence board, Gigabit network switching module, video acquisition terminal and other functional hardware modules, which is the basic platform for software operation.

The foundation layer provides the distributed optoelectronic payload software running environment and basic services, including embedded operating system, file management, neural network model, deep learning acceleration library, basic library and hardware driver dependent on application program.

The scheme uses embedded Linux as the operating system, and the system supports network communication and deep learning reasoning acceleration library. The deep learning acceleration library provides software calls in the form of software API to obtain the acceleration of application software deep learning algorithm. The embedded operating system, file system and other basic software libraries mainly play the role of the operating system, mainly provide the running environment for the application software, and be responsible for the resource scheduling of the hardware system. The bottom driver software is mainly hardware driver and network driver, so that the application software can interact with the outside world. The deep learning acceleration library is mainly used to support the special acceleration chip of neural network to accelerate the forward reasoning of deep learning, and is called by the main control software in the form of standard API.

The application layer mainly includes the scene simulation software, the video splicing software and the target detection and recognition software to realize the simulation of the target optical dynamic characteristics and the background characteristics in typical land, sea and air visible light simulation scene, as well as the functions of real-time image data acquisition, splicing, target detection and recognition, image data transmission and so on.

3.2. Video splicing algorithm
The video splicing software is responsible for splicing the video stream of the connected multi-channel network camera. The video splicing software is responsible for decoding and encoding the network
video stream, splicing multiple videos according to the splicing algorithm, and scaling and cutting the spliced images for output. The video splicing software includes video codec function and video panoramic splicing module. The video splicing module is shown in the figure below.

In order to realize the video transmission and processing, the video codec module needs to pull and decode the network video stream generated by the camera, transmit it to the video splicing module for splicing, and output the video through the target recognition module for encoding and streaming.

The video splicing module mainly realizes the splicing of multi-channel 1080p video, and then divides and displays the panoramic video according to the demand, including LUT table initialization, splicing, fusion, uniform brightness, image scaling and cutting.

LUT table is used to splice multi-channel video. The accuracy of LUT table determines the quality of splicing. In order to obtain high-precision LUT table, registration algorithm is used to obtain the internal and external parameters of each camera for splicing. The internal parameters of the camera can be obtained according to the provided camera focal length and photosensitive element size. The camera can be registered according to the image of the camera at the same time. The external parameters of the camera can be obtained from the matching points, and then the LUT table required for splicing can be generated. The flow chart is shown in the following figure.

![Fig.3 Initialize LUT algorithm](image)

FAST characteristic point extraction algorithm is a simple and fast feature extraction algorithm, which is widely used because of its fast operation speed and good feature extraction effect. However, for some images with different local contrast due to uneven illumination or shadow, the adaptability and feature extraction effect of fast algorithm are not ideal. BREF algorithm uses binary coding to obtain descriptors. It only compares simple gray values without involving a large number of complex calculations, whose operation speed is very fast. However, because it is based on the specific pixel value of the image, the algorithm is very sensitive to the noise, which cannot meet the rotation invariance and scale invariance. The ORB (Oriented FAST and Rotated BRIEF) characteristic operator is used in our paper to find the characteristic points. The ORB characteristic detection algorithm combines FAST characteristic point detection method with BRIEF characteristic descriptor, and improves and optimizes their shortcomings.

For the two images, after using the ORB algorithm, the detected characteristic point descriptors are generated. Assuming that \( \alpha_1 \) and \( \alpha_2 \) are descriptors of the characteristic point of the two images respectively, the image matching is carried out by calculating the Hamming distance \( D \):

\[
D(\alpha_1, \alpha_2) = \alpha_1 \oplus \alpha_2
\]

The smaller the value of \( D \), the higher the similarity between the two characteristic points.

When the scale difference is small, it has a good effect on characteristic matching. In video mosaic, the camera image scale is similar, and the characteristic point extraction and matching with good real-time effect can be realized by ORB. FLANN matching algorithm is used for characteristic point matching. FLANN matching algorithm can search the nearest neighbor of large data sets and high-dimensional features. Combined with better characteristic description in orb, it can obtain as few false matching points as possible. The RANSAC algorithm is used to get the external parameters, and the wrong matching points are removed to obtain higher precision registration.
3.3. Target detection and recognition algorithm

The target detection and recognition software is responsible for detecting and recognizing the target in the image and obtaining the target structured information, including target location box, target category and target confidence. The incoming image information is the original real-time image collected by multi-channel cameras and the spliced image.

For the small target and multi-scale target detection and recognition, it is necessary to comprehensively consider two factors: network structure design and training mode design. In the design of network structure, based on the traditional feature extraction backbone, more feature fusion is needed to comprehensively judge the location and category of targets by combining shallow features and deep features. In the design of training method, the processing of training data and the design of training trick can effectively improve the accuracy of small target detection and recognition of parameter model on the premise of the same network structure.

The mainstream method of deep learning algorithm in the field of target detection and recognition is the one stage series of target detection algorithms, such as Yolo. This paper draws lessons from the overall structure design of Yolo, which uses a lightweight backbone to strengthen the running speed of embedded board, and adds a stronger feature fusion network structure design. The structure design of the target detection network mainly has two characteristics. One is to introduce a feature fusion mechanism with stronger feature extraction ability to ensure the full utilization of the shallow and deep features of the image. The network structure has strong adaptability to small targets and multi-scale changes in the scene, and the accuracy in the scene is higher than that of the open source algorithm. Secondly, lightweight backbone selection ensures the effectiveness of feature extraction while fast down sampling. The network has the characteristics of short forward computing time and is suitable for embedded platform computing.

4. Experiments

The target detection and recognition technology based on hardware in the loop simulation training mainly includes two parts: algorithm training and algorithm reasoning. The first part is the algorithm model training. The target types identified are detected according to the needs, the relevant data sets are collected, and the target scene simulation is carried out at the same time. The collected data is projected through the projector, and the front-end camera video image is collected in real time through the 100Mbps network interface. After the input multi-channel video is decoded in real time, the image is spliced. In the process of stitching, high-resolution stitched images with large field of view angle will be generated. Frame the video to get the target detection and recognition data samples. Then, the spliced video is extracted to obtain target recognition data samples. After obtaining the data set, the data set needs to be sorted and labeled, and the network needs to be repeatedly trained, optimized and trained through the deep learning training platform. The second part is the algorithm reasoning. In this step, the trained neural network model is used to detect the target in the data, and the structured data is output for further analysis.

To sum up, after the original splicing data is calibrated and expanded, the training model is obtained through optimization training. The real-time spliced video image data is detected and recognized by loading the trained model. Finally, the information about the target in the video image is obtained, and the structured data is output for further analysis.

A photoelectric sensing simulation training principle system applied to UAV swarm is built in the laboratory. In this system, four distributed photoelectric cameras with brackets are used to simulate the distributed photoelectric payloads carried by four UAVs in the swarm. The target video to be perceived and recognized is projected by the projector for capture by the distributed photoelectric payload. The embedded board based on the deep learning is used to recognize the target data captured by the distributed photoelectric payload in real time. The target recognition effect of photoelectric perception simulation training principle system is shown in the figure below. Figure (a) shows the built hardware system, and figure (b) shows the results of stitching and recognition of four captured video images. The results show that the embedded target recognition algorithm successfully
recognizes the aircraft target projected by the projector. At the same time, 4-channel video data splicing effect also worked perfectly.

![Effect display of simulation training system](image1)

![Effect display of simulation training system](image2)

Fig.4 Effect display of simulation training system

5. Conclusion

The UAV swarm can realize wide area reconnaissance with a larger field of view in the target area by distributed with photoelectric payloads. Aiming at the UAV swarm lack of effective target sample images during perceptual training, the distributed photoelectric sensing and training technology based on the loop simulation is studied. This paper discussed the hardware architecture and software architecture of distributed photoelectric sensing and training system. Finally, a simple distributed photoelectric sensing hardware in the loop simulation training system applied to UAV cluster was built in the laboratory. Later, with the deepening of research, the architecture and constituent units of the hardware in the loop simulation training system will be improved.

References

[1] Berenguer, Y., Payá, L., Ballesta, M., Reinoso, O. (2015) Position Estimation and Local Mapping Using Omnidirectional Images and Global Appearance Descriptors. Sensors, 15:26368-26395.

[2] Valiente, D., Gil, A., Fernández, L., Reinoso, O. (2014) A comparison of EKF and SGD applied to a view-based SLAM approach with omnidirectional images. Robot. Auton. Syst., 62: 108–119.

[3] Maohai, L., Han, W., Lining, S., Zesu, C. (2013) Robust omnidirectional mobile robot topological navigation system using omnidirectional vision. Eng. Appl. Artif. Intell. 26:1942–1952.

[4] Erhan, D., Szegedy, C., Toshev, A., Anguelov, D. (2014) Scalable object detection using deep neural networks. In CVPR.

[5] Girshick, R. (2015) Fast R-CNN. arXiv:1504.08083.

[6] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR.