Design and realization of agricultural intelligent inspection robot

Chen Zhichang[1,2], Gao Wanlin[1,2*]

(1. *College of Information and Electrical Engineering, China Agriculture University, Beijing 100083, China*;
2. *Key Laboratory of Agricultural Informationization Standardization, Ministry of Agriculture, Beijing 100083, China*)

Abstract: In view of the agricultural production process, many scenes require human monitoring, an agricultural intelligent inspection robot was put forward. The robot can detect and track pedestrians around the production site and transfer real-time monitoring information to the host computer for storage. Robot consists of three parts: the hardware platform, software systems and intelligent algorithms. They help the robot to achieve common intelligent inspection tasks. The hardware platform is mainly composed of sensors, actuators, mechanical structures and electronic systems. Like human body, the hardware platform configuration is the basis of robot design. The software system is mainly composed of ROS system, like human brain to deal with coordination of various kinds of information. ROS as an open source system, has multi-platform support, multi-language support, toolkit and other advantages. Intelligent algorithm is the robot's "soul", from the intelligent algorithm, we can see the robot targeted feedback to the operator. In this paper, the robot obtained the color image, then the pedestrian detection was carried out by combining the channel characteristic and the faster R-CNN algorithm, and the pedestrian was distinguished from the background and positioned accurately. Then, the pedestrian location information was detected to be transmitted to the depth channel Boosting pedestrian tracking algorithm to track pedestrians. Meanwhile, robot sent back the monitored pedestrian information to the master computer to save through WiFi.

Keywords: robot, pedestrian detection, pedestrian tracking, ROS composed

Citation: Chen, Z. C., and W. L. Gao. 2017. Design and realization of agricultural intelligent inspection robot. International Agricultural Engineering Journal, 26(3): 320–326.

1 Introduction

With the continuous development of computer science and technology, people's way of life has undergone enormous changes, more and more tasks that have been done by people who spend a lot of manpower replaced by the computers in nowadays. With the rapid development of robot technology, artificial intelligence and robot combination of intelligent robots in more and more areas have very good applications.

In the agricultural sector, there are many aspects in agricultural production and processing of agricultural products need human to handle. Such as some food quality requirements of the strict cultivation of agricultural products need to be monitored to prevent agricultural products from damaging and ensure the quality and safety; keeping the idlers away from the important experimental fields to prevent damages to the tests; seed bank and pesticide library should be protected to prevent theft; keeping idlers out of the flour mills to prevent the occurrence of danger etc. Briefly speaking, using pedestrian tracking detection technology and robot technology, designing agricultural intelligent inspection robots can replace manpower in the above-mentioned areas, and will have better results and benefits. It is the direction of future development.

Received date: 2017-06-27 **Accepted date:** 2017-08-17

* **Corresponding author: Gao Wanlin,** PhD, Professor, Research interests: agricultural informationization and application specific chip. Address: College of Information and Electrical Engineering, China Agriculture University, Beijing 100083 P. R. China. Tel: +86-010-62736755, Email: cau_szmtyjs@163.com.

2 Materials and method

2.1 Pedestrian detection

Pedestrian detection is the image or video appear in the pedestrian from the background area and precision positioning. From the time of the issue put forward to the present, pedestrian detection, both detection accuracy and detection speed have been greatly improved, especially in

the past ten years. Machine learning-based pedestrian detection improved the detection performance to a very high level. However, in solving practical problems, the current study is subject to the influence of the environment and pedestrians, and in some cases, cannot achieve the desired results.

Pedestrian detection, from the method, can be divided into the moving object detection method and feature matching method. In the early study of pedestrian detection, the method is mainly based on moving object detection. There are background elimination method (Jain, 1979), frame difference method (Meier and Nagank, 1998) and optical flow method (Lipton et al., 1998). But these methods have great limitations, cannot detect the object in the static image or detect an object that does not move in the video

In recent years, the pedestrian detection algorithm has been developed rapidly. In this paper, the pedestrian detection algorithm, based on machine learning, is used to extract the pedestrian characteristics, and the classification model is constructed according to the characteristics. Then, the pedestrian image data are trained and identified. The At present, the successful features of Haar (Papageorgiou et al., 1998) wavelet feature, LBP (Ojala et al., 1996), HOG (Dalal and Triggs, 2005), and edgelet (Bo and Nevatia R, 2005) are used in pedestrian detection. The general detection framework is to extract features and SVM, adaboost Combine.

Based on deep learning, from the beginning of 2012, pedestrian detection research gradually entered into the people's vision, a variety of pedestrian detections for the deep learning methods and deep learning network structures have been put forward, and it greatly improved the effect of pedestrian detection. Especially in the past three years, the deep learning in the object detection has made great progress, led the pedestrian detection performance and speeded up its upgrading. The great development of object detection was the beginning of the R-CNN (Girshick et al., 2014) method for object detection and segmentation published by Girshick et al. in 2014, followed by a large number of the same series of algorithms, this pedestrian detection method has achieved very good results, and this paper also used this method and channel characteristics to achieve pedestrian detection.

2.2 Object tracking

The goal of the object tracking is to estimate the position and motion parameters of the object after the initial position of the object in the first frame of the image and it is given in the image sequence. The typical object tracking system consists of three parts: object visual model, dynamic model and search strategy (Wang et al., 2011). The objective model of the goal is the most important part of the object tracking, which determines which objective function is used to search for the object of interest in the frame. The dynamic model is pre-defined or obtained through training data learning in order to reduce the object search space and calculate the load. The common dynamic model method has linear regression (Seber and Lee, 2003), particle filter (Ali et al., 2014), Kalman filter (Chang and Ansari, 2005), Mean shift (Comaniciu et al., 2000), the hidden Markov model (Chen et al., 2003) and so on. Finally, the apparent model and the dynamic model are combined, and then the optimal search strategy is used to obtain the most possible position of the object, and the positioning of the tracking object is realized.

In order to deal with the change of the object appearance, the object tracking algorithm in recent years mostly uses the adaptive representation model based on the generated or discriminant formula. The generation method is based on the objective detection, the background of the object after the modeling, according to a certain tracking strategy to track the optimal location of the object tracking. The tracking process and the detection process are independent, and the object is tested and then tracked. The discriminant tracking method is to obtain the object state by the object detection of each frame image, and to see the classification problem of the object and the background. By training the classifier, the object region is selected in the current frame and the background is the largest, detection and tracking will be done at the same time. The discriminant method has better tracking performance than the generating method, and the method of using the incremental learning to distinguish the object and the background is less costly and the tracking effect is better.

2.3 Robot

The robot is a programmable and versatile machine, or a specialized system with computer changes and programmable actions for performing different tasks. It is generally composed by the implementing agencies, drive devices, detection devices and control systems, complex machinery and other components. The future direction of the robot is to research the intelligent systems. At present, the mainstream robot operating systems on the market are ROS, Android and Ubuntu, providing some methods for writing robot navigation and limb procedures. These methods do not need to consider the specific hardware, and they even provide a higher level of command. When the robot operating system started on the robot's computer, it will get robot attribute data, such as the robot's limb length and motion data, and then pass the data to a higher-level algorithm. The data are then processed by these algorithms and returned to the motion instructions of the robot.

3 Hardware platform construction

Intelligent robot mainly consists of two parts: the hardware platform and software system. The "hardware platform" in the intelligent robot component refers to a robot entity, like a human body. Hardware platform mainly involves two major areas of mechanical and electronic technologies, including sensors, actuators, mechanical structure, electronic systems and other components. The hardware platform is the basis of the software system. It provides real-time feedback information to software and receives system information to complete the corresponding action.

3.1 Agricultural intelligent inspection robot hardware platform

The components of intelligent robot hardware platform are mainly four categories: sensors, actuators, mechanical structure and electronic systems. The configuration and performance of hardware platform are directly related to the possibility of a robot application, therefore, a robot project to carry out the first step is based on the project objectives, the appropriate robot hardware platform should be selected. This paper started with the configuration of sensors and actuators in hardware, and the hardware platform of agricultural intelligent inspection robots was designed.

3.2 Sensor configuration

The sensor is a sensing device of the robot, which is the main way for the robot to get the information of the external environment. Whether the sensor type is rich in configuration and the performance or not, its accuracy and reliability are directly related to the appropriate environment data the robot obtained and also the data processing the robot handled.

Sensors, such as accelerometer, angular speedometer, encoder, pressure, temperature, displacement, ultrasonic (infrared) ranging, visible light, microphone and so on, can be obtained from the environment, and detection of information can be regarded as sensors. A variety of sensors of microelectronics miniaturization, integration, high precision, high reliability, high stability and other direction are endless new sensors, appeared in recent years.

Intelligent robot sensor configuration should meet the robot's intelligent needs, and use intelligent technology data sources. The sensor configuration of the agricultural intelligent inspection robot mainly includes the following categories: RGB-D sensor, wide-angle HD RGB sensor, laser radar sensor and infrared distance measurement sensor.

RGB-D sensors are visual sensors that are widely used as somatosensory game devices and as research tools for computer vision. It is mainly used for collecting depth information of near depth images, and is used for robot obstacle avoidance and following tasks. Wide angle HD RGB sensor is used to collect high-definition images, and these high-definition images are used to achieve long-distance pedestrian tracking detection, thus to expand the scope of robot surveillance. Laser radar sensors are used for real-time map establishment and robot position information acquisition, and it is an essential tool to realize path navigation in the process of robot patrolling. The infrared ranging sensor is used to acquire the distance information of the near obstacle, and is used for obstacle detection of moving robot.

3.3 Actuator configuration

The implementing agency has the ability of the

robot to act and react physically to the outside world. Such as robots and mobile robotic chassis drive mechanism are the robot's actuator. The performance indicators of the implementing agency are directly related to whether the expected directive can be carried out smoothly.

In this paper, the agricultural intelligent inspection robot mainly means the mobile chassis, does not involve the implementation of mechanical arm and other institutions. The mobile chassis is not a pure actuator, but a complete robot system that contains sensors, mechanical structures and electronic systems, used as an actuator. The main body of the mobile chassis has a three-wheeled structure, and the first two driving wheels. The last one as a driven wheel is to ensure the accuracy of its turning angle. The response of the mobile chassis to the instruction is guaranteed by the system of the chassis itself, and the mobile chassis can also provide the position and attitude information to the system through its own sensor.

3.4 Hardware platform architecture

The intelligent platform design of the agricultural intelligent inspection robot mainly includes the sensor configuration and the actuator configuration. The sensor is used to obtain the information of the external environment. The actuator provides the mobile ability, the position and attitude information to the system through its own sensor. It is foreseeable that the robot needs a powerful host to process all kinds of information in real time to ensure that the robot can respond to the command in a timely manner. The hardware platform of agricultural intelligent inspection robot has simple interactive function and autonomous navigation function. The hardware platform adopts a distributed architecture, which consists of two hosts, one is an airborne host, the other is a master host, and two hosts are connected directly with WiFi, and all the Ubuntu systems based on ROS are running. The on-board host is used to process information of sensors and actuator, feedback the information to the host computer, and receive commands from the host computer. The host receives the feedback information from the airborne host and issues commands to the on-board host.

4 Software system

The software system in the intelligent robot component is the robot's data processing system, just like a human brain, which is used to collect and process various information. Software system is the bridge between hardware platform and intelligent algorithm, which provides a good foundation for the intelligent robot. The coordinated control of the hardware platform provides information interaction for the upper level intelligent algorithm, and is an integral part of the intelligent robot design.

4.1 Software system selection

With the robot becoming more and more intelligent, robot software system is more and more huge, including the underlying hardware drivers, environmental awareness, motion planning, machine vision, logical reasoning and behavior decision making. For the realization and management of these functions are more and more complex. In order to effectively shorten the development cycle of the robot, reduce the R & D cost and improve the development efficiency, many companies and university research institutions have developed a unified robot software development platform; there is special software to achieve the simulation function. In addition, there is mature algorithm software. These software platforms have their own characteristics, covering all aspects of the field of robots.

ROS is an open source robot software platform, its purpose is to integrate different research results, achieve the algorithm release and code reuse robot software platform. It provides functions like operating systems, including hardware abstraction, underlying driver management, inter-process messaging, and program release management. Applications based on ROS development are currently up to 2000, covering hardware drivers, simulation, motion planning, motion control, environmental awareness and other aspects. As these features meet the needs of the majority of developers, ROS has been widely used. Therefore, this paper will use ROS system as a software system, and it can greatly speed up the robot design and development speed.

4.2 Operating mechanism

ROS processing data is a point-to-point network structure, so the program is running through the following: node, message, topic to achieve indoor autonomous mobile robot function.

A node is a process that executes a task in the ROS operating system and is an executable file in each function package. The modularity of the code in the ROS helps the richness of the intelligent function. The indoor autonomous mobile robot operating system is composed of many nodes. The function packs related to these nodes can be called software modules.

The communication between nodes is carried out through a message mechanism, and each message has its own data structure. The message can contain any array of nested structures, similar to the structure. The node announces a message to a specific topic, and the other node can subscribe to the subject by subscribing to the subject.

The publish/subscribe model in the main body is a very flexible communication paradigm, but this broadcast mode is not suitable for simplifying the node design synchronization mode. ROS, with a pair of more stringent message definition paradigm: request/response. This is similar to the operation of a web server, just like a remote call. The service call is bidirectional: a node sends a request to another node and waits for a response. Each service is initiated by a node, and the response to this service returns the same node, so that the differences between the message and the message can be seen.

A service is a paired message that occurs when a node requests information and another node needs to respond. The service can be used to send message between nodes. A providing ROS node offers a service under a string name, and a client calls the service by sending the request message and awaiting the reply.

ROS has a lot of nodes, messages, services, library files, development tools, and need the effective mechanisms to manage the code, such package, stack to play a role. ROS functional modules are organized in packages. There are nodes in the package, depending on the library, configuration files, etc. and the package provides a convenient portability and reusable software structure.

4.3 Automatic navigation

Robot autonomous navigation is mainly composed of two steps, the real-time positioning and map construction. The use of environmental information after the processing of the data to create and locate the map, it will create a good map to calculate the probability of the map, and get the global and local weight map. Global path planning to accept the weight of the map generated by the global weight map planning from the starting point to the target point of the path as a local path planning reference. Local path planning uses local weight map which is generated by weight map to plan robot's movement information. Then use the package to convert the data obtained by the sensor into real-time environmental information.

Map-based navigation: the built map information and real-time environmental information obtained by the sensor are transmitted to the server, and then refer to the global path to the robot real path for local planning to achieve the purpose of autonomous navigation.

5 Implementation of pedestrian tracking detection algorithm

Artificial intelligence algorithm, is the robot's soul, can make the robot behave more like people. The realization of the intelligent algorithm can give the machine owner some of the ability to replace the human to complete certain tasks. The pedestrian tracking algorithm used in this paper is to use robots to obtain images, first pedestrian detection. If there are pedestrians with a rectangular box circle, and for each pedestrian to create a tracker to track.

5.1 Pedestrian detection algorithm

In recent years, the R-CNN method for target detection and segmentation has been initiated by R Girshick et al., 2014, for the first time using the selective search (Uijlings et al., 2013) and the convolution neural network to combine the target detection. Subsequently, Kaiming He et al. can only read the fixed size of the problem, put forward the SPP-NET (He et al., 2014) method to further improve the detection results. After that, fast R-CNN (Girshick, 2015) and faster R-CNN (Ren et

al., 2015) have also been proposed to achieve end-to-end detection.

In this paper, a pedestrian detection method combining channel characteristics with faster-RCNN was used. Firstly, the VGG16 network pre-trained on ImageNet was used as the main network to extract the original image. After convolution, the original image generated a feature map (Conv4_3) with 256 channels.

Secondly, the channel feature of the image is extracted. The channel feature is preprocessed for the input image. A series of linear and non-linear transformations of the original image are carried out in the pretreatment stage. Each image is transformed into a channel by different transformations. In pedestrian detection, the LUV channel, the gradient intensity channel, and the gradient histogram channel are all very efficient channels.

Then, the channel feature is input into a two-layer convolution neural network. The convolution kernel size is 3, the step size is 1, the maximum pooling layer is 2, the step size is 2, and the output is a 128 channel of the activation map, the size of the input channel characteristics of 1/8.

And then all the obtained channel feature maps are in series, input to the RPN model, RPN model relative to the original version of some improvements, each anchors used a fixed aspect ratio of 0.41, and then from 40 pixels to 1.3 times the steps to generate 9 different sizes of anchors. The RPN model outputed the candidate target area.

The generated candidate target area is put into the Fast R-CNN model. The Fast R-CNN model has two full connection layers, a softmax classifier and a Bbox regressor. The softmax classifier output pedestrian detection results and the Bbox regressor output location information of pedestrian.

5.2 Pedestrian tracking algorithm

In recent years, the target tracking algorithm mostly uses the adaptive representation model based on the generated or discriminant formula. The discriminant method has better tracking performance than the generating method, and the method of using the incremental learning to distinguish the target and the background is less costly and tracking better. In this paper, we chose the online boosting algorithm and pedestrian detection process to extract the depth of channel features to achieve pedestrian tracking.

The tracking method uses a detection-based tracking strategy, that is, uses a sliding window to obtain a candidate region, and calculates the visual feature only once in the search area, and then extracts the features in each sliding window. And, this method uses a coarse to fine search strategy, you can quickly achieve tracking. First, at the beginning of the trace, the initial position is obtained by means of a detector or manually annotated, and then a number of positive and negative samples are collected and the boosting classifier is trained. In each frame of the subsequent sequence, the tracking area is a candidate region with the highest classification scoring. Some positive and negative samples are then collected after the end of the tracking and the classifier has been updated until the end of the tracking.

The tracking method uses the depth channel feature, and the depth channel feature is a feature of the channel feature combined with the faster RCNN method. The network structure and training mode are the same as in the previous section, but need to combine the network output before the full connection layer of the faster-RCNN network with the underlying channel characteristics as the depth channel feature. And then put it into the boosting classifier for training.

6 Conclusion

In this paper, for agricultural production process, many scenes need to monitor the situation, design an agricultural intelligent inspection robot. The hardware platform is mainly composed of three parts: hardware platform, software system and intelligent algorithm. The hardware platform is mainly composed of sensors, actuators, mechanical structures and electronic systems. It is the foundation of robot design. The software system is mainly composed of ROS system. The intelligent algorithm is mainly the pedestrian tracking detection algorithm, use the robot to obtain the image, make the first pedestrian detection, and then the pedestrian has been detected to the tracker for pedestrian tracking.

Finally, the monitored pedestrian information is transmitted back to the master computer through WiFi. Agricultural intelligent inspection robot has high practical value.

[References]

[1] Ali, N. H., and G. M. Hassan. 2014. Kalman filter tracking. *International Journal of Computer Applications*, 89(9): 15–18.

[2] Bo, W., and R. Nevatia. 2005. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. *Proceedings of the IEEE International Conference on Computer Vision*, Beijing.

[3] Chang, C., and R. Ansari. 2005. Kernel particle filter for visual tracking. *IEEE Signal Processing Letters*, 12(3): 242–245.

[4] Chen, F. S., C. M. Fu, and C. L. Huang. 2003. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing*, 21(8): 745–758.

[5] Comaniciu, D., V. Ramesh, and P. Meer. 2000. Real-time tracking of non-rigid objects using mean shift. *In Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition.* Hilton Head, SC, USA: IEEE, 142–149.

[6] Dalal, N., and B. Triggs. 2005. Histogams of oriented gradients for human detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, IEEE.

[7] Girshick, R. 2015. Fast r-cnn. *Proceedings of the IEEE International Conference on Computer Vision.* 1440–1448.

[8] Girshick, R., J. Donahue, T. Darrell, and J. Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. June 24-27, 2014: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 580–587. Ohio, USA.

[9] He, K., X. Zhang, S. Ren, and J. Sun. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9): 1904–1916.

[10] Jain, R., and H. Nagel. 1979. On the analysis of accumulative difference pictures from image sequences of real world scenes. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1(2): 20–214.

[11] Lipton, A., H. Fujiyoshi, and R. Patil. 1998. Moving target classification and tracking from real-time video. *Proc. Fourth IEEE Workshop Applications of Computer*, Princeton, NJ, pp.8–14.

[12] Meier, T., and K. Nagank. 1998. Automatic segmentation of moving objects for video object plane generation. *IEEE Trans. Circuits and Systems for Video Technology*, 8(6): 525–538.

[13] Ojala, T., M. Pietikainen, and D. Harwood. 1996. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1): 51–59.

[14] Papageorgiou, C., M. Oren, and T. Poggio. 1998. A general framework for object detection. *Proceedings of the International Conference on Computer Vision*, Bombay, India, IEEE.

[15] Ren, S., K. He, R. Girshick, and J. Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137–1149.

[16] Seber, G. A. F., and A. J. Lee. 2003. Linear regression analysis (Second Edition). New York: *John Wiley & Sons.*

[17] Uijlings, J. R. R., K. E. A.V. D. Sande, T. Gevers, and A.W.M. Smeulders. 2013. Selective search for object recognition. *International Journal of Computer Vision,* 104(2): 154–171.

[18] Wang, Q., F. Chen, W. Xu, and M. H. Yang. 2011. An experimental comparison of online object-tracking algorithms. SPIE Optical Engineering+ Applications. *International Society for Optics and Photonics*, 8138(3): 815–822.