Design and Implementation of Parts Visual Recognition System for Intelligent Manufacturing

Jihong Yan*, Ph.D.
Professor in Industrial Engineering
Deputy Dean of School of Mechatronics Engineering
Head of intelligent Manufacturing Scientific Research Team
Harbin Institute of Technology
Harbin, China
jyan@hit.edu.cn

Yuhao Zhang
School of Industrial Engineering
Harbin Institute of Technology
Harbin, China
zyhfromxj@qq.com

Abstract

Sorting task is one of the main activities in manufacturing. Traditional industrial sorting technology is laborious, time-consuming and inefficient, and it is difficult to meet the needs of automated long-term operations. Therefore, this paper designed a vision-based workpiece recognition system for intelligent manufacturing, which applied deep learning methods to realize the recognition and localization of workpieces to drive the robotic arm to sort multiple types of workpieces. In this paper, the transfer learning method was used to train the enhanced data to recognize new images. Filtering methods such as Gaussian, Bilateral, and morphological transformation methods such as expansion, erosion, and opening operations was applied to achieve image denoising and distortion elimination. Finally, through the feature matrix calculated by processed image data, the information such as the centroid position and the deflection angle can be obtained, which lay the foundation for accurate localization and rapid sorting of workpieces.

Keywords

Sorting robot; Transfer learning, Edge detection; Filtering;

1. Introduction

In recent years, China's technological level has advanced by leaps and bounds, and the development trend of intelligence and automation in the manufacturing industry is very rapid. A highly intelligent and automated manufacturing plant can greatly improve efficiency, reduce the proportion of manual labor, and bring great economic benefits [1].

In traditional industrial and logistics sorting systems, most of them adopt a manual work-oriented model, which is costly and time-consuming. It is difficult to meet the long-term and high-intensity work requirements of the industrial production, which greatly restricts the industry production efficiency. Nowadays, more and more researchers combine machine vision with industrial sorting robots to provide "eyes" for industrial sorting robots, allowing the robots to autonomously recognize the types and poses of workpieces, and can handle complex workpiece positions, which shows the robot has a higher level of intelligence. Therefore, this article will design a kind of automatic parts recognition and positioning system based on the background of industrial sorting robots.

After fully considering the requirements of parts classification and recognition and positioning, this article will mainly focus on the convolutional neural network model and the related visual positioning algorithm based on OpenCV. At present, convolutional neural networks have developed very maturely and have been widely used. As long as a large amount of image data can be provided for training, the excellent performance of the convolutional neural network is sufficient to meet the recognition of multiple types of parts, and it is sufficient to meet the real-time performance of dynamic capture of parts recognition. In addition, OpenCV-based image processing methods are very rich. By using the calculation method of feature moments, the contour information, centroid information, and pose information of the part can be accurately captured.

2. Literature Review

After fully considering requirements of parts classification, recognition and positioning, this article will mainly focus on the Machine vision and the related visual positioning algorithm based on OpenCV. The Machine vision have developed very maturely and have been widely used and the convolutional neural network is one of it. The convolutional neural networks have developed very maturely and have been widely used. As long as a large amount of image data can be provided for training, the excellent performance of the convolutional neural network is sufficient to meet the recognition of multiple types of parts, and it is sufficient to meet the real-time performance of dynamic capture of parts recognition. The OpenCV-based image processing methods are very useful for the obtainment of the Image contour information. By using the calculation method of feature moments, the contour information, centroid information, and pose information of the part can be accurately captured.

2.1 Development status of machine vision and convolutional neural network

Since the last century, the machine vision industry has developed vigorously abroad, and the corresponding technology has been applied more widely, and the supporting equipment has become more advanced.

In the manufacturing field, Wu [2] applied machine vision to the PCB board manufacturing production line to monitor and optimize the quality of the PCB board; Wenda [3] used image segmentation technology to detect surface defects of air bearings; Liu [4] proposed a new sparse coding model to improve the real-time and accuracy of machine vision in surface defect detection; Hashemi [5] used machine vision to check the roundness of the end of the pipe, which reduced the problems of pipe welding and later connection, and detected the roundness in a non-contact way to ensure the quality of the pipe; Based on fuzzy logic, Lee [6] selected gray-scale angles as local features, and obtained information to make models through gray-scale angle detectors. The actual industrial components have been identified and detected as objects, and good application effects have been achieved.

Foreign developed countries started to study machine vision earlier. Relying on its leading computer science and technology, machine vision is relatively mature in experimental research and practical applications, and its application range is quite extensive.

Deep Learning belongs to the category of machine learning. In recent years, it has achieved good results in classification recognition and speech recognition. Deep learning uses the method of simulating the connection mode of human brain neurons, processing the features at a level, and then abstracting the deep-level feature information [7-9] .

Convolutional Neural Network (CNN) is one of the deep learning network models. It has good classification and recognition performance. It consists of an input layer, a convolutional layer, a down-sampling layer (pooling layer), a fully connected layer, and an output layer.

The input layer is the input of the neural network model, and that is the original image data. The convolutional layer is to process the original image data of the input layer, and perform convolution operations through the convolution kernel to extract the features of the original image data. The down-sampling layer is generally placed after the convolutional layer to perform a pooling operation to process the original image data. On the one hand, it can reduce the dimensionality of the feature map, and on the other hand, it can also keep the feature scale unchanged. The fully connected layer, structurally speaking, is a simple neural network layer. Each neuron on it is fully connected to the neuron in the previous layer of network structure. It serves as a function of integrating and summarizing the training data information in the previous layer. The output layer is located at the end of the convolutional neural network and is used to output the result [10] .

2.2 Basic knowledge of image processing based on OpenCV

OpenCV contains many image processing and computer vision algorithms, such as motion analysis, target detection, edge detection, color recognition, target segmentation and recognition, and 3D reconstruction, etc.

In recent years, OpenCV has been widely used in mathematical morphology image processing, and

has shown great power in the field of image recognition, which is mainly reflected in the following aspects:

(1) Image feature detection:

Among the existing examples of recognition technology, a large proportion of scholars and researchers have adopted OpenCV-based image feature detection technology. Image defogging technology, adaptive threshold noise reduction technology, image segmentation technology, etc. have been widely used.

Liu Huiying and Wang Xiaobo proposed a new method of contour detection. The goal is to obtain a perfect contour line for vehicle recognition. After the OpenCV image processing operation, the target curve is successfully extracted. Lei Jianfeng and Wang Wei realized a fixed threshold segmentation method and an adaptive threshold segmentation method for images. This method provides great convenience for subsequent image analysis and feature extraction, and also provides technical support for multi-target real-time detection, which has strong practical value. Li Huachen studied a variety of image edge detection technologies, and compared a variety of edge detection technologies with VS and OpenCV related algorithm codes, and successfully optimized some of them in actual application scenarios.

(2) Image content recognition:

Zhao Jian and Zhang Dongquan studied number gesture recognition algorithms. Using the HSV color recognition method, the recognition of ten digits by human hands is realized, and the accuracy rate is 97.4%. Lü Kun et al. proposed a license plate recognition method, which uses edge detection, color detection, license plate feature detection and SVM to achieve precise positioning. Guo Zhonghua and others have conducted research on face recognition systems, and successfully applied face recognition technology in community access control systems, which has high practical value.

This article mainly uses OpenCV technology for image processing, and obtains the pose information of the part through grayscale, binarization and feature moment calculation. However, because the spots and scratches in the background of the original image interfere with image processing to a certain extent, it is necessary to adopt related morphological conversion and filtering methods for noise reduction processing, so that the subsequent calculation results are better.

3. Data Collection

3.1 Collect data

By Using the Hikvision camera Windows driver MVS , we can collect the data. The video is captured in the MVS program first, and the frame is intercepted from the video as the original data. Figures1. are the raw video datas captured by the Hikvision camera.

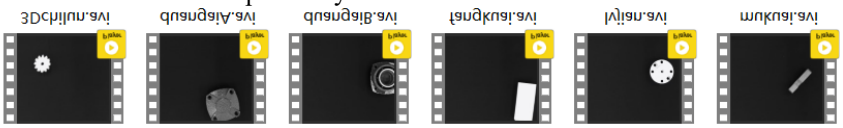


Figure 1. Video data

3.2 Data enhancement

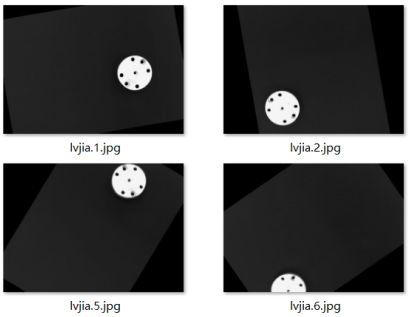


Figure 2. Image data after data enhancement

The richness of picture data has a great influence on the training effect of deep learning. When the sample space is too small or the sample size is seriously insufficient, the results of model training will be seriously affected, the generalization degree of the model will be greatly reduced, and the recognition rate and accuracy rate will be seriously reduced.

Therefore, because we can only obtain limited dataset, so the data should be enhanced, which can greatly increase our data abundance and improve the effect of model training to a certain extent.

This article uses python code for data enhancement, mainly through geometric transformation to enhance the original image data. Figure 2. shows the image data that has been enhanced. You can see that the posture of the workpiece in the picture is more abundant after being processed.

3.3 Build the dataset

For the pictures in the dataset, we divide them into a training set (train) and a test set (test). The training dataset is used to train the model, and the test dataset is used to test the training effect, so as to test the accuracy and loss to find the optimization direction during the training of the convolutional neural network. In the dataset, the ratio of the training set to the test set reaches 8:2 to achieve the optimal training results.

4. Methods

4.1 The hardware conditions and the development environment

First of all, we should choose the camera. The camera we chosen is the Hikvision Industrial Camera, MV-CE100-31GM which has the characteristics of low noise, excellent image imaging, and high cost performance. The computer is equipped with AMD Ryzen 7 CPU, graphics card 2060 MAX-Q, and 16G RAM. The development environment adopts the method of configuring the Anaconda manager, and the language is selected as python3.6.2.

4.2 The VGG16 model for training

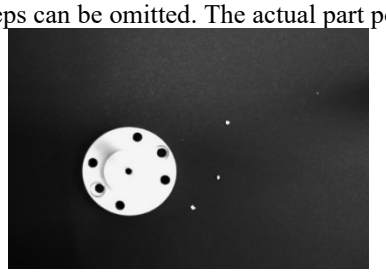
Import the dataset constructed in 2.3 into the VGG16 model for training. The training process is shown in Figure 1-10 above. It can be seen that the loss gradually decreases and the accuracy rate gradually increases. In the training process, the parameters need to be adjusted to determine the reasonable number of training layers (epoch). When the loss does not increase but decreases, it can be judged that the model has overfitting (Refers to that the training set is too small, the prediction data is too similar to the training data, resulting in the poor training prediction effect of the model). Finally, the H5 weight file is obtained, which is used to predict the new data.

In the training process, in order to unify the output results of recognition and positioning in the following text, we named the six types of artifacts respectively. The specific naming conditions are as follows:

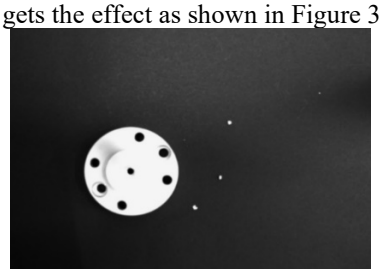
Part A---Aluminum parts;
Part B---3D Print Gear;
Part C---3D Print Block;
Part D---End cover A-face;
Part E---End cover B;
Part FF---Wood block.

4.3 Part attitude acquisition based on OpenCV

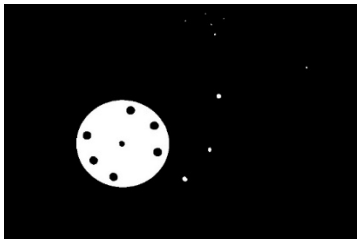
Because the Hikvision MV-CE100-31GM industrial camera is a black-and-white camera, the grayscale steps can be omitted. The actual part position gets the effect as shown in Figure 3. below:



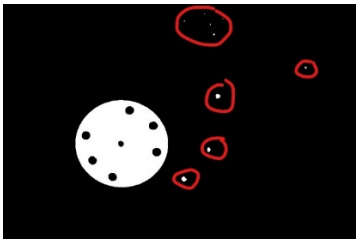
a)The original picture read by the camera



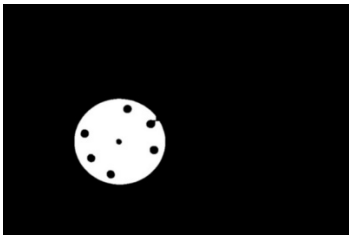
b) A picture that has been filtered by Gauss



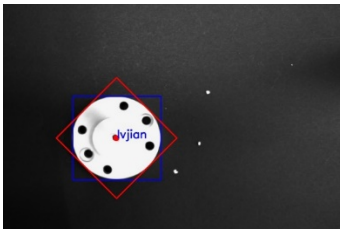
c)A picture that has been treated with dualization



d) Mark the red part as light spot and bottom plate spot



e) After open operation and closure operation



f) Full effect of identifying and positioning

Figure 3. The processing process of identifying and positioning the overall image

Figure a) is the original picture read by the camera. Since the background plate is a black base plate, it can be seen that the light will change the color of the background to a certain extent, and the background plate cannot be completely non-reflective.

Figure b) is a picture after Gaussian filter processing. The Gaussian filter used in this article is 5×5, which can maintain a certain filtering effect while lowering the degree of image blur and retaining better image clarity.

Figure c) is a picture after the binarization operation. It traverses every pixel in the picture, resets the pixels within the threshold to 1, and resets the remaining pixels to 0, which is convenient for subsequent processing and calculation.

Figure d) marks the position of the spots on the picture after the binarization operation. After Gaussian filtering, some subtle features in the image (such as small spots and textures on the bottom plate) have been filtered out. However, larger spots will still be recognized, showing white spots in the field of view. In addition, areas with strong light will also be identified, which is directly above the field of view in the picture. When collecting data, for the low-light intensity, it also appears as small white spots. For changes in lighting conditions, binarization operation processing is very sensitive at low thresholds, and it is prone to misrecognition problems caused by changes in lighting conditions. This article has made certain optimizations for this problem, which will be described in subsequent chapters.

Figure e) is the picture after the "open operation-close operation" of the morphological conversion which can eliminate all the spots in the binarization picture and achieve a good recognition effect.

Figure f) is a display picture of the complete effect of identification and positioning. The identified parts are aluminum parts, and the identification results are printed in real time in the figure. In the figure, the blue rectangle is a right-angled rectangle, the red rectangle is a rotating rectangle, and the red dot is the center of mass position. Since the calculation of the characteristic moment covers all the pixels, although the part in the figure is the center of the circle, the rotating rectangle still has an angle of approximately 45°, which is mainly affected by the hole of the part.

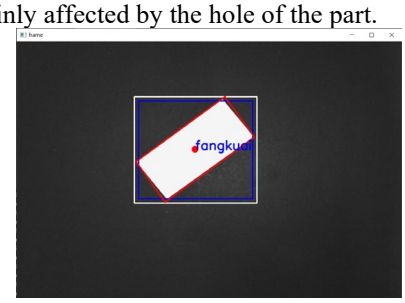


Figure 4. Block Parts Identify the results

Figure 4. is the recognition result of the block part, and the recognition and positioning real-time results are printed in the figure. This figure adds a new white rectangle, which is the block diagram of the multi-target detection and recognition of the subsequent program.

In summary, the program in this article implements the basic functions, completes the recognition of the workpiece category, the real-time feedback of the workpiece position (the position of center of mass) and the deflection angle (the angle between the right-angled rectangle and the rotating rectangle).

4.4 Camera calibration

1) Get internal parameter matrix and external parameter matrix

The internal parameter matrix is an inherent parameter of the camera, which will not change after the camera leaves the factory, so its value should be unique. Two methods are used to determine the internal parameter matrix: the first one is through the OpenCV-Python code to obtain the internal parameter matrix and the second one is through the relevant calculation software the to obtain it.

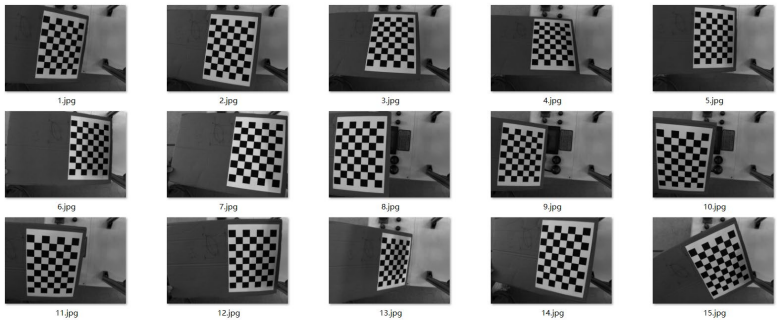


Figure 5. Obtain the calibration diagram of the internal parameter matrix

Figure 5. shows the calibration diagram. In the calculation process, the corner points of the checkerboard are selected for calibration. Each checkerboard calibration drawing can obtain a rotation vector rvec and translation vector tvec. The rotation vector and translation vector of each checkerboard itself refer to the checkerboard angle. The point is the transformation of the origin coordinate system relative to the world coordinate system.

Furthermore, when obtaining the internal parameter matrix, we take multiple photos of different angles for calibration. Under the combined action of these calibration pictures, the internal parameter matrix can be obtained.Then through the rotation vector and translation vector of each picture itself, you can transform to the world coordinate system and get the real coordinates.

2) Data transmission

Use MySQL related sentences to realize the function of uploading workpiece type and location information in real time, which is convenient for the robotic arm to grasp.

5. Conclusion

This paper has completed the design of the parts visual recognition system for intelligent manufacturing, completed the recognition, classification and positioning of the parts, explored the relevant theories of convolutional neural networks in the field of deep learning, and explored the relevant image processing methods and related methods based on OpenCV Algorithm, explored the method of camera calibration and related coordinate system conversion. The main work content is as follows:

1) Under the tensorflow machine learning framework, the keras deep learning library is used, and the VGG16 convolutional neural network model is selected to achieve accurate recognition of the workpiece, and the accuracy rate can reach 99%.

2) Using OpenCV-based image processing methods, Gaussian filtering and morphological conversion methods are used to denoise the image, edge detection methods are used to extract part contour information, centroid position and rotation angle are calculated through feature moments, and the contour extraction success rate of parts can reach 99%, and the accuracy of pose acquisition can reach 99%.

3) Complete the coordinate transformation from the pixel coordinate system to the world coordinate system, establish the difference capture strategy relative to the reference point, and realize the sorting of multiple types of parts by the driving robot arm. The overall effect is good.

On the basis of realizing the basic functions, there are several optimization points to the system:

1) Optimized the lighting conditions based on the industrial production environment to improve the anti-light interference of the program;

2) Using contour recognition and image segmentation technology, based on the single-output weight model of the VGG16, multi-target detection and real-time tracking are realized, which has a good effect and can meet the real-time capture function of the robot.

The content completed in this article is highly adaptable to industrial sorting scenarios, and can be used for assembly line sorting parts and AGV sorting parts, which can improve the efficiency of industrial sorting to a certain extent and increase economic benefits. What's more, the relative reference point calibration scheme provides another feasible scheme for the realization of the general precision grasping scheme.

In addition, this system still has certain improvements:

1) Because the camera is black-and-white camera, the gray-scale processing process is omitted. But the black-and-white image training model is generally robust, and the situation of misrecognition is serious, so color cameras can be considered as a substitute for identification.

2) The stability of the relative reference point calibration scheme is a little weak. The hand-eye calibration method , which means that the origin of the world coordinate system can be established in

the world coordinate system centered on the robot, should be further studied, which is convenient for the robot to understand and helps to improve the stability of the system.

Acknowledgements

This work was supported in part by the JCKY Project of China (#JCKY2019603C016).

References

[1] Rao AR. Future directions in industrial machine vision: a case study of semiconductor manufacturing applications[J]. Image and Vision Computing, 1996. 14(1): 3-19.
[2] Shuangxi W, Mengxia H, Jiaxing Y, et. al. An Auto Backlight Screen Inspection Equipment Based on Machine Vision[C]//Seventh International Conference on Measuring Technology & Mechatronics Automation. IEEE, 2015.
[3] Wenda C, Ruilin B, Feng J I, et al. Bearing shield surface defect detection based on machine vision[J]. Computer Engineering and Applications, 2014.
[4] Liu Z, Hu J, Hu L, et al. Research on On-Line Surface Defect Detection for Steel Strip Based on Sparse Coding[J]. Advanced Materials Research, 2012, 548: 749-752.
[5] Hashemi S H, Moradpour M A, Karimi E, et al. The use of machine vision inspection for out of roundness measurement in steel pipes[C]//Pattern Recognition and Image Analysis (PRIA), 2013 First Iranian Conference on. IEEE, 2013.
[6] Lee K J, Bien Z. A model-based machine vision system using fuzzy logic[J]. International Journal of Approximate Reasoning, 1997, 16(1): 119-135.
[7] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1) : 1-127.
[8] SERRE T, KREIMAN G, KOUH M, et. al. Aquantitative theory of immediate visual recognition[J]. Progress in Brain Research, 2007, 165: 33-56 .
[9] Hubel D H, Wiesel T N. Receptive Fields, Binocular Interaction and Functional Architecture in the Cats Visual Cortex. Journal of Physiology, 1962, 160(1): 106-154
[10] Fukushima, K. (1980) Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. Biological Cybernetics, 36, 193-202.