Research on the method of cement bag palletizing system based on machine vision

Zujia Zheng, Ruyi Ma
Infrastructure and Logistics Management, Wuhan University of Science and Technology, Wuhan 430081, China
zhengzujia@qq.com

Abstract. Aiming at the problem that the traditional cement bag palletizing robot can only accomplish the simple task of handling operation, but can not meet the production requirements, this paper proposes a kind of cement bag palletizing robot based on machine vision, which applies machine vision technology to the traditional cement bag palletizing robot. Firstly, the camera is used to calibrate the visual system of the target, then the image is collected and preprocessed. Finally, the cement bag is recognized according to the SIFT matching algorithm. This method improves the intelligence level of traditional cement bag palletizing robot, makes it have the ability of detection, analysis and decision-making, improves the operation quality of the cement bag palletizing robot, and expands its application scope.

1. Introduction
Construction industry is a major industry in our country. It is closely related to people's lives and is a very important part of our national production. Since the 21st century, with the development of society and the progress of science and technology, the application of artificial intelligence in the construction industry is more and more[1]. There are many drawbacks in the traditional building mode. AI can solve many problems in the traditional building mode. So the application of AI in the human construction industry is an inevitable trend.

The cement bag palletizing is an indispensable type of work in our country's existing buildings. In the past, the cement bag palletizing robot has the problems of low efficiency and low quality, which makes the importance of the cement bag palletizing robot based on machine vision very prominent[2]. Machine vision is to extract information from the image of the object, process and complete the judgment, and then guide the operation and action of the robot in a wide range. It is widely used in automatic optical inspection, face recognition, tracking and positioning of products and other aspects. The cement bag stacking robot is to stack the cement bags in the designated position in accordance with a certain pattern, so as to realize the storage, handling, loading and unloading, transportation and other logistics activities of the cement bags[3].

However, at present, in the construction industry, the cement bag palletizing robot can only complete the simple task of handling operation, but can not analyze the size, shape, quality of the palletizing object and complete the corresponding judgment, nor can not complete the intelligent palletizing of the cement bag. The traditional cement bag stacking robot can not meet the actual production needs. If machine vision is applied to the cement bag robot stacking operation, the intelligent cement bag stacking task can be completed[4]. Aiming at a series of problems existing in traditional cement bag stacking robot, this paper will mainly discuss the research and application of machine vision in cement bag palletizing system.
2. Visual system calibration

When building machine vision system, it is necessary to establish the corresponding relationship between the position of image pixels in the camera and the position of three-dimensional scene points, and then extract three-dimensional spatial information from two-dimensional image information. The corresponding relationship of position is determined by the camera's imaging model and its internal and external parameters. The camera's internal and external parameters are obtained by experiment and calculation. The process of obtaining them is called camera calibration. The key step of camera calibration in vision system also determines the positioning accuracy of the system[5]. The transformation relationship between camera coordinate system and robot world coordinate system can be deduced by the result of camera parameter calibration and matrix transformation between coordinate system and robot world coordinate system.

The planar target chosen in this paper is 10x7 black-and-white chessboard target. And the size of the square is 25x25mm. A set of calibration images were taken with Kinect camera for planar chessboard targets with different positions and postures. Zhang Zhengyou's calibration method only needs three calibration images to obtain camera parameters[6]. To reduce the calibration error of the camera, nine calibration images were taken in this group, as shown in Fig.1.

![Fig.1 Planar chessboard target image](image)

The camera calibration experiment in this paper is carried out under the programming environment of Matlab, using the calibration toolbox of Matlab. The main function in the toolbox is provided by Dr. Jean-Yves Bouguet of California Institute of Technology. The toolbox can be applied to different calibration environments and methods[7]. Therefore, on the basis of the toolbox, the plane target image is calibrated by Zhang Zhengyou calibration method.

3. Image acquisition and preprocessing

3.1. Image acquisition

Digital image is composed of pixels, which is a numerical matrix consisting of the values of pixels as elements. Image acquisition is also called image digitization, that is to say, it receives the image of the target object through the sensitive device in the optical lens, and converts it into the electrical signal that can be processed by the computer, and then the electrical signal is sampled and quantized to become a digital image. The camera selected in this paper is Kinect camera, which has the characteristics of low
power consumption, good stability and high resolution. Through the application interface program of the camera, it can easily control the camera to collect the image of the target object[8].

3.2. Image preprocessing

Digital image will inevitably have various kinds of noise and distortion signals after various conversions and processing, so it is not suitable for direct application in subsequent image analysis and visual recognition. In order to extract useful information from digital images more accurately, the image preprocessing process must be used to eliminate the interference of useless information such as noise to the maximum extent, and enhance the weak image signal to make it easy to process. The image preprocessing in this paper mainly includes three parts: image graying, image denoising and image enhancement.

(1) Image graying

The Kinect camera used in this paper is a color camera. The original image is a color image, while the gray image is needed for subsequent image analysis and visual recognition. So it is necessary to convert the color image into a gray image.

(2) Image denoising

Image denoising is mainly to remove or reduce the interference of noise signals and improve the quality of images. This process is actually a low-pass filtering of images. The main methods of image smoothing are low-pass filtering, mean filtering, median filtering and so on. By comparing various filtering methods, this paper finds that the median filtering method has the best effect of voice removal. It not only removes the interference noise in the image, but also retains the clear contour of the original image, which meets the requirements of recognition application.

(3) Image enhancement

Image enhancement is to enhance and highlight the feature information such as contour, edge and contrast in the image for further analysis and recognition. The simplest and most commonly used technique in image enhancement is contrast enhancement, which can process the gray value of each pixel and correspond to each other one by one. The typical one is the linear transformation of gray level. Linear transformation can expand the dynamic range of the image and increase the contrast of the original image. After linear transformation, the image is clearer and clearer, and the overall quality of the image is effectively improved.

4. Type Recognition of Cement Bags

4.1. Analysis of Recognition Method

When the cement bag enters the visual recognition area, the PC drives the camera to collect the cement bag image and preprocess the randomly collected cement bag image, then extracts the SIFT feature vector of the cement bag image, and matches the SIFT feature vector of the cement bag image to be recognized with the SIFT feature vector of the template image in the template image library. If the matching is successful with one of the template images, the type of cement bag can be identified. If the matching fails, the next template image can be matched until the matching is successful. If the matching between the identified image and all template images is not successful, the system collects the image to be recognized again and modifies the matching parameters to match again. Before matching processing, the system needs to save the pre-processed image to be recognized to prepare for the position recognition of cement bags. The correct matching feature pairs should also be preserved after the type identification in order to identify the angle information of the cement bag.

The recognition of cement bag type by image matching is based on SIFT algorithm, so the principle of SIFT matching algorithm needs to be described.

4.2. SIFT Matching Algorithms

SIFT (Scale Invariant Feature Transform) calculation method was first proposed by David G. Lowe in 1999, and it was improved and summarized in 2004[9]. The SIFT method makes use of the local
invariance of the image to recognize the feature points in the image, and the method can achieve high results in translation, rotation and radiation transformation between images. The algorithm has been widely used in many fields such as image sensing, image fusion, pattern recognition and machine vision[10].

SIFT matching algorithm consists of SIFT feature point detection and SIFT feature vector matching. Specifically, it can be divided into building image scale space, determining key points in image scale space, eliminating bad key points, assigning the direction of key points, generating feature descriptors of key points and matching SIFT feature points.

1. Establishment of image scale space
   The construction of scale space as the initialization operation of SIFT matching algorithm is mainly to simulate the multi-scale characteristics of images. Among them, the Gauss convolution kernel is the only linear transformation kernel to complete multi-scale spatial transformation, so a two-dimensional image can be obtained.

2. Determining Key Points in Image Scale Space
   To find out the key points in scale space, it is necessary to compare each sampling point with all neighboring points around the point and observe the size relationship between the sampling point and its image and scale domain. If a detection point is detected to be the maximum value in the three layers of DOG scale space, it is determined that this point is the key point in the image scale.

3. Eliminate the key points
   The location and scale space of key points can be accurately inferred by fitting the three-dimensional quadratic function. However, in order to enhance the stability and anti-jamming of matching, the key points with relatively low contrast must be removed. In addition, because DOG operator produces strong edge response in the process of processing, which results in the instability of the edge response points, Hessian matrix can be used to calculate the principal curvature of the Gauss difference operator, and the unstable edge points can be removed by setting appropriate thresholds.

4. Direction of key points allocation
   After detecting the feature points in each image, it is now necessary to assign a direction to each feature point in the image. The gradient distribution of the pixels around the feature points can be regarded as the direction of each feature point, and it can also have the invariance of rotation.

5. Matching of SIFT feature points
   When the feature vectors of two images to be matched are generated, Lowe uses a method of comparing the nearest neighbor distance with the next nearest neighbor distance. By setting a threshold, when the ratio of the distance between the two images is less than the threshold, it can be determined that the two feature points can be matched successfully. That is, using the Euclidean distance of the feature vectors to distinguish the feature points of different images. Specifically, a feature point is selected in one image to be matched, and two feature points with the smallest distance from the point are found in another image. If the ratio of the distance between the point and the other two points is set, the matching points are accepted.

5. Posture Recognition of Cement Bags
   Different types of cement bags fall randomly on the conveyor belt and their postures are not fixed, mainly referring to the rotation angle and geometric center position on the conveyor belt. Therefore, it is necessary to recognize the rotation angle and geometric center position of the cement bags immediately after identifying the type of cement bags. Accurate recognition of the position and posture information of the cement bags is the foundation and key for the robot to successfully complete the operation of the conveyor belt. In order to obtain the position information of the target object, many scholars have done a lot of research. Dong Tantan of Nanjing Agricultural University calculated the centroid position and inertia spindle of mature tomatoes by moment method, and identified the growth posture of tomatoes from the angle of inertia spindle. Wu Yiquan et al. used the corner information of the characters in the license plate and the inertial spindle to detect the tilt angle of the license plate. Fan Jiangyan used Hough line detection method to identify the deflection angle of profiles[11].
5.1. Identification of Center Position of Cement Bag

For the identification of the center position of the cement bag, firstly, the image of the cement bag stored in the early stage of matching the type of the cement bag is read, and the image has been preprocessed. Then, the gray image is binarized by using the maximum variance between classes, and the target image is segmented from the background image. There will be some small interference areas and small holes in the binary image of cement bags separated, so we need morphological open operation and morphological close operation in image processing to eliminate interference and holes, get clear and complete binary image of cement bags, and finally calculate the centroid pixel coordinates of cement bags in binary image, that is, the center of cement bags image[12]. After identifying the center coordinate of the cement bag in the cement bag image, the real center position of the cement bag can be obtained by calibrating the internal and external parameters and the transformation relationship between the image coordinate system and the world coordinate system through the front camera. The process of identifying the center position of the cement bag is shown in Fig. 2.

Fig. 2 Central Location Recognition Process

A lot of experiments have proved that the center position recognition program used in this paper can accurately identify the center position of the cement bag. Because the edge of the cement bag is irregular, some images of the cement bag can not represent the real edge characteristics after morphological operation, but its range is relatively small compared with the whole cement bag, so the center of mass position of the binary image of the whole cement bag can be identified. Detection effects can be neglected.

5.2. Rotation Recognition of Cement Bag

For the recognition of rotation angle of cement bags, this paper identifies the rotation angle of cement bags based on SIFT feature matching. In particular, it uses a series of key points matched correctly when the type of cement bags matches, carries on affine transformation to the key points, finds out the spatial transformation relationship, and then finds out the rotation angle of cement bags. The principle of affine transformation is briefly introduced below[13].

Affine transformation is an important linear geometric transformation. It can describe the imaging process from the object in three-dimensional space to the two-dimensional plane, and has the invariance of translation, rotation, and scaling.
The program of angle recognition in this paper is completed in the environment of Matlab programming. Firstly, it loads the correct matched feature points pairs into the workspace through the program. In the program, the affine transformation relation and tilt angle of the feature points are deduced by using the space transformation function in the image processing toolbox. Finally, the recognition results are displayed in the cement bag image.

6. Conclusion
A machine vision recognition system is constructed for the type and pose recognition of cement bags. The camera calibration experiment of vision system is completed by Zhang Zhengyou calibration method. The automatic matching between the characteristics of the identified cement bags and the features of the template image is realized by SIFT algorithm, and the type of cement bags is accurately identified. On the basis of matching feature points correctly, affine transformation of feature points is used to identify the rotation angle of cement bags. The centroid position of cement bags is calculated by image segmentation, binarization and morphological operation, and the type and pose of cement bags are automatically recognized.

References
[1] Li Gongfa, Jiang Du, et al. Human Lesion Detection Method Based on Image Information and Brain Signal [J]. IEEE Access, 2019, 7: 11533-11542.
[2] Sun Ying, Li Cuiqiao, et al. Gesture Recognition Based on Kinect and sEMG Signal Fusion [J]. Mobile Networks and Applications, 2018, 23(4):797-805.
[3] He Yang, Li Gongfa, et al. Numerical Simulation-based Optimization Of Contact Stress Distribution and Lubrication Conditions in the Straight Worm Drive[J]. Strength of Material, 2018, 50(1):157-165.
[4] Li Yongjie, Zang Chen, et al. Motion Control and Visual Compensation of Spraying Robot [J]. Machinery Design & Manufacture, 2018, (12):257-261(in Chinese).
[5] Xie Fenglong, Han Jianhai, et al. A Fast Way of Stable Camera Calibration with Robot [J]. Machinery Design & Manufacture, 2018, (11):237-240.
[6] Zhong Xungao, Xu Min, et al. Image-based Uncalibration Visual Feedback Control Method for Robot Global Positioning [J]. Natural Science, 2018,57(3):413-419(in Chinese).
[7] Gao Li, Hu Hong, et al. Research of the Calibration Method for Auto Plug in Machine Based on Machine Vision [J]. Machinery and Electronics, 2016,34(11):77-80(in Chinese).
[8] Sun Jiangang, Gao Na, et al. Research of the building slabstone installing robot vision system and the calibration method [J]. Journal of Hebei University of Technology, 2013,42(4):64-68(in Chinese).
[9] Du Qing, Miao Yanhua, et al. Design of vehicle automatic classification algorithm based on machine vision [J]. Electronic Test, 2019,(1):56-58(in Chinese).
[10] Liu Wenchao, Xia Zhenqiao, et al. Research on Position Measurement of Vision-Guided Handling Manipulator [J]. Chinese Journal of Sensors and Actuators, 2018,31(11):1674-1678(in Chinese).
[11] Wang Xiangming, Liu Mingchun, et al. Visual identity system for automatic feeding robot [J]. Journal of Shenyang University of Technology, 2018,40 (5):564-570(in Chinese).
[12] Tan Chong, Sun Ying, et al. Research on Gesture Recognition of Smart Data Fusion Features in the IoT [J]. Neural Computing and Applications, 2019.
[13] Chang Wenjun, Li Gongfa,. et al. Thermal Mechanical Stress Analysis of Ladle Lining with Integral Brick Joint [J]. Archives of Metallurgy and Materials, 2018,63(2):659-666.