

# Study on Dynamic Target Positioning and Grabbing Based on Binocular Vision

**Jiatong Han**

*Hohhot Municipal Engineering Technology Service Center, Inner Mongolia, 010023, China*

**ABSTRACT.** *Because binocular stereo vision technology is a hot technical content applied to computer research in the field of computer and vision technology. It can quickly obtain and restore 3D image information about people and objects in computer scenes, and is widely used in robot recognition and positioning, industrial manufacturing and automation, and many other computer fields. Therefore, the research and development of computer binocular stereo vision related technology has important scientific theory and technical practical significance.*

**KEYWORDS:** *image binocular target vision, monocular image capture, feature extraction, stereo matching, image target visual feature image positioning*

## 1. Introduction

In recent years, the development and research of binocular stereo assembly line vision has gradually begun to attract the attention of scholars and researchers at home and abroad. Generally speaking, binocular stereo vision mainly analyzes and derives the information of the entire three-dimensional space of the scene where the image is located from an external environment by collecting one or more assembly line images. And the basic position and nature of objects existing in the external environment can be obtained by accurately quantified analysis. The technology of binocular stereo assembly line vision has unparalleled technical advantages. It can be seen that it has great academic significance and broad development prospect to apply high-precision binocular stereo vision technology to the automatic identification and three-dimensional grasping technology of assembly line parts.

## 2. Research status at home and abroad

The foreign stereo vision measurement technology has an early origin in China and has many typical applications. Minuasada et al. independently developed a set of advanced moving object adaptive path tracking stereo vision measurement system. The measurement system does not need to calibrate the path of the moving camera

in advance, and uses matlavs theory to accurately predict the adaptive path of the object in motion. Dhavalk.patel et al. independently designed a set of measurement algorithm system for stereo vision, used matlab theory to carry out the comprehensive design of the algorithm, estimated the position of centroid distance of the object to be measured, and used the algorithm to combine the calibration of the camera with the principle of polar triangulation to achieve accurate measurement of the centroid distance of the object. The system design has the advantages of high accuracy and convenient measurement [11]. Micuongcaopham et al. from South Korea conducted in-depth research on the application of stereo matching measurement algorithm in Korean binocular vision. They agreed that the search time and cost consumption requirement with excessive stereo matching are the difficulty and bottleneck in the practical application of stereo matching measurement technology. A new filter cost clustering calculation method realizes the conversion from the traditional 2d filter cost to the one-dimensional filter cost conversion, which reduces the stereo matching from 2d search cost to 1d search, greatly reducing the matching time.

In China, the development of robot depth map based on monocular vision and the calculation technology of target visual positioning have attracted great attention from researchers. Jakobengel et al. proposed a representation and calculation method of robot depth map based on the estimation of image disappearance probability, tracking the robot image sequence to restore the information of the robot in the disappearing area and realize the real-time visual positioning of single target by the robot. But the accuracy of visual estimation on the non-smooth road is not high, because the robot needs to use the camera to take the height away from the ground level of the robot as the positioning parameter. Chen Chen and Jin Hui proposed the feature description of screening scale-invariant feature transformation to realize the positioning of robot monocular vision. Hanwang et al proposed a visual positioning method based on rgb-d, using canny operator to extract the edge of the robot ceiling, and using hof conversion algorithm to detect the straight line of the robot. And iterative icp algorithm can extract the position and posture of robot. But the visual positioning method cannot be widely used in indoor and outdoor occasions where the linear features of the robot cannot be extracted. The research of target identification positioning and grabbing currently is based on the robot servo system of robot camera and system architecture. The robot servo system can be divided into two categories: hand-eye system with camera as sensor and eye-hand system. The hand-eye camera configuration in this system is directly and rigidly mounted on the actuator at the end of the vision robot system. In the eye-hand system, the camera is configured as that a robot and an object that is directly fixed in the sensor workspace for observation. The configuration of one or more cameras in the system is used as the configuration of two cameras or a combination of both. Based on the limitation of sensor testing and application, these configurations and structures can be used for a wide range of visual servo tasks. The results show that the monocular system can use the camera configuration in a system as the "eye-hand" of the global sensor as the bare configuration or as the hand-eye configuration of the sensor.

If a system uses simple hand-eye configuration, it can effectively use feature-based depth tracking estimation technique. It is the shortest processing method and time to extract 3D visual depth information by using a single 3D camera. However, when the pose of three-dimensional object and model is completely unknown, the operation of servo machine and the design of complex computer control system are greatly limited due to the visualization and loss of depth features and information. Stereoscopic 3D visual depth information can be widely used to provide more accurate 3D visual depth information of scenes and 3D objects. One of the commonly used estimation methods is to estimate the parallax of objects and apply it to the estimation of vision and depth. Because the stereo matching system used in the current stereo matching research is based on binocular stereo camera vision. Compared with monocular or stereo camera vision system, the stereo vision system of multi-purpose camera can provide users with more additional information. However, the basic method of stereo matching still requires difficult operations, so the time and process of stereo matching calculation are too complicated to apply directly to real-time system.

Through the investigation and analysis of the development status of key technologies such as binocular stereo vision positioning and robot arm visual servo at home and abroad. On the basis of domestic and foreign research, the subject is based on the in-depth research on key technologies such as binocular stereo vision system, target three-dimensional positioning method, robot arm kinematics and hand-eye calibration system. And a dynamic target positioning system based on binocular vision is designed to realize the task of grasping the target with robot arm combined with the robot arm kinematics.

### **3. Establishment of binocular vision system model**

#### **3.1 Camera model**

##### **3.1.1 Monocular camera model**

The first step in calibrating the parameters of three-dimensional camera is to establish the model of the camera, that is, the three-dimensional mathematical imaging principle and model of the three-dimensional camera. The meaning of stereo mathematical imaging principle process is to restore the three-dimensional image of the objective object and the scene and the object projection information to the plane of the two-dimensional mathematical image. In essence, it is to conduct corresponding geometric transformation for the parameters of four common image pixel coordinate systems, including physical coordinate system in the world, camera coordinate system, image pixel coordinate system and physical coordinate of two-dimensional image. The three-dimensional image projection information obtained by the three-dimensional camera and generated and lost on the two-dimensional image projection is recovered from the three-dimensional mathematical imaging principle model we established, combined with the information calibrated by the camera.

### ***3.2 Establishment of coordinate***

#### **3.2.1 World coordinate system**

As in practical applications, the camera and target object to be measured are placed randomly, and the relative pose relationship between them is unknown, so a reference coordinate system needs to be established to describe the position of the camera in global coordinate system, that is, the world coordinate system ( $x_w, y_w, z_w$ ), and it can be used to describe the positional relationship of any object relative to the camera in the environment. The world coordinate system is often selected according to the specific application situation.

#### **3.2.2 Camera coordinate system**

According to the geometric dynamic relationship between the real object and object image during the imaging process with the camera, a coordinate system that looks similar to the camera can be simply established to accurately depict the relative pose between the camera and the object in the imaging center. The coordinate system is shown in Figure 2-2. The origin of coordinate system is the imaging optical center of the camera, and the optical axis is  $z_c$  axis. The two axes  $x_c$  and  $y_c$  are the two horizontal axes parallel to the image plane of the subject to be filmed and the other vertical axis on the optical axis.  $E$  and  $f$  in the focal length are the relative physical distance between the imaging plane and the optical center of the subject to be filmed.

### ***3.3 Camera distortion model***

The lens distortion of camera is mainly from radial and tangential distortion. They may be non-linear distortion. The radial distortion is mainly caused by the lens deformation of the camera. The main reason of tangential distortion is that the camera is difficult to achieve an ideal installation state where the camera lens is completely parallel to the traditional imager lens due to its manufacturing and process limitation in the installation process.

There is no radial distortion in the photosensitive center. But the closer to the edge, the greater the radial distortion is. For ordinary CCD camera, radial distortion is usually described using two coefficients, where the first term is usually  $k_1$  and the second term is  $k_2$ . For cameras with large radial distortion, such as fisheye camera and panoramic camera, the third radial distortion sparseness can be introduced.

## **4. Conclusion**

The research focus is mainly on the direction that two-dimensional robot is developed to intellectualization. In the process, the research on the collaborative grabbing and operation of vision and robot arms is conducted is paid more attention. The main content of the research includes the following aspects: Through in-depth reading of the literature, the current status and development trend of binocular machine vision of domestic and foreign application research in recent years is

summarized. And the current status of the research of robot target object identification feature positioning and robotic arm grabbing technology is investigated deeply, focusing on the analysis of the importance of binocular vision for robot application and the development history of the theoretical research of robotic arm and two-dimensional vision, as well as related research results of its collaborative grasping operation with the two-dimensional robotic arm. Through the in-depth analysis of the working principle of binocular imaging, a two-dimensional linear robotic arm imaging model based on small-hole robotic arm imaging as the main model is established. The process of calibration for the camera internal parameters is studied, and verification of reliability is carried out for experimental results.

According to the requirements of 2D robotic arm for grasping the target object, the smooth preprocessing of 2D image acquisition is completed, and the method for matching 2D template coordinate based on the feature positioning points of the robotic arm is studied. In the robot arm image, it automatically finds the object specified by the robot arm to grab, and outputs the coordinates in the two-dimensional robot arm image in the image, which provides the position information for the robot to obtain the image of the target object posture in the subsequent three-dimensional image.

## References

- [1] Xia Bo. Research on dynamic target positioning and grabbing based on binocular vision [D]. 2016.
- [2] Zhu Daixian. Research on Workpiece Positioning and Grasping Based on Binocular Vision [J]. Computer Measurement and Control, 2011 (1): 92-94.
- [3] Li Shengli. Research on workpiece recognition and positioning technology based on binocular stereo vision [D]. 2016.