

Building a model of the external environment of a robot based on stereo vision

Alexander N. Goritov

the Department of ACS

Tomsk state University of control system and

Radioelectronics

Tomsk

ang@asu.tusur.ru

Sergei I. Yakovchenko

the Department of ACS

Tomsk state University of control system and

Radioelectronics

Tomsk

sergei.yakovchen@mail.ru

Abstract— In this article, the problems of constructing an external environment of a robot based on stereo vision are considered. Finding ways to solve problems related to building a three-dimensional model of the working environment. The theoretical part describes the basic formulas and definitions of stereoscopic vision. The optical distortions of the camera images and the alignment of images of the stereo system have been eliminated. The problem of classifying objects was solved using the Hough transform. A method is proposed based on the solution of the classification method for constructing a three-dimensional model of the working stage. The research and the results of the obtained method are presented.

Keywords — stereo vision, object classification, disparity map, three-dimensional model.

I. INTRODUCTION

In many problems of automation of operations with the help of robots, it is necessary to have information about the location of objects not only on the plane of the working space, but also in height. This - the manipulation of objects of different heights, the location of objects at different heights. In this case, to automate the manipulation of objects in the working environment, knowledge of the spatial location of objects is required.

The solution of the problem can be accomplished by classifying objects and restoring three-dimensional coordinates of objects based on images from two video cameras (stereo vision). With the help of solving these problems, you can build a model of the working stage. The information about the spatial arrangement of objects in the working stage serves as the basis for the formation of robot control commands.

II. FORMULATION OF THE PROBLEM

To work with complex three-dimensional objects, it is necessary to develop a method for constructing a three-

dimensional model of the external environment of the robot. To solve this problem, we use stereo vision, image processing methods, and image recognition methods.

III. THEORY

A. Stereo vision

The basis of the method for constructing a three-dimensional model of a working stage is a stereoscopic system. The geometry of stereo vision is shown in Fig. 1.

The task of stereovision consists in finding three-dimensional coordinates of objects with the help of two cameras [1].

Let Y be an object located on the working stage (Fig. 1). It is necessary to find the Z -distance to the object. To solve the problem it is necessary to find Z (distance to the object).

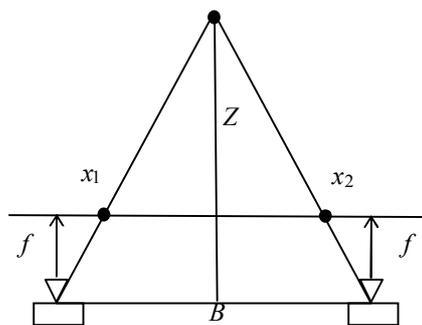


Fig. 1. Geometry of stereo vision

Using the reduction of such triangles, one can arrive at the relation use (1):

$$B / Z = ((B + x_2) - x_1) / (Z - f) \quad (1)$$

Where B it distance between the centers of the cameras, Z - distance to the object, x_1, x_2 - points on the planes of the left and right cameras, f - focal length of the cameras. Substituting these terms, we obtain the relation use (2):

$$x_2 - x_1 = (B \cdot f) / Z. \quad (2)$$

The difference between x_1 and x_2 is called disparity (or OX offset relative to the left image) and is denoted by d . Taking into account the disparity, expression (2) can be rewritten in expression (3). Those to find the distance to the object, it is necessary to know the distance between the centers of the cameras, the focal length of the cameras and disparity.

$$d = x_2 - x_1 \Rightarrow Z = (B \cdot f) / d. \quad (3)$$

In order to find the magnitude of disparity, it is necessary to find the same points on the two images (from the left and right cameras) that belong to the same object in the real world (epipolar points).

B. Elimination of image distortion and rectification

The translation of the real coordinates into the coordinates of the image obtained from the camera is described by the use (4):

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (4)$$

where x, y are the coordinates of the point on the image, f - the focal length, c_x, c_y - the coordinates of the image center, r - the rotation matrix coefficients, X, Y, Z - coordinates of the point in the real world (c_x, c_y is the center of coordinates). But due to optical distortions with the help of use (4), it is impossible to compare the image points to real world points. These distortions are described by nonlinear expressions (5):

$$\begin{aligned} x_{out} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2 (r^2 + 2x^2) \\ y_{out} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 xy + p_1 (r^2 + 2y^2), \end{aligned} \quad (5)$$

Where k_1, k_2, k_3, p_1, p_2 are distortion coefficients, r is the sum of squares of undistorted coordinates. These distortions are not related to the distance to the object. I.e. distortions are only projective. To eliminate them, we usually use a template on which the coordinates in the real world and the projective plane are known in advance. By repeatedly shooting the pattern and determining the coordinates of the distortion, the distortion coefficients are calculated. In addition to distortion of optics

due to the non-ideal installation of cameras, vertical desynchronization occurs, i.e. points on the images of the left and right cameras are not on one line. This leads to even greater difficulties in the search for epipolar points. To eliminate vertical desynchronization, a process called rectification is used.

Rectification is the process of aligning the images of the left and right cameras, so that the epipolar point on the left image is on the same line of the right image. It is necessary to simplify the search for epipolar points. For rectification, it is only necessary to know the position of the right chamber with respect to the left, i.e. the rotation matrix and the coordinate transfer vector. In Fig. 2 shows an example of eliminating distortions and rectification of images.

C. Classification of objects

Classification of objects is the assignment of the original data to a certain class of objects, which most characterizes the original data.

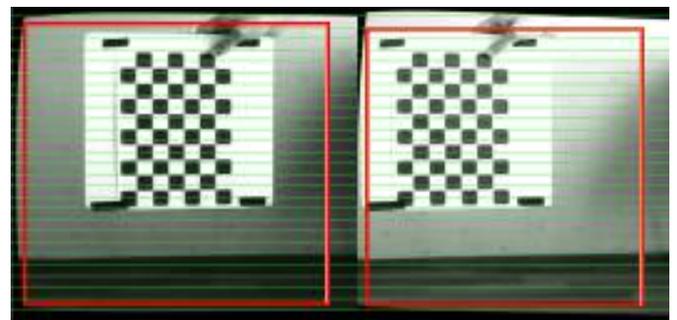


Fig. 2. Elimination of image distortion and rectification

Cameras are mounted above the working stage strictly vertically, so in the image you can see only the top of the objects. For their classification, it is advisable to use algorithms for delineating boundaries, since the contours contain all geometric parameters. But such an approach poses high requirements to the contours of objects - the contours of objects must be closed and the image must be cleaned from noise that worse the results of the algorithm for delimiting boundaries. Therefore, it is necessary to improve the image quality by using noise elimination algorithms. There are a number of noise elimination algorithms. For example, Gaussian blur algorithms, Bokeh effect, depth blur. But since these filters blur the image in every pixel, including contours, such algorithms are not suitable for eliminating noise, as objects lose their outlines. Therefore, it is advisable to use segmentation algorithms. They allow you to divide the image into homogeneous areas and get rid of noise. One of such algorithms is meanshift. In Fig. 3, Fig. 4, Fig. 5 shows a comparison of the results of the algorithm, the selection of boundaries after processing by blur algorithms and the algorithm of segmentation meanshift.

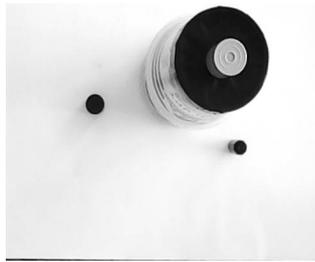


Fig. 3. Source image

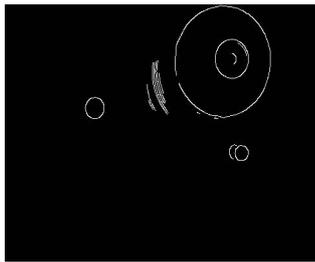


Fig. 4. Edge detection (with blur algorithms)

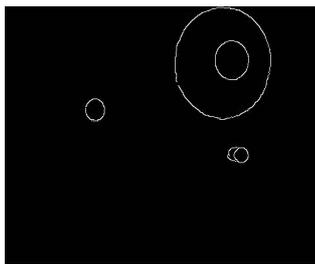


Fig. 5. Edge detection (with meanshift algorithm)

After the image is improved, you can proceed to retrieve the information obtained after the processing of the border allocation algorithm. To search for simple objects, the Hough transformation algorithm was chosen [2]. Target of Hough's transformation consists in finding parametrically given figures, such as: a line, a circle, an ellipse, and so on. After processing the transformation, Hough displays a list of objects with their centers and geometric parameters. In Fig. 6 shows the sequence of image processing for classifying objects of figures by the developed method based on the Hough transformation.

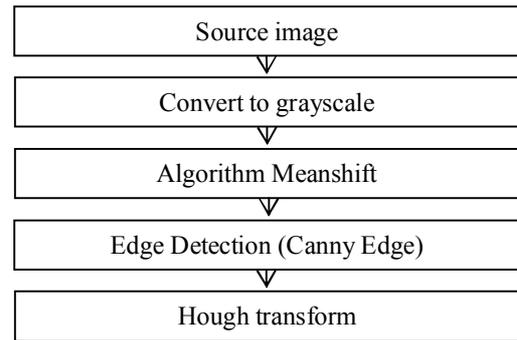


Fig. 6. Method of classification objects

D. Algorithm for calculating disparity based on the Hough transformation

To solve the problem of calculating object disparity, a method based on the classification method was proposed (see Fig.6). For the left and right images of cameras, the classification method is applied. The output of the classification algorithm is two lists. One list for the left image of the camera, the second for the right one. Both lists are sorted by position on the working stage first along the OY axis, then along the OX axis, also sorted by object type and their geometric parameters. Next, sorting by geometric parameters, such as radius (if it's a circle), height and width (if it's a rectangle). In the loop, we search for paired elements for each object. After the new list is formed, in which the elements are references to the same object in the left image and the same object on the right image, disparity values between the objects are calculated. Subtract the position value along the x axis of both objects. In Fig. 7 shows a diagram of the algorithm for calculating disparity based on the Hough transform.

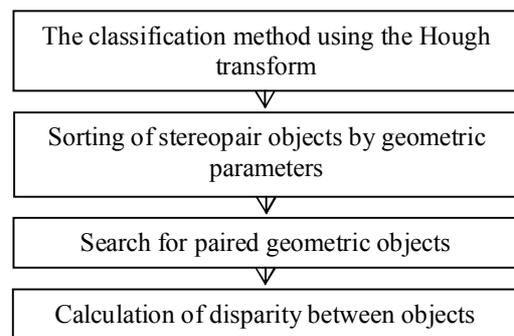


Fig.7. Algorithm for calculating disparity based on the Hough transformation

IV. RESULTS AND EXPERIMENTS

To evaluate the effectiveness of the proposed method, a program was written in the C ++ programming language. At the input of the image processing program, two source images of Fig. 8 and Fig. 9 (for the left and right cameras, respectively).

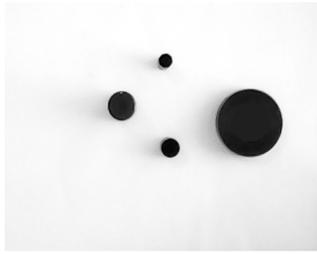


Fig. 8. Source image of left camera

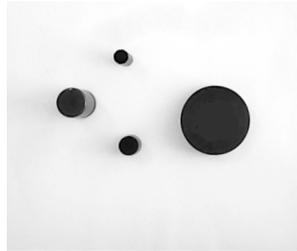


Fig. 9. Source image of right camera

After the images were taken, a method for classifying and selecting objects is performed for each image. In Fig. 10 and Fig. 11 shows the results of the algorithm for classifying and selecting objects.



Fig. 10. The result of classification and selection of objects for the left camera



Fig. 11. The result of classification and selection of objects for the right camera

To calculate the distance to objects, the developed method (see Fig. 7) is performed based on the classification results. Knowing the distance to the working stage (650 mm), types and heights of objects, a three-dimensional model of the working stage is built.



Fig. 12. Three-dimensional model of the working stage

The resulting three-dimensional model of the working stage is shown in Fig. 12.

To visualize the constructed three-dimensional model of the working stage, the graphic editor MeshLab is used. For this purpose, the data was exported to the .xyz format (a list of three-dimensional points) of the graphical information representation of the MeshLab editor.

V. THE DISCUSSION OF THE RESULTS

Summarizing the results of numerical experiments, it can be noted that the implemented method showed a high level of recognition and acceptable accuracy in determining the distance to objects. It can be noted that the use of short-focus cameras is not recommended for distance determination. As this leads to perspective distortions increasing at different distances and viewing angles to the object.

VI. CONCLUSIONS AND CONCLUSION

As a result of the work done, a method for constructing a three-dimensional image of the robot's working stage is proposed. The objects located on the working stage of the robot are selected and classified. The distances to the selected objects are determined.

The constructed three-dimensional image of the working stage can be used to generate robot control commands.

REFERENCES

- [1] R. Sveliski, "Computer Vision: Algorithms and Applications", Springer-Verlag London, 2011, 812 p.
- [2] R.Gonzalez, R.Woods, "Digital image processing", Technosphere: Moscow, 2005, 1072 p.