

Palinko Oskar, dipl. eng.

Faculty of Technical Sciences, Department of Industrial Systems Engineering and Management,  
21 000 Novi Sad, Dositej Obradovic Square 6, Serbia & Montenegro

## STEREOSCOPIC ROBOT VISION SYSTEM

### Abstract

The visual system is one of the most important sensors in robotics. It is used by the robot to acquire information about the world, to be able to navigate in it. A stereoscopic system allows the robot to easily determine the distance of the objects in its vicinity. This paper discusses a stereoscopic system, which allows the robot to locate a flight of stairs and to determine its orientation. The main steps in solving this task are as follows: edge detection is applied on the stereo image; then the analytic form of these edges are calculated using Hough transformation; stereo matching is done using the starting points of the edges; the distance of the stairs is derived using triangulation. Finally the orientation of the stairs is calculated using geometric equations.

**Keywords:** machine vision, robotics

## 1. INTRODUCTORY REMARKS

Robotics in the modern world is gaining more and more significance. The field of robotics, which aims to create human-like robots is called humanoid robotics.

Robot vision is a very broad scientific field. Some of its interests are: visual servoing, pattern recognition, stereo vision, etc. Stereo vision is one of the more important parts of robotics, because of its significance for moving in an unstructured environment.

This paper introduces a stereo system designed for use in anthropomorphic robots as for navigation in an unknown environment. The emphasis was set on detecting and analyzing simple, prismatic objects like: stairs, holes, prismatic obstacles, etc. Special attention was devoted to scenes containing stairs, which the robot analyses and then approaches.

A virtual simulation environment is programmed for testing the designed algorithms in every phase of development. Finally the system is tested on a small-scale mobile robot in a real-life human environment.

## 2. ELEMENTS OF MACHINE VISION

This chapter introduces some of the basic elements of machine vision, which are of use later in this paper, like edge detection, feature extraction and triangulation.

### 2.1 Edge detection

Edge detection is an image processing method. At the places, where a function has an intense inclination, the first derivative will have a local extreme value [5]. In the case of 2D images, the mathematical operator gradient is used. It is a two dimensional derivative which is directed towards the biggest rate of change in the vicinity of the point considered.

So as to detect edges, local extreme values of the gradient must be found. In the case of digital signals, the gradient is approximated with the following equations:

$$\frac{\partial f(x, y)}{\partial x} = \Delta_x = \frac{f(x + d_x, y) - f(x, y)}{d_x}, \quad (1.1)$$

$$\frac{\partial f(x, y)}{\partial y} = \Delta_y = \frac{f(x, y + d_y) - f(x, y)}{d_y}, \quad (1.2)$$

where  $d_x$  and  $d_y$  is the horizontal and vertical distance between two adjacent samples (usually equal to 1). The magnitude and orientation of the gradient can be expressed as:

$$M = \sqrt{\Delta_x^2 + \Delta_y^2}, \quad \theta = \arctan\left(\frac{\Delta_y}{\Delta_x}\right). \quad (1.3)$$

#### 2.1.1 The Canny method

This algorithm, in addition to the gradient calculation, also contains other steps for improving the results of the detection. Its distinguishing marks are the two threshold values [6], which are to be explained in the following.

The algorithm can be divided into 6 steps:

1. In the first, a digital Gauss filter is applied to the image as to suppress the possible noise
2. After the elimination of noise, a 2D gradient is used with extended convolution matrices:

$$\Delta_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad \Delta_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

3. The orientation of the gradient is calculated using the formula (1.3).

4. In the fourth step the orientation of the gradient of each point is classified into one of 4 groups. For example, lines with orientation between 0 and 22.5 degrees and the ones between 157.5 and 180 are assigned into the class of 0°. In this way, the exact value of the orientation is substituted with classes.

5. Next, the non-maximum edge points are suppressed. In this step the edge line is followed by the class information. Every edge point that is not in the orientation of the previous points class, gets eliminated. Only those points are left which have corresponding orientation class.

6. In the last step, the continuity of the edge line is assured. Two thresholds are introduced,  $p1$  and  $p2$ , where  $p1 > p2$ . All the points on the edge, which have intensity larger than  $p1$  are automatically confirmed. If its intensity is less than  $p1$  but more than  $p2$  then the point is going to be confirmed only if it has a confirmed point in an adjacent square. Otherwise it will be deleted.

This last step is the main innovation of the Canny method, which makes it very effective. In this paper, this method will be used, because it gives narrow (5th step) and continuous (6th step) edges.

## 2.2 Hough transformation

This transformation belongs to the group of feature extraction methods. It is intended to extract regular features [4] from the images, e.g. lines, circles, ellipses, etc. All of these forms can be expressed analytically (i.e. through an equation).

In this paper the Hough transformation is used to extract analytic expressions of lines. Because of this, the methods subtype designed for lines will be explained in the following. A line can be parameterized in the x-y plane like:

$$x \cos \theta + y \sin \theta = r, \quad (2.1)$$

where  $r$  is the length of the normal of the line, which normal intersects the coordinate origin (0,0), while  $\theta$  is the angle between the normal and the  $x$  axis. For any point on a particular line, the values of  $r$  and  $\theta$  are constants.

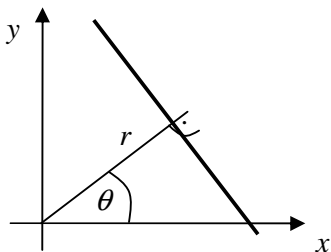


Figure 2.1 Parametric description of a line

In this way, Hough transformation is a projection from the  $(x, y)$  space into  $(r, \theta)$  space. Equation (2.1) shows that points in the first space are really sinusoidal curves in the

second space. If Hough transformation is applied on an image derived by edge detection (a binary image), then the brightest points in the  $(r, \theta)$  plain will be the ones corresponding to the straight edge lines [4]. Finding local maximal points will yield in edge detected line equations. The familiar form of an equation  $y = ax + b$  is gotten by expressing the parameters in the following way:

$$a = -\frac{\cos \theta}{\sin \theta}, \quad b = \frac{r}{\sin \theta}. \quad (2.2)$$

## 2.3 Stereoscopy - triangulation

Stereoscopy is a way of seeing objects in 3D. The goal of it is to be able to determine depth of view, object distance, object proportions, etc. in the scene. Triangulation is a well-known method of calculating the distance of objects knowing the angles under which the object is seen from at least two positions and knowing the distance between those two positions [3]. In this way triangulation is closely related to stereoscopy, because it gives information about distance, one of the most important elements of stereoscopy.

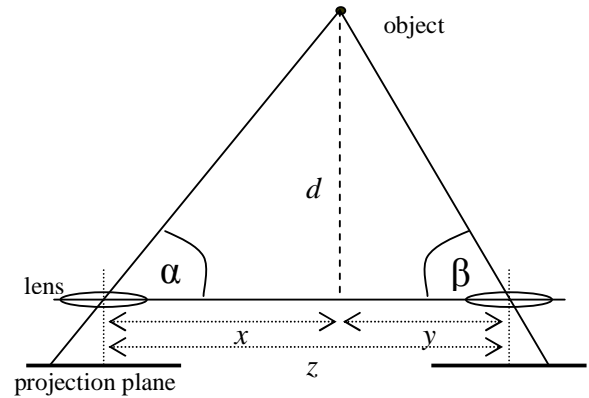


Figure 2.2 Triangulation using two cameras

Knowing the angles  $\alpha$  and  $\beta$  as well as the distance  $z$  between the two cameras, the next equation can be derived to give the distance of the object:

$$d = z \frac{\operatorname{tg} \alpha \cdot \operatorname{tg} \beta}{\operatorname{tg} \alpha + \operatorname{tg} \beta}. \quad (2.3)$$

## 3. SIMULATION SYSTEM – VIRTUAL SCENE

The virtual scene is a 3D environment, which exists only as a software simulation. It contains virtual objects, lights and virtual cameras. Cameras are used to get 2D images of the scene, just like in real world.

In the following, the main steps will be given to explain how the simulation works. The first step is the acquisition of images from two stereo cameras, which are then processed and analyzed as get information on objects that are present in the scene. Under analysis, we mean the determination of distance of objects and their orientation compared to the camera. Finally the virtual robot approaches the object (stairs) so as to be parallel with the

front edge of it. It is important to emphasize, that the robot knows the geometry of the system only through the camera images.

### 3.1 Virtual cameras

Knowing the coordinates of a point on the camera image, it is easy to determine the angle under which it is seen by the camera:

$$\alpha = \arctan\left(\frac{\tan(\alpha_{\max}/2) \cdot x}{x_{\max}/2}\right),$$

where  $\alpha_{\max}$  is the width of the field of view,  $x_{\max}$  is the resolution of the camera. In this way, the necessary angles are gotten for triangulation.

The cameras are positioned in a so-called canonical configuration. That means that their optical axes are parallel, the projection surfaces are in the same plane and their upper edges belong to the same line. In this way, the stereo pairing of a point on the stereo images is done on the same horizontal line.

### 3.3 Image processing

The first step after acquisition is the transformation of color into intensity (grayscale) images, because the following algorithms can work only on such pictures. Then the Canny detection is invoked (described in 2.1.1). The binary pictures (black-white) of the detected edges are sent to the blocks for Hough transformation, which extracts parametric information on the lines present in the scene (as described in 2.2). The result of this method is a continuous 2D grayscale image from which the local extremes must be extracted. As explained in 2.2 the maximum points represent the straight lines in the x-y plane. In this way, the parametric line equations are gotten.

### 3.4 Analysis and reasoning

The analysis begins with searching for the beginning and ending points of the parametric lines on the left-side stereo image. It is done in the following way: the analytic lines are followed until a discontinuity in the edge is reached. If it is a discontinuity from black to white dot, then the edge begins, otherwise it ends. These points are then stereo paired with the right-side image. Pairing search is done on the same horizontal line as in which the point lies on the left image. Stereo pairing is a demanding process, that uses 2D correlation calculation for each point of the line [1]. The point with the best correlation result will be declared as the right pair of the point on the left image. Correlation usually gives good results, because the images are quite similar due to the small parallaxes.

#### Calculating the angle $\theta$

The algorithm for calculating the angle  $\theta$  entirely based on geometric equations. No approximations were used.

Equation 2.2 shows how to calculate the angles in the field of view of the camera knowing its position on the camera image. This calculation is valid for both x and y-axis.

In order for the robot to be able to approach the stairs, it must know what is the angle between him and the object in the horizontal plane. This angle is designated  $\theta$  which must be found knowing only the angles in the image ( $\alpha$ ), ( $\beta$ ) and also the inclination and height of the camera.

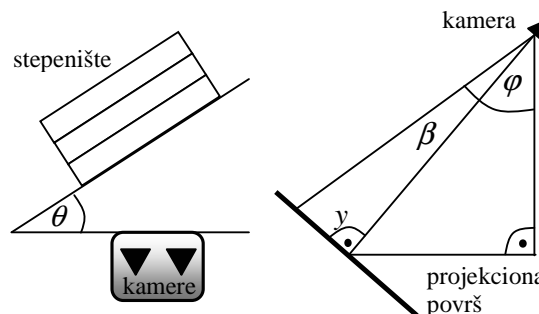


Figure 3.1 Angles in a) horizontal and b) vertical plane

After the deduction of a series of geometric equations, we get:

$$\theta = \arctan\left(\tan \alpha \frac{\cos \beta}{\cos \varphi}\right), \quad (3.1)$$

where  $\alpha$  is the horizontal angle in the picture,  $\beta$  is the vertical angle in the picture and  $\varphi$  is the camera inclination.

#### The phenomenon of inclined edges

Using perspective projection, an interesting effect can be noticed: near the border of the image the lines that are in real-life parallel to the border, appear to be inclined. It doesn't happen with the lines going through the center of the image. The closer the line is to the border, the phenomenon is more emphasized.

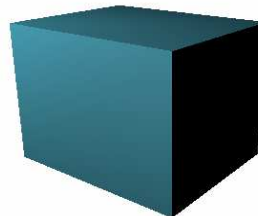


Figure 3.2 The phenomenon of inclined lines

This effect appears when the lines are not parallel with the projection surface, because then some points on the line are situated closer to the surface than others. I.e. when the camera is inclined forward, the upper part of the projection plane is getting closer to the scene while the lower part is getting more distant. In that case when a ray of light comes from one of the upper corners, inclining the camera will cause the light to move up and away from the center. This deviation can be corrected with the following equation:

$$\alpha = \arctan \frac{x}{\sqrt{\left(\frac{x_{\max}/2}{\tan(\alpha_{\max}/2)}\right)^2 + y^2}}. \quad (3.2)$$

### Calculating the distance of objects

With the basic method of triangulation the right-angle distance of the object is gotten.

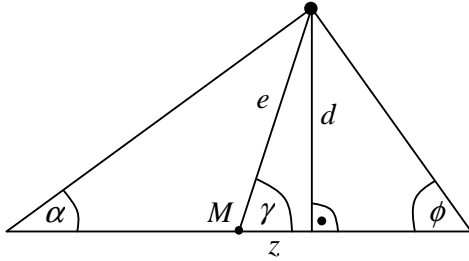


Figure 3.3 Geometry in the triangulation plane

But in this work the distance of the object from the central point  $M$  is needed:

$$e = \frac{d}{\sin \left( \arctan \frac{d}{z/2 - \frac{d}{\tan \phi}} \right)}, \quad (3.3)$$

where  $d$  is gotten from equation (2.3) and the other elements are explained on figure 3.3. The calculated distance must be projected on the x-y plane. Knowing the angle of inclination of the camera  $\beta$ , the task is trivial:

$$e_{xy} = e \sin \beta. \quad (3.4)$$

Finally the projection of the angle  $\gamma$  on the x-y plane must be expressed:

$$\tan \lambda = \frac{\tan \gamma}{\sin \beta}. \quad (3.5)$$

### 3.5 Simulation of the robot's actions

To perform the action of approaching the stairs the following information is needed:

- the angle  $\theta$  in the horizontal plane between the robot and the object
- the projection (on the x-y plane) of the distance of the object from the central point,  $e_{xy}$
- the angle  $\lambda$

It is enough to know only these elements so as the robot can perform its actions in case a flight of stairs is in front of it. The robot must approach it so, that the front edge of the stairs must be parallel with the line connection the two cameras and every time the robot must be on a constant distance. The starting position is in the general case as follows:

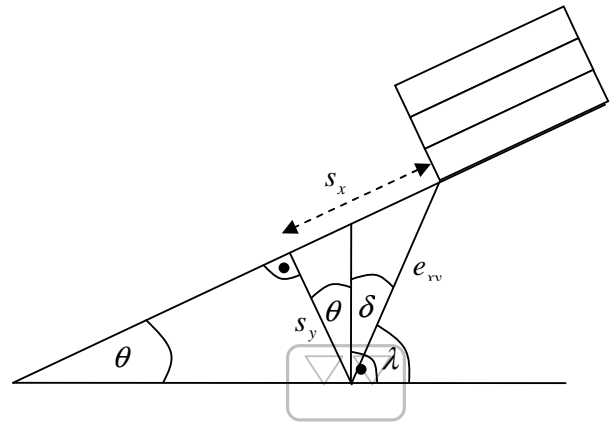


Figure 3.4 Calculation of the operational lengths

The operational lengths  $s_x$  and  $s_y$  for approach are calculated as follows from the figure:

$$s_x = e_{xy} \sin(\theta + 90 - \lambda) \quad (3.6)$$

$$s_y = e_{xy} \cos(\theta + 90 - \lambda) \quad (3.7)$$

Knowing them, the robot can perform its actions. These can be divided into the following steps:

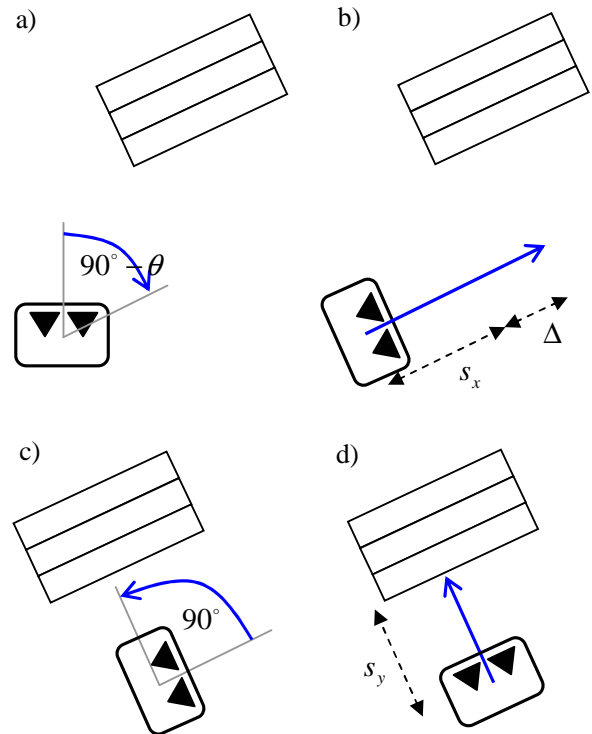


Figure 3.5 Robot actions

In step a) the robot turns in an angle of  $90^\circ - \theta$  so as the optical axes will be parallel to the front edge of the stairs. Then, in step b) the robot is moving straight forward. The

distance it should prevail is equal to the sum of  $s_x$  and some value,  $\Delta$ . This value is added so the robot doesn't approach exactly the left edge of the stairs, but somewhere in the middle. Step c) is turning back in a right angle. This way, the robot is parallel to the front edge of the stairs. Finally the last step d) is performed in which the robot passes the straight distance of  $s_y$ .

#### 4. EXPERIMENTAL VALIDATION

The goal of experimental realization of the visual system is to check the theoretical and simulation algorithm in real-life situation. The experimental system consists of a mobile robot, that analyzes the scene and then approaches the stairs. It is not a walking biped (as the goal platform) but a wheeled robot. The reason for this is of course that, the biped is not produced yet, but the validation must be performed.

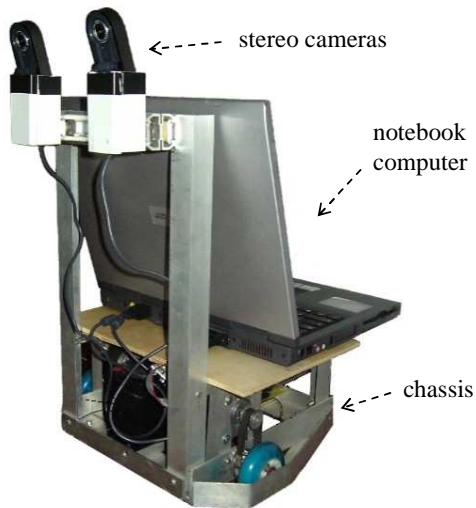


Figure 4.1 Mobile robot

The main parts of the robot are:

- the stereo cameras – which are the most essential part of the system; two Logitech's QuickCam for Notebooks Pro web cameras were selected for the task with a field of view of  $30^\circ \times 40^\circ$  and a resolution of  $640 \times 480$  pixels; the cameras must be mounted in the canonical way
- a notebook computer which performs all the signal processing; an Acer Aspire 1312 is used with AMD Athlon 2000+ 1.66 GHz processor and with 256Mb of memory
- electronic circuits – they are comprised of a AT89C52 controller (with its environment and RS232 communication with the PC) and a driver electronics board
- motors – step motors are used for greater precision.

The analysis of the scene is performed exactly like in the simulation. A test image and the results of the analysis are shown in the following:



Figure 4.2 Camera image of the real-life stair model

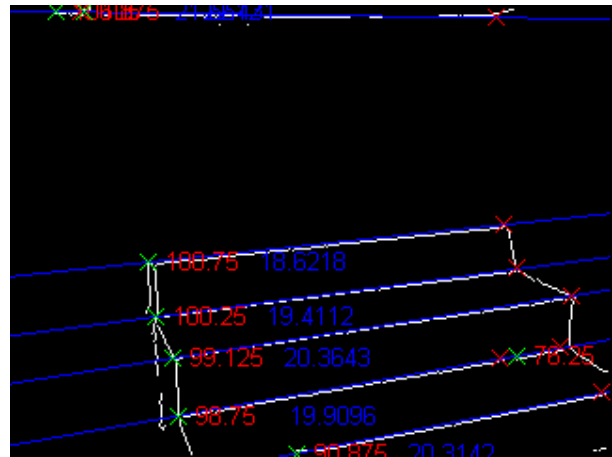


Figure 4.3 Detected edges, Hough lines, distances, angles

The experimental results of determining the distance of an object based on the above gotten pictures gave the following results:

$e_{racun}$ [cm]	$e_{nom}$ [cm]	$\Delta e$ [cm]	$\delta e$ [%]
97,43	97,1	0,33	0,40
119,39	120,5	1,11	0,92
89,31	88,5	0,81	0,91
88,93	88,4	0,53	0,60
91,12	90,8	0,32	0,35
90,84	89,0	1,8	2,02
101,46	99,5	1,96	1,97
102,88	101,2	1,68	1,66
78,21	76,0	2,21	2,91
73,72	71,8	1,92	2,67

Table 4.1 Distance measurement results

In further testing, eight cases of robot approach were conducted. The results were as follows: the robot succeeded in six cases to precisely position itself in front of the stairs, while in two cases it didn't. The errors were in tests number 4. and 6. Analysing the errors it made, it was concluded, that in the first case the stereo matching system did not succeed in its task due to high noise in the images, while the other fault happened because an imperfection in the electro-mechanical part of the mobile robot.

## 5. CONCLUSION

This paper demonstrated that the proposed stereoscopic visual system is a viable solution of the required task. It was shown, that the problem of approaching a staircase can be solved solely with the use of the stereo images and knowing only the camera inclination and height. The deduced geometrical equations in combination with some well-known image processing methods (like Canny edge detection, Hough transformation, etc.) are well suited for solving problems of this type.

## 6. DIRECTIONS OF NEXT RESEARCHES

The presented visual system was projected for work with prismatic objects. In further development the existing algorithm should be generalized to be able to work with objects that don't have straight edges. After that, when the robot would be able to recognize a large range of different obstacles, a cognitive system should be developed using artificial intelligence. That would mean that the robot would be able to classify and learn to recognize new, unfamiliar types of objects.

## 7. REFERENCES

- [1] M. Sonka, V. Hlavac, R. Boyle, "*Image Processing, Analysis and Machine Vision*", Brooks and Cole Publishing, Iowa City, 1998.
- [2] D. Ballard, C. Brown, "*Computer Vision*", Prentice-Hall Inc., Englewood Cliffs, 1982.
- [3] A. Marshall, "*Vision Systems*", [http://www.cs.cf.ac.uk/Dave/Vision\\_lecture](http://www.cs.cf.ac.uk/Dave/Vision_lecture), last access: 2.9.2004
- [4] R. Fisher, "*Hough Transform*", [http://www.dai.ed.ac.uk/HIPR2/hipr\\_top.htm](http://www.dai.ed.ac.uk/HIPR2/hipr_top.htm), last access: 2.9.2004.
- [5] B. Green, "*Edge Detection Tutorial*", <http://www.pages.drexel.edu/~weg22/edge.html>, last access: 9.11.2004.
- [6] B. Green, "*Canny Edge Detection Tutorial*", [http://www.pages.drexel.edu/~weg22/can\\_tut.html](http://www.pages.drexel.edu/~weg22/can_tut.html), last access: 9.11.2004.