STEREO VISION SENSOR FOR 3D MEASUREMENTS

A complete solution to produce, calibrate and verify the accuracy of the measurements results

Liviu Toma, Fangwu Shu, Werner Neddermeyer
Informatics Department, University of Applied Sciences Gelsenkirchen, Germany

Alimpie Ignea
Optical Electronics Department, University "Politehnica" Timisoara, Romania

Keywords: calibration, camera model, correspondence problem, distortion, stereo-vision, sub-pixel accuracy

Abstract: The goal of this paper is to build a stereo sensor to be used as a 3D measurement tool with direct application in automotive industry. The distance between the object to be measured and the stereo sensor is between 200 mm and 300 mm. This paper presents the solutions developed in order to produce, calibrate and verify a stereo sensor used to measure 3D coordinates with an accuracy of 0.1 mm. The measurement area is defined by a square with a side of 100 mm. The contribution of this paper to the extant literature is twofold. First, it presents a new method to compute the coefficient of the radial distortion. Second, it develops an image-processing algorithm, in order to minimize the errors that occur from the non-correspondence problem. The most important issues that need to be addressed are the following: defining a camera model in order to best simulate a real camera, and identifying the same point with both cameras of the stereo sensor (correspondence problem), in order to reduce the measurement errors.

1 INTRODUCTION

The camera calibration problem has been extensively studied over the past 25 years. The extant literature addressing this topic can be divided into two categories: the calibration of zooming and rotating cameras (Agapino, 2001) and the calibration of fixed cameras (Armangue, 2000). The calibration developed and presented in this paper belongs to the latter category.

Armangue, Salvi and Balle (2000) is a very good survey of existing calibration methods of fixed cameras. According to this study, there are four calibration methods: the method of Hall (Hall, 1982), the method of Faugeras-Toscani (Faugeras, 1986), the method of Tsai (Tsai, 1987) and the method of Weng (Weng, 1992). We used the method of Tsai and the method of Faugeras-Toscani, as main references in our work.

The remainder of this paper is organized as follows. Section 2 presents the camera model we consider. Section 3 describes the image segmentation alghorithm. We present the calibration procedure that we developed in Section 4, and the measurement procedure in Section 5. The analysis of the errors obtained with our sensor is presented in Section 6, while Section 7 briefly describes two possible industrial applications of the stereo sensor. Section 8 concludes.

2 CAMERA MODEL

Our goal is to find that set of parameters which best simulates the behavior of a real camera. Generally, the camera parameters are divided in two categories: extrinsic parameters and intrinsic parameters (Faugeras, 1993).

There are six extrinsic camera parameters. We denote them t_x, t_y, t_z, α, β, γ. The first three give the position and the last three the orientation of the camera frame with respect to a reference frame or a world frame. In our case we call this reference frame the stereo sensor frame. The position of the stereo sensor frame is in the middle of the calibration plate and the orientation is as one can see in Figure 1.

Toma L., Shu F., Neddermeyer W. and Ignea A. (2004).
STEREO VISION SENSOR FOR 3D MEASUREMENTS - A complete solution to produce, calibrate and verify the accuracy of the measurements results.
In *Proceedings of the First International Conference on Informatics in Control, Automation and Robotics*, pages 410-416
DOI: 10.5220/0001138204100416

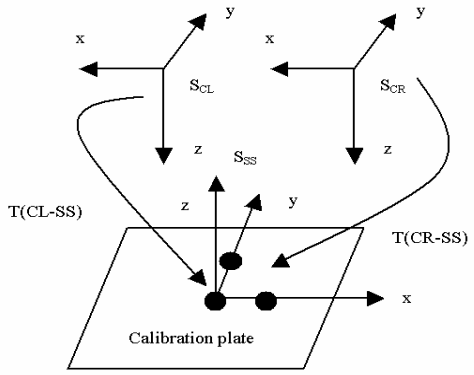

Figure 1: Stereo sensor frames

More details about the calibration plate are provided in Section 4. Axes *x* and *y* are in the same plane with the calibration plate and axis *z* is orthogonal to this plane.

Things become a little more complicated with respect to camera intrinsic parameters. The simplest model is the pinhole model. This is a distortion-free model, which includes four independent parameters: $s_x f$, $s_y f$, C_x, and C_y, where we denote the focal length by *f*, the scale factors by s_x and s_y and the center of the image (the intersection of the optical axis with the CCD chip plane) by C_x and C_y. A better simulation of a real camera is given by a model which includes the radial distortion. We denote the coefficient of the radial distortion by *k*. There are camera models which also include other types of distortions, such as decentering and thin prism distortion (Weng, 1992). Theoretically, we should also consider the skew factor (Faugeras, 1993). The skew factor is a function of the angle between the axes defined by two adjacent sides of the CCD chip. Normally, this angle is 90 degrees, and in this case the skew factor will have no influence on the perspective matrix. Other intrinsic parameters can be introduced to model the fact that the optical axis is not orthogonal to the CCD chip. This is one of the next problems to address in our future work.

In order to reach the required accuracy (see abstract), we consider a model that includes the four standard intrinsic parameters $s_x f$, $s_y f$, C_x, and C_y and the coefficient of the radial distortion *k*. Because of technological progresses in manufacturing lenses and CCD chips, the effects of distortions, other than the radial distortion, and that of the skew factor are very small.

We present below a short description of the mathematical model considered for the radial lens distortion. There are two types of radial distortion, one positive, called pincushion distortion, and one negative, called barrel distortion (Landsberg, 1958). The lenses we have used are affected by barrel distortion.

We consider two points, P_u which is the ideal point, undistorted, and P_d, which is the real point, distorted. The coordinates of these points are X_u, Y_u and X_d, Y_d, respectively. We approximate the distortion by the following relations

$$X_u = X_d + X_d \cdot f(R_d), \quad (1)$$

$$Y_u = Y_d + Y_d \cdot f(R_d), \quad (2)$$

where R_d is defined by the following relation

$$R_d = \left((X_d)^2 + (Y_d)^2\right)^{\frac{1}{2}}. \quad (3)$$

According to the literature for this radial distortion, only the second order term has a significant value (Weng, 1992).

We can then approximate the function *f* by the following expression

$$f(R_d) = k \cdot R_d^2, \quad (4)$$

where *k* is the coefficient of radial distortion, as described at the beginning of this section. Using Eqs. (1) – (4), we obtain, after some mathematical computations

$$X_u = X_d \cdot \frac{1}{1 + k \cdot R_d^2}, \quad (5)$$

$$Y_u = Y_d \cdot \frac{1}{1 + k \cdot R_d^2}. \quad (6)$$

We use these two relations in our further calculations to model the radial distortion.

3 IMAGE PROCESSING ALGORITHM

The first problem was to decide what type of marks one can use in order to identify them in the pictures obtained from the CCD cameras. One solution was to use crosses. The other was to use circles. We decided to use circles and the reasons why we did so are presented further in the section.

The accuracy of the information obtained from the marks is directly dependent to the accuracy of the detected edge points. If we take a circle having a radius *r*, the total length of the edges is given by

$$L_{Circle} = 2\pi * r \,. \tag{7}$$

For a cross, with the dimensions *2r* horizontal and *2r* vertical, the total length of the edges is given by

$$L_{Cross} = 8 * r \,. \tag{8}$$

Based on these two relations, we show that the number of the edge points for a circle is smaller than the number of the edge points for the corresponding cross. Thus, we have to detect more edge points in the case when a cross is used than in the case of a circle. Each one of the edge points is detected with an error and affects the information used in our further calculations. Therefore, more edge points means more errors and this ultimately will have a higher effect on the amount of useful information.

The next problem was to find a segmentation method, by which to decide which pixels belong to the circle and which belong to the background. Initially, we worked at the pixel level using segmentation methods based on dynamic thresholds (Parker, 1997, and Gui, 1999). The results are better than when we use only the functions from the image processing software. However, an alternative method, which consists in working at the sub-pixel level, can further improve these results. We explain further in more details why we chose this method, and present the algorithm we developed.

Consider the following example based on a real situation: a simple plate, which is half white and half black, as in Figure 2.a1. If we take the image of this plate through a camera, and store it on the chip, it will be a little distorted, as one can see in Figure 2.a4. We consider that the transfer from the chip image to the computer image takes place without errors. Therefore, the situation presented in Figure 2.a4 is also valid at the pixel level.

In most cases, the border between an object and the background in the pixel image should be situated on the surface of one pixel, and not at the border between two pixels. Due to physical considerations, we cannot have two different levels of electricity in one cell of the chip. Furthermore, the corresponding pixel cannot have two grey levels. We develop a mathematical algorithm, which determines a sub-pixel value associated with the location of the border between two grey levels.

Our goal is to reach an accuracy of a tenth of a pixel. To obtain that level of precision, we have to explore each “circle” in the following way: we start from the weight point of the “circle” with lines to the edges of the “circle”. One thing to be mentioned here, the term “circle” denotes the calibration mark, which by projection to the image becomes an ellipse. There are two problems that must be solved. The first one is to decide how many lines to use. The second one is to compute the grey level in certain sub-pixel positions situated on this line.

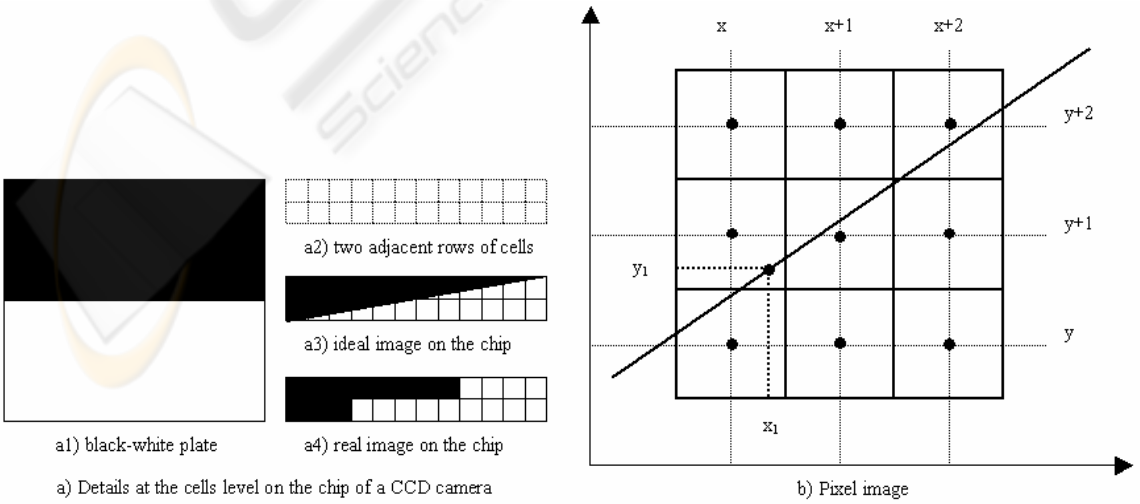

Figure 2: Details concerning the sub-pixel resolution

The number of lines to use is determined by the value of the angle between two consecutive lines. The length of the circle is computed using the following equation

$$L_C = 2\pi \cdot R \cdot \Delta p\,, \tag{9}$$

where Δp is the length associated with one side of a square pixel. The angle between two consecutive exploration lines will be computed using the following relation

$$\Delta\alpha = \frac{180}{\pi \cdot R \cdot n}[grad]. \tag{10}$$

where n is the number of parts in which we want to divide a pixel. In our case, we take n equal to 10 and the maximum value for the radius, R, equal to 20. This way, we obtain a value of 0.28 for $\Delta\alpha$, suggesting that we use approximately 1285 exploration lines.

For each of these lines we analyze a part of it with a length equal to the length of 5 pixels. The middle of this part is situated at a distance equal to the circle radius R. Between the Cartesian coordinates of one point situated in this part of the line and the polar coordinates of the same point, the following two relations can be written

$$x = C_x + (R+d)\cdot\cos\alpha\,, \tag{11}$$

$$y = C_y + (R+d)\cdot\sin\alpha\,, \tag{12}$$

where d takes values between –2.5 and +2.5. The difference between two consecutive values of d is 0.1. C_x and C_y are the coordinates of the "circle" weight point. As indicated before, our goal is to have the grey level of the points situated at any location d on the exploration line. Using Eqs. (11) and (12), we compute the corresponding coordinates (x, y) for each of the 51 values of d. The problem now becomes that these coordinates (x, y) have float values and we only know the grey level for those with integer values.

Next, we present a solution for computing the grey level of a point whose coordinates take float values. In Figure 2.b, we show a square formed by nine pixels.

The values taken by x any y are positive integers. They represent the location of the pixel in the image. With small circles we have denoted the grey levels of the pixels, and placed them in the middle of their corresponding pixels. We are next interested in computing the grey level of the point situated at the location (x_1,y_1), as one can see in Figure 2b. The following notations are made in order to simplify the calculations

$$\Delta x = x_1 - x\,, \tag{13}$$

$$\Delta y = y_1 - y\,. \tag{14}$$

Next, the grey level of the point situated at location (x_1, y_1) can be calculated using the relation

$$\begin{aligned} g(x_1,y_1) = {} & g(x,y)\cdot(1-\Delta x)\cdot(1-\Delta y) + \\ & + g(x,y+1)\cdot(1-\Delta x)\cdot\Delta y + \\ & + g(x+1,y+1)\cdot\Delta x\cdot\Delta y + \\ & + g(x+1,y)\cdot\Delta x\cdot(1-\Delta y) \end{aligned}\,. \tag{15}$$

Following the algorithm described before, we generate 51 pairs (d_i, $g(d_i)$), where d_i takes values between –2.5 and +2.5, and i takes values between 0 and 50.

In order to simplify the mathematical calculations and avoid working with float numbers, we define a variable D as follows:

$$D = 10\cdot d - 25\,. \tag{16}$$

where D takes integer values between –25 and +25. Using the relation

$$G(D) = g\left(\frac{D-25}{10}\right) = g(d) = g(x,y) \tag{17}$$

one can next compute the function $G(D)$.

So far, we managed to divide an interval of five pixels in fifty sub-pixels intervals, and to compute for each sub-pixel position the corresponding grey level.

Next, our approach is to find a mathematical relation, which best approximates the function $G(D)$. The following relation defines this function:

$$G(D) = G_0 + K_G\cdot\arctan\left(K_D\cdot(D-D_0)\right). \tag{18}$$

Our final goal is to compute D_0. Rewriting the Eq. 18 as follows

$$F_{D_i,G_i}(D_0,G_0,K_D,K_G) = 0\,, \tag{19}$$

where D_i and G_i are those variables calculated in the previous steps, we will obtain an over determined system of nonlinear equations. To solve this system, we first use the Newton algorithm, to make the system linear, and then least square methods (Manusar, 1981, Naslau, 1999).

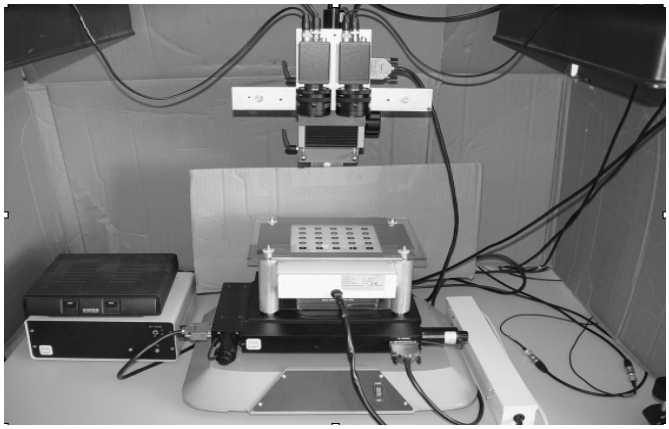

Figure 4: The stereo sensor and the calibration device

4 CALIBRATION OF THE STEREO SENSOR

In this section, we show the relation between the 3D coordinates and the 2D pixel coordinates of a calibration point and the camera parameters. We start from the equations

$$\frac{1}{1+k\left(\frac{(X_p-C_x)^2}{s_x^2}+\frac{(Y_p-C_y)^2}{s_y^2}\right)}\frac{X_p-C_x}{s_x}=f\frac{x}{z}, \quad (20)$$

$$\frac{1}{1+k\left(\frac{(X_p-C_x)^2}{s_x^2}+\frac{(Y_p-C_y)^2}{s_y^2}\right)}\frac{Y_p-C_y}{s_y}=f\frac{y}{z}, \quad (21)$$

where (X_p, Y_p) are the pixel coordinates and (x, y, z) are the 3D coordinates of a calibration point with respect to the camera frame. The following notations are made

$$p_x = s_x f , \quad (22)$$

$$p_y = s_y f , \quad (23)$$

$$d = kf^2 . \quad (24)$$

Using these notations, Eqs. (20) and (21) become

$$(X_p-C_x)z - p_x\left(1+d\left(\frac{(X_p-C_x)^2}{p_x^2}+\frac{(Y_p-C_y)^2}{p_y^2}\right)\right)x = 0 \quad (25)$$

$$(Y_p-C_Y)z - p_y\left(1+d\left(\frac{(X_p-C_x)^2}{p_x^2}+\frac{(Y_p-C_y)^2}{p_y^2}\right)\right)y = 0 \quad (26)$$

Between the 3D coordinates of a calibration point with respect to the camera frame and the 3D coordinates of the same point, but with respect to the stereo sensor frame, one can write the relation

$$\begin{bmatrix} x & y & z & 1 \end{bmatrix}^T = {}^{Cam}_{SS}T \cdot \begin{bmatrix} x_S & y_S & z_S & 1 \end{bmatrix}^T , \quad (27)$$

where (x_s, y_s, z_s) are the 3D coordinates of the calibration point with respect to the stereo sensor frame. The transformation from the camera frame to the stereo sensor frame is a function of t_x, t_y, t_z, α, β and γ (Paul, 1981). We denote this function by 28.

$${}^{Cam}_{SS}T = \begin{bmatrix} \cos\gamma\cos\beta & \cos\gamma\sin\beta\sin\alpha-\sin\gamma\cos\alpha & \cos\gamma\sin\beta\cos\alpha+\sin\gamma\sin\alpha & t_x \\ \sin\gamma\sin\beta & \sin\gamma\sin\beta\sin\alpha+\cos\gamma\cos\alpha & \sin\gamma\sin\beta\cos\alpha-\cos\gamma\sin\alpha & t_y \\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Using Eqs. (27) and (28), we can rewrite Eqs. (25) and (26) in the following way

$$F^x_{X_p,Y_p,x_S,y_S z_S}(\alpha,\beta,\gamma,t_x,t_z,p_x,p_y,C_x,C_y,d) = 0 , \quad (29)$$

$$F^y_{X_p,Y_p,x_S,y_S z_S}(\alpha,\beta,\gamma,t_y,t_z,p_x,p_y,C_x,C_y,d) = 0 . \quad (30)$$

The last two equations are non-linear. For each calibration point, we obtain one pair of non-linear equations. Using N (N>10) calibration points we obtain an over-determinate system of non-linear equations. To solve this system, we first use the Newton algorithm, to make the system linear, and then least square methods (Manusar, 1981, Naslau, 1999).

Figure 4 shows the calibration plate. It was made out of glass, in order to reduce possible modifications due to the temperature variation. The accuracy of the circle positions is between –0.01mm and +0.01mm.

One can also see, in Figure 4, that the calibration plate is fixed on a special device. This device can provide movements in three orthogonal directions (*x,y,z*) with an accuracy situated between -0.01mm and +0.01mm. The alignment between the frame and the calibration plate frame is done mechanically, and is adjusted and controlled using Leica 3D measurement system with an accuracy of 0.01mm. Finally, the total accuracy of the position of the circles is between -0.025mm and +0.025mm.

5 MEASUREMENT PROCEDURE

Our goal is to measure the 3D coordinates of a point with respect to the stereo sensor frame. We consider a point *P* with coordinates x_S, y_S, z_S with respect to the stereo sensor frame. This point will have the coordinates x_R, y_R, z_R with respect to the camera right frame and the coordinates x_L, y_L, z_L with respect to the camera left frame. With these notations one can write the next relation:

$$ {}^{SS}_{L}T \cdot \begin{bmatrix} x_L \\ y_L \\ z_L \\ 1 \end{bmatrix} - {}^{SS}_{R}T \cdot \begin{bmatrix} x_R \\ y_R \\ z_R \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (31) $$

where ${}^{SS}_{L}T$ represents the transformation from the stereo frame to the camera left frame, and ${}^{SS}_{R}T$ is the transformation from the stereo sensor frame to the camera right frame.

From the two images made with the stereo sensor we find the pixel coordinates of the point P. We denote these coordinates (X_L, Y_L) and (X_R, Y_R), for camera left and right, respectively. Between the 3D coordinates and the pixel coordinates of the point *P* one can write the relations

$$ \frac{1}{\left(1+d^L\left(\frac{(X_p-C_x)^2}{p_x^2}+\frac{(Y_p-C_y)^2}{p_y^2}\right)\right)}\frac{X_L-C_x^L}{p_x^L}\cdot z_L = x_L \quad (32) $$

$$ \frac{1}{\left(1+d^L\left(\frac{(X_p-C_x)^2}{p_x^2}+\frac{(Y_p-C_y)^2}{p_y^2}\right)\right)}\frac{Y_L-C_y^L}{p_y^L}\cdot z_L = y_L \quad (33) $$

$$ \frac{1}{\left(1+d^R\left(\frac{(X_p-C_x)^2}{p_x^2}+\frac{(Y_p-C_y)^2}{p_y^2}\right)\right)}\frac{X_R-C_x^R}{p_x^R}\cdot z_R = x_R \quad (34) $$

$$ \frac{1}{\left(1+d^R\left(\frac{(X_p-C_x)^2}{p_x^2}+\frac{(Y_p-C_y)^2}{p_y^2}\right)\right)}\frac{Y_R-C_y^R}{p_y^R}\cdot z_R = y_R \quad (35) $$

where $p_x^L, p_y^L, C_x^L, C_y^L, d^L$ are the intrinsic parameters for the left camera and $p_x^R, p_y^R, C_x^R, C_y^R, d^R$ are the intrinsic parameters for the right one. The values of these parameters are known because they were computed in the calibration procedure. We solve Eqs. (31) – (35), and compute x_S, y_S, and z_S, which in fact is our goal.

For measuring purposes, we use the same device as in the calibration procedure (see Figure 4). We move the plate in different positions, and we measure with the stereo sensor the 3D coordinates of the points from the calibration plate. The big advantage of this device is that we can control very precisely the x, y, and z movements of the plate (0.01 mm). This way, the accuracy of measurements made with the calibrated stereo sensor can be verified.

6 ANALYSIS OF THE MESUREMENT RESULTS

According to Tsai (1987) and Weng (1992), there are three ways to analyze the accuracy of the camera calibration process. We use the first method from their classification, which consists in analyzing the accuracy of 3D measurements though stereo triangulation.

We present three plots with the errors of the measurement results. They present the errors Δx, Δy, and Δz, obtained for each coordinate *x*, *y* and *z*.

We have measured 25 points situated in a plane. In our plots, the coordinates *x* and *y* indicate the position of the measured point in this plane, and the coordinate *z* indicates step by step the four errors presented before.

Figure 5.a shows the distribution of errors for the x coordinate. The errors are between -19μm and +24μm. Similarly, Figure 5.b shows the distribution of errors for the y coordinate. These errors take values between -16μm and +19μm. The errors for z coordinate, which are between -78μm and +91μm, are presented in Figure 5.c. Having obtained these errors, we have reached our goal of building a stereo sensor with the features described in the abstract.

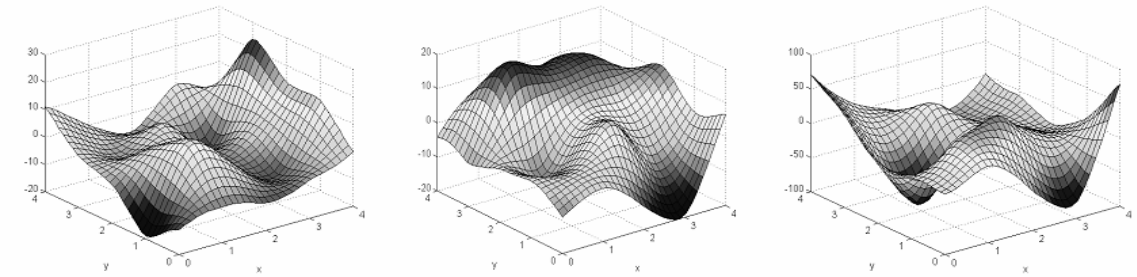

Figure 5: The error distribution for the measurement results

7 INDUSTRIAL APPLICATIONS

In industrial applications, a stereo sensor can be used in two configurations: as a fixed sensor or, as a mobile sensor mounted on the robot hand.

The first configuration can be employed in measuring the angle between the axles of a vehicle and the plane in which the wheels are rotating. The accuracy in such applications has to be very high. The solution developed in this paper, using a stereo sensor, provides this high accuracy. It can replace the current solution, which uses very expensive laser devices.

The second configuration, mobile sensor, is found useful in automatic processes, such as robotic hands mounting of windows for passenger cars. Here as well, this solution with a stereo sensor mounted on the robot hand can replace, with better results, the current solution. It needs only two cameras instead of four or eight, which are needed for the multi-camera method, which is presently used.

8 CONCLUSIONS

One of the main conclusions of this paper is that in order to obtain high accuracy and stable measurement results with a stereo sensor, it is necessary to include the radial distortion as a parameter in the camera model, and to make the image processing at sub-pixel level. We present in details the reasons why we need a sub-pixel approach. Furthermore, we develop an algorithm, which detects the position of the edge, by using a mathematical function to approximate the grey level for those points situated in the edge vicinity. The next step in the future research is to mathematically model the fact that the optical axis is not orthogonal to the CCD chip.

REFERENCES

Armangue, X. Salvi, J., Balle, J., 2000. A comparative review of camera calibrating methods with accuracy evaluation. V Ibero American Symposium on Pattern Recognition.

Agapito, L., Hayman, E., Reid, I, 2001. Self-calibration of rotating and zooming camera. Department of Engineering Science, Oxford University

Faugeras, O., Toscani, G., 1986. The calibration problem for stereo. Proceedings of IEEE Computer Vision and Patern Recognition. pp. 15-20

Faugeras, O., 1993. Three dimensional computer vision. Massachusetts Institute of Technology. London, 1993

Gui, V., 1999, Image Processing (in Romanian). Editura Politechnica Timisoara, 1999

Hall, E., Tio, J., McPherson, C., Sadjadi, F., 1982. Measuring curved surfaces for robot vision. Computer Journal. vol. December, pp. 42-54

Landsberg, G. S., 1958. Optics (in Romanian). Editura Technica Bucuresti, 2nd edition

Manusar, St., 1981. Numerical Methods to Solve Non-liniar equations. Editura Technica Bucuresti, 1981

Naslau, P., 1999. Numerical Methods (in Romanian). Editura Politechnica Timisoara, 1999

Parker, J. R., 1997. Algorithms for image processing and computer vision. Copyright © 1997 by John Wiley & Sons, Inc.

Paul, R., 1981. Robot Manipulators: Mathematics, Programing and control. The MIT Press Cambridge, Massachusetts. London, 1981

Tsai, R., 1987. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Int. Journal on Robotics and Automation. pp.323-344

Weng, J., Cohen, P., Herniou, M., 1992. Camera calibration with distortion models and accuracy evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence. vol 14, pp.965-980