ABSTRACT

The purpose of this paper is to compare three methods for three-dimensional measurements of line position used for the vision guidance to navigate an autonomous mobile robot. A model is first developed to map three dimensional ground points into image points to be developed using homogeneous coordinates. Then using the ground plane constraint, the inverse transformation that maps image points into three dimensional ground points is determined. And then the system identification problem is solved using a calibration device. Calibration data is used to determine the model parameters by minimizing the mean square error between model and calibration points. A novel simplification is then presented which provides surprisingly accurate results. This method is called the magic matrix approach and uses only the calibration data. A more standard variation of this approach is also considered. The significance of this work is that it shows that three methods that are based on three-dimensional measurements may be used for mobile robot navigation and that a simple method can achieve accuracy to a fraction of an inch which is sufficient in some applications.

Key Words: Autonomous, mobile robot, calibration, navigation, magic matrix.

1. INTRODUCTION

Humans receive a large amount of their information through the human vision system since it enables them to adapt quickly to changes in their environment. An intelligent machine such as mobile robot that must adapt to the changes of its environment must also be equipped with vision system so that it can collect visual information and use this information to adapt to its environment. The Center for Robotics at the University of Cincinnati has been involved in a nation wide competition to building a small-unmanned autonomous ground vehicle that can navigate around an outdoor obstacle course. The major components of the vehicle are: the supervisor control computer, speed control, steering control, obstacle avoidance, breaking system, emergency controls, and vision system. The purpose of the vision system is to obtain information from the changing environment - the obstacle course. The robot then quickly adapts to this information through its controller that guide the robot to follow the obstacle course. Modeling of the vision system is done with a CCD camera and that is the focus of this paper.

The obstacle course that the robot is supposed to follow is bounded by a solid and dashed lines ten feet apart and it can assume different shapes. Via the medium of the CCD camera, the lines of the obstacle course are digitized by the vision system from a 3-D coordinate system to a 2-D coordinate system. An image-processing tool for the vision system displays the image of the line. A 3-D coordinate system has been reduced to 2-D coordinate system by the camera system. The robot from the vision system easily obtains the information about the 2-D image coordinates. In an autonomous situation, how can the 3-D information or coordinates of a line be determined given its image coordinate? A mathematical as well as geometrical transformation occurs via the camera parameters in transforming a 3-D coordinate system to a 2-D system. If these mathematical and geometrical relations are known, a 3-D coordinate point on a line can be autonomously determined from its corresponding 2-D image point. To establish these mathematical and geometrical relationships, the camera has to be calibrated. The calibration of the camera or the vision system is the main task of this project. This is because if the vision system is well calibrated, accurate measurements of the coordinates of the points on the line with respect to the robot will be made. From these measurements, the orientation of the line with respect to the robot can be computed. With these computations, the next task is to guide the robot and this is a control problem.
In the fundamental theorem of robot vision is given. The manipulation of a point in space \( x_1 \) by either a robot manipulator that moves it to another point \( x_2 \) or through a camera system that images the point onto a camera sensor at \( x_2 \), is described by the same matrix transformation, which is of the form: \( x_2 = Tx_1 \). The transformation matrix \( T \) can describe the first-order effects of translation, rotation, scaling, and projective and perspective projections. The robot vision theorem suggests that the sensing of a point or collection of points on an object have some relation. In an effort to exploit this relation, the calibration of the sensing becomes very essential.

Camera calibration is a complex problem because of the following problems: (1) all the intrinsic and extrinsic parameters should be computed from the two-dimensional projections of a limited number of feature points, (2) the parameters are inter-related, and (3) the formulation is non-linear due to the perspective of the pin-hole camera model.

Several works have been done on calibration of cameras for various applications. Lovenitti, et al. [6] presented a 3-D coordinate measurement technique that uses a single 2-D image of four coplanar points, which are arranged in a square of known size, to measure geometric features on an object, some of which may be hidden from the view of the camera. In this approach, a hand-held probe is positioned on the object and in the view of a camera. The perspective projection of the probe is used to determine the 3-D coordinates of the point of contact. The image of the square’s four vertices is used to calculate the probe position and point of contact on the object. Caution is required during calibration. Hong, et al. [7] lists two points that should be considered in camera calibration: 1. Reducing the location error of image features as far as possible, by exploiting image processing techniques. 2. Compensating system error by the optimal pattern of approximating residual error of image points, namely the posterior compensation of the system error. Based on these two points, the calibration process discussed in [6] are of three parts: (1) the direct transformation error approximation camera calibration algorithm; (2) the subpixel image feature location algorithm combined with the 3D control point field delicate design and fabrication; (3) the precisely movable stage, which provides the reliable means of accuracy checking.

A method based on locating a few three-dimensional coordinates and their corresponding image coordinates on the image plane may be used to obtain the perspective transformation matrix elements. The use of the corner points of a cube known dimensions as the reference points to measure object surface is described by Parke. In Renner described a method that uses 23 miniature light bulbs on a pyramid base to serve as the reference points. Measurements of these reference points require the use of accurate and reliable measuring devices. In a study done by Tio, an object was partitioned into grids and the selection of points was based on a widely dispersed set of points that provide information on each major grid line. These reference points were used with the image points on sets of matrix equations to obtain the parameters for the calibration of the camera.

Tsai presented an algorithm that decomposes a solution for 12 transformational parameters (nine for rotation and three for translation) into multiple stages by introducing a radial alignment constraint. The radial alignment constraint assumes that the lens distortion occurs only in the radial direction from the optical axis \( Z \) of the camera. Using this constraint, six parameters are computed first, and the constraint of the rigid body transformation is used to compute five other parameters. The remaining parameters are computed by radial lens distortion parameter and estimating it by a nonlinear optimization procedure.

Liu, et al. first suggested the use of straight lines and points as features for estimating extrinsic camera parameters. Line features are usually abundant in indoor and some outdoor environments and less sensitive to noise than point features. The constraint used for the algorithms is that a three-dimensional line in the camera coordinate system \((X, Y, Z)\) should lie in the plane formed by the projected two-dimensional line in the image plane and the optical center. This constraint is used for computing nine rotation parameters separately from three translation parameters. They present linear and nonlinear algorithms for solutions. According to Liu et al’s analysis, eight-line or six-point correspondences are required for the linear method, and three-line or three-point correspondences are required for the nonlinear method. A separate linear method for translation parameters requires three-line or two point correspondences. Haralick, et al. reported their comprehensive investigation for position estimation from two-dimensional and three-dimensional sensed features. Other approaches based on different formulations and solutions include Kumar, Yuan, and Chen.

In this study, the calibration of the vision system was done with a specially constructed calibration device. First, three mathematical and transformational models were formulated to represent the camera parameters. For each of the models, a set of physical world coordinates and their corresponding image coordinates were obtained to solve for the unknown variables in the mathematical models. Accurate measurement of the physical points was very essential. The calibration device, which is made of a wooden base and six pin-pong balls, enabled us to achieve high accuracy. The three calibration models were evaluated and the one with the best performance was implemented on the robot.
Section 2 presents the mathematical and transformational models of the vision system. The geometrical model of the robot is also illustrated in this section. In Section 3, an experimental method, which describes the construction of the calibration device, was related. Results of the calibration process for each of the calibration models are given and illustrated.

2. MATHEMATICAL MODEL of the VISION SYSTEM

The purpose of the vision system is to guide the robot to follow a line using a digital camera. To do this, the camera needs to be calibrated. Camera calibration is a process to determine the relationship between a given 3-D coordinate system (world coordinates) and the 2-D image plane a camera perceives (image coordinates). More specifically, it is to determine the camera and lens model parameters that govern the mathematical or geometrical transformation from world coordinates to image coordinates based on the known 3-D control field and its image. The CCD camera digitizes the line from 3-D coordinate system to 2-D image system. Since the process is autonomous, the relationship between the 2-D system and the 3-D system has to be accurately determined so that the robot can be appropriately controlled to follow the line. The objective of this section is to show how three different mathematical models were developed to calibrate the vision system so that, given any 2-D image coordinate point, the system can mathematically compute the corresponding ground coordinate point.

2.1 Model of Robot Geometry

The model of the mobile robot illustrating the transformation between the image and the object is shown in Figure 1. The robot is designed to navigate between two lines that are spaced 10 feet apart. The lines are nominally 4 inches wide but are sometimes dashed. This requires a two-camera system design so that when a line is missing, the robot can look to the other side via the second camera. Measurements are referenced to the robot centered as a global coordinate system. For navigation, the cameras must be located with respect to this centroid.

![Figure 1. Top view model of the robot in the obstacle course](image)

2.2 Vision Processing Devices

Point or line tracking is achieved through the medium of a digital CCD camera. An Iscan[14] tracking device does image processing. This device finds the centroid of the brightest or darkest region in a computer-controlled window, and returns its X, Y image coordinates as well as size information. This information is updated every 16 ms, however the program must wait 10 ms after moving the window to get new data. This results in a 52 ms update time for tracking two points in sequence [15].

2.3 Mathematical Models

Three mathematical models were considered for developing the vision system. The models are the forward and inverse homogeneous matrix transformation, stereo vision model approach, and direct coefficient computation approach.

2.3.1 The Homogeneous Matrix Transformation
The homogeneous matrix transformation approach maps the physical coordinates to the image coordinates using matrix transformation. A point in global coordinates is related to the corresponding point in image coordinates using exactly the same transformations that are encountered in the physical system. The global coordinates are first translated to the center of the image plane, then rotated counter clockwise about the x-axis by the tilt angle, $\theta$, then rotated counter clockwise about the z-axis by the pan angle, $\phi$, then mapped through a perspective transformation with lens center located at $y=y_0$, and finally projected onto the x-z plane. These series of mappings are implemented as shown below. The imaging system model is shown in Figure 2.

![Imaging system Model](image)

**Figure 2. Imaging system Model**

### 2.3.1.1 Forward Transformation

The forward transformation is developed using homogeneous coordinates for the parameters shown in Figure 2.

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & \frac{1}{y_0}
\end{bmatrix}
\begin{bmatrix}
\cos \phi & \sin \phi & 0 & 0 \\
-\sin \phi & \cos \phi & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
y_0
\end{bmatrix}
\]

(1)

Multiplying the matrices and converting from homogeneous to physical coordinates gives:

\[
z_w = \frac{z - y_0 \sin \theta + z_2 \cos \theta + k \sin \theta + \ell \cos \theta}{w}
\]

(2)

\[
x_w = \frac{x - x_0 \cos \theta + y \sin \theta + z_2 \sin \theta - k \sin \theta - \ell \cos \theta + l \cos \theta + \ell \sin \theta + y_0}{w}
\]

(3)

### 2.3.1.2 Inverse Transformation

The inverse transformation equations were developed using the ground plane constraint. That is, three-dimensional measurements are made by back projecting and intersecting an image ray with the ground plane. The equations were developed symbolically using the computer program Mathcad. The equations permit computation of ground points from the image points.

\[
x_g = \frac{(y_z - y_0 \sin \theta \sin \phi + XPI \sin \theta \cos \phi + XPI \sin \phi \cos \theta + XPI \sin \phi \cos \theta - ZPI \sin \theta - ZPI \cos \theta)}{(y_0 \cos \theta \sin \phi - XPI \sin \theta \cos \phi - ZPI \sin \theta - ZPI \cos \theta)}
\]

(4)

\[
y_g = \frac{-(y_0 \cos \phi - ZPI \cos \phi + ZPI \sin \phi + XPI \cos \phi + XPI \sin \phi + XPI \sin \phi - y_0 \cos \phi + y_0 \sin \phi - y_0 \sin \phi - y_0 \cos \phi)}{(y_0 \cos \phi \sin \phi - XPI \sin \phi \sin \phi - ZPI \cos \phi)}
\]

(5)
2.3.1.3 Determining Model Parameters - System Identification

To make use of the forward and inverse transformations, the system parameters must be determined. The above equations have six explicit parameters. Three are the translation locations of the center of the image plane (h, k, and l). Two others are the pan, θ, and tilt angles, φ, of the camera. The sixth parameter is the lens center distance of the camera lens, f. Direct measurement can be used to give approximations to the parameters; however, the parameters (h, k, l) are measured to the center of the image plane while the others are measured at the lens center. Since the image plane is internal to the camera and the lens center is internal to the lens, accurate direct measurement is difficult. There are also implicit parameters in the actual system. For example the scaling between the optical image and the digital image must be determined. The digital image has a certain number of pixels in the horizontal and vertical directions. The camera sensor also has certain physical dimensions.

2.3.2 Stereo Vision Principles Model

The stereo vision principle approach uses the scaling between the two coordinate systems to determine the relationship between the physical and image coordinates. The model equation relating the two coordinates is:

\[ WX_{PI} = A_{11}x_g + A_{12}y_g + A_{13}z_g + A_{14} \] .........................(6)

\[ WY_{PI} = A_{21}x_g + A_{22}y_g + A_{23}z_g + A_{24} \] .........................(7)

\[ W = A_{31}x_g + A_{32}y_g + A_{33}z_g + A_{34} \] ......................... (8)

where \( W \) is the scaling, \( A_{nm} \) are coefficients, \( x_{PI} \) and \( y_{PI} \) are \( x \) and \( y \) image coordinates, and \( x_g, y_g, \) and \( z_g \) are the ground coordinates. Eliminating the scaling, \( W \), and utilizing six matching calibration points compute the coefficients for the above equations. Below are the resulting two equations:

\[ A_{11}x_g + A_{12}y_g + A_{13}z_g + A_{14} - A_{11}x_{PI} - A_{12}y_{PI} - A_{13}z_{PI}x_{PI} = 0 \] ........................(9)

\[ A_{21}x_g + A_{22}y_g + A_{23}z_g + A_{24} - A_{21}x_{PI} - A_{22}y_{PI} - A_{23}z_{PI}y_{PI} = 0 \] ........................(10)

The above equations represent two equations in twelve unknown coefficients, \( A_{nm} \). The coefficients are computed using magic matrix techniques by utilizing six matching calibration points. Since we are dealing with homogeneous system, the matrix will include an arbitrary scale factor. If the coefficient \( A_{14} \) is set as unity, the resulting transformation matrix will be normalized. With six calibration data points and \( A_{14} = 1 \), the following matrix equation was formulated.

\[ QA = 0 \] ..............................................................................(11)

Where

\[
Q = \begin{bmatrix}
  x_{g1} & y_{g1} & z_{g1} & 1 & 0 & 0 & 0 & 0 & -x_{g1}.x_{PI} & -y_{g1}.x_{PI} & -z_{g1}.x_{PI} & -x_{PI} \\
  0 & 0 & 0 & 0 & x_{g1} & y_{g1} & z_{g1} & 1 & -x_{g1}.y_{PI} & -y_{g1}.y_{PI} & -z_{g1}.y_{PI} & -y_{PI} \\
  x_{g2} & y_{g2} & z_{g2} & 1 & 0 & 0 & 0 & 0 & -x_{g2}.x_{PI}2 & -y_{g2}.x_{PI}2 & -z_{g2}.x_{PI}2 & -x_{PI}2 \\
  0 & 0 & 0 & 0 & x_{g2} & y_{g2} & z_{g2} & 1 & -x_{g2}.y_{PI}2 & -y_{g2}.y_{PI}2 & -z_{g2}.y_{PI}2 & -y_{PI}2 \\
  x_{g3} & y_{g3} & z_{g3} & 1 & 0 & 0 & 0 & 0 & -x_{g3}.x_{PI}3 & -y_{g3}.x_{PI}3 & -z_{g3}.x_{PI}3 & -x_{PI}3 \\
  0 & 0 & 0 & 0 & x_{g3} & y_{g3} & z_{g3} & 1 & -x_{g3}.y_{PI}3 & -y_{g3}.y_{PI}3 & -z_{g3}.y_{PI}3 & -y_{PI}3 \\
  x_{g4} & y_{g4} & z_{g4} & 1 & 0 & 0 & 0 & 0 & -x_{g4}.x_{PI}4 & -y_{g4}.x_{PI}4 & -z_{g4}.x_{PI}4 & -x_{PI}4 \\
  0 & 0 & 0 & 0 & x_{g4} & y_{g4} & z_{g4} & 1 & -x_{g4}.y_{PI}4 & -y_{g4}.y_{PI}4 & -z_{g4}.y_{PI}4 & -y_{PI}4 \\
  x_{g5} & y_{g5} & z_{g5} & 1 & 0 & 0 & 0 & 0 & -x_{g5}.x_{PI}5 & -y_{g5}.x_{PI}5 & -z_{g5}.x_{PI}5 & -x_{PI}5 \\
  0 & 0 & 0 & 0 & x_{g5} & y_{g5} & z_{g5} & 1 & -x_{g5}.y_{PI}5 & -y_{g5}.y_{PI}5 & -z_{g5}.y_{PI}5 & -y_{PI}5 \\
  x_{g6} & y_{g6} & z_{g6} & 1 & 0 & 0 & 0 & 0 & -x_{g6}.x_{PI}6 & -y_{g6}.x_{PI}6 & -z_{g6}.x_{PI}6 & -x_{PI}6 \\
  0 & 0 & 0 & 0 & x_{g6} & y_{g6} & z_{g6} & 1 & -x_{g6}.y_{PI}6 & -y_{g6}.y_{PI}6 & -z_{g6}.y_{PI}6 & -y_{PI}6
\]

and \( A^T = \begin{bmatrix}
  A_{11} & A_{12} & A_{13} & A_{14} & A_{21} & A_{22} & A_{23} & A_{24} & A_{31} & A_{32} & A_{33} & A_{34}
\end{bmatrix} \)

There are 12 unknowns in the matrix equation. However, since the matrix equation is homogeneous, there is an arbitrary scale. Moving the last column in the matrix \( Q \) to the right-hand side and applying the least square regression method can solve the transformation coefficients. Therefore, one coefficient can be arbitrarily selected leaving 11 coefficients to be determined. Since each image point (x, y) gives two equations, a minimum of five and one half image points could give a solution. A greater number of points permit a least square solution. After the coefficients are determined, \( W \) is computed and for any image coordinate, \( X_{PI} \) and \( Y_{PI} \), the corresponding ground coordinates are computed.
2.3.3 Direct Coefficients Computation Approach

The vision system was modeled by the following equations.

\[ x_{PI} = A_{11} x_g + A_{12} y_g + A_{13} z_g + A_{14} \] (12)

\[ y_{PI} = A_{21} x_g + A_{22} y_g + A_{23} z_g + A_{24} \] (13)

Where \( A_{nm} \) are coefficients, \( x_{PI} \) and \( y_{PI} \) are \( x \) and \( y \) image coordinates, and \( x_g, y_g, \) and \( z_g \) are the ground coordinates. In transforming the ground coordinate points to the image coordinate points the following transformation operations occur on the points: scaling, translation, rotation, perspective, and projective. Solving for the transformation parameters to obtain the image and ground coordinate relationship is a difficult task as noted in section 3.2. Fortunately, in the model equations given above, the transformation parameters are imbedded into the coefficients. To compute the coefficients, a calibration device was constructed to obtain 12 data points. With the 12 points, a matrix equation was yielded as shown below:

\[ X_{PI} = CA_{1k} \] (14)

\[ Y_{PI} = CA_{2k} \] (15)

where

\[ C = \begin{bmatrix} x_{g1} & y_{g1} & z_{g1} & 1 \\ x_{g2} & y_{g2} & z_{g2} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{g12} & y_{g12} & z_{g12} & 1 \end{bmatrix}; \quad X_{PI} = \begin{bmatrix} x_{PI1} \\ x_{PI2} \\ \vdots \\ x_{PI12} \end{bmatrix}; \quad Y_{PI} = \begin{bmatrix} y_{PI1} \\ y_{PI2} \\ \vdots \\ y_{PI12} \end{bmatrix} \]

Equations (3) and (4) consist of 12 linearly independent equations and four unknowns, the least-square regression method is applied to yield a minimum mean-square error solution for the coefficients. Below are the equations for the solution:

\[ A_{1k} = (C^T C)^{-1} C^T X_{PI} \] (16)

\[ A_{2k} = (C^T C)^{-1} C^T Y_{PI} \] (17)

Given an image coordinate \( x_{PI} \) and \( y_{PI} \), and \( z \) ground coordinate (the \( z \) coordinate of the points with respect to the centroid of the robot is maintained constant because of the ground constraint) the corresponding \( x_g \) and \( y_g \) ground coordinates are computed as indicated by the following matrix equations.

\[ \begin{bmatrix} x_g \\ y_g \end{bmatrix} = Q^{-1} B \] (18)

where

\[ Q = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}; \quad B = \begin{bmatrix} x_{PI} - A_{14} - (A_{13} z_g) \\ y_{PI} - A_{24} - (A_{23} z_g) \end{bmatrix} \]

Note that equation (1) and (7) can be modified to accommodate the computation of \( z_g \) when an elevation of the ground surface is considered.

3. EXPERIMENTAL METHODS

A calibration device was constructed to permit measurement of corresponding three dimensional object points and image points.

3.1 Calibration Device
The calibration device comprises a wooden base, painted black for contrast, six white ping-pong balls, and two five inch long poles. The wooden base is 14.5” x 11.5” x 0.75”. Four of the balls are placed on the same plane (the surface) of the wooden base while the poles that are pinned to the wooden base support the other two balls. Six other darkened points are spotted on the graph sheet, bringing the total points considered for the calibration to twelve. The points on the graph sheet are considered to be on the ground level.

3.2.1 The Homogeneous Matrix Transformation

Previously it was noted that the six parameters associated with this model are difficult to measure. This is due to the fact that the measurement is made in reference to the center image plane and lens center that is internal to the lens and inaccessible. For these reasons, arbitrary values were assumed for the parameters. These values were iteratively changed in order to minimize the mean square error between the model calculated data points and the measured calibration data. Figure 3 shows the data point plots for the model and calibrations.

Despite the fact that the parameters chosen gave the least minimum mean square error, the deviation (error) of the model points from that of the calibration is too high to be neglected. This discrepancy was due to lack of direct access to the camera lens which made almost impossible to measure the required parameters; as such the forward and inverse homogeneous transformation model is being given further investigation and was not implemented to solve the vision problem.

3.2.2 Stereo Vision Principles Model

This model involved the computation of the coefficients and the scale factor, W. Because the number of unknowns in Equations 10 and 11 is more than the number of equations, there are infinitely many solutions to this model. Even with the generated matrix equations from the calibration data, one of the values for the coefficients had to be assumed, in this case it was assumed to be 1. Figure 4 shows the data point plots for the model and calibrations.
The results of this model are way off the mark. As demonstrated by the plot in Figure 5, the model points have been scattered in disarray. The accuracy of this model was dependent on the coefficients. However because there are infinitely many solutions to the matrix equation, arriving at the best solution is extremely difficult if not impossible. This model is being given further thoughts and was not implemented to solve the vision problem.

### 3.2.3 Direct Coefficient Computation Approach

Unlike the stereo vision principles model where the scaling was expressed as a discrete equation, this model treated all the transformational operations, including scaling to be embedded into the coefficients. Figure 5 shows the data point plots for the model and calibrations.

The deviation of the model points from the calibration points shown in Figure 6 is very minimal compared to the plots in Figures 4 and 5. With an accuracy of about a tenth of an inch, this model has proven to be the best among the three models considered in this study. As a result of this reliable performance, the direct coefficient computation model was thus implemented to solve the vision problem.

A correlation plot for the original and the computed x coordinates is shown in Figure 6. The linearity of the plot means that the difference between the original coordinates and the computed ones is very small. Also computed to ascertain or test the discrepancies between the two sets of coordinates is the mean square error. The mean square error was 0.242 for the x-axis and 0.295 for the y-axis. The mean square error of a few tenths of an inch was considered accurate and reliable enough to compute the physical coordinates for our application. Further research is needed to provide more accurate three-dimensional imaging methods.
Correlation Plot

Figure 6. Correlation plot for the X-axis. (Mean square error = 0.242)

REFERENCES