Real World Coordinate Estimation using Homography Matrix and Plane Model Construction

Barath CSS[1] and Dr.A.Babiyola[2]

[1]Student, B.E. Electronics and Communication Engineering
[2] Professor, Department of Electronics and Communication Engineering
Meenakshi Sundararajan Engineering College
Kodambakkam, Chennai, Tamil Nadu, India.

---***---

Abstract – *Images represent information and play an increasingly important role in a wide variety of disciplines and fields for manipulating and interpreting visual information of a particular subject. The process of extracting essential information from the images has burgeoned in the recent years. The world we live is 3D. However, there has been very less insight on capitalizing the determininstic 3D scene model for estimating the real world coordinates of any point in the image mainly because it demands human intervention proving this to be a time consuming process. Therefore, this paper elucidates a simplified and a radically different approach for the efficient computation of the real world coordinates of any point in the image using the Homography matrix and also by constructing a 3D plane which is made to pass through the respective real world points in order to estimate the model coefficients of the fitted plane and use the determined coefficients to compute the real world coordinates.*

***Key Words*: PCL, RANSAC, Homography, real world coordinate, image coordinate, Plane Construction, inliers, outliers.**

1. INTRODUCTION

The world we live in today is influenced by technology to such drastic measures that imagining life without it seems to be impossible. Although, it's worth noting that the history of camera started even before the advent of photography, it was the invention of modern day cameras that gave various dimensions of how a photograph is perceived. Photographs allow us to communicate and facilitate information in an artistic manner and move people in ways that sometimes words cannot. Many recent developments haven't just stopped at capturing high quality images but have also paved the way for many remarkable and exciting developments in the field of Image Processing. The determination of real world coordinate from image coordinate has many applications in the field of Computer Vision. The conversion of image coordinate to world coordinate can be done by using a transformation after camera parameters that are known in Image Acquisition. But, there occurs a loss of information about the depth of the position when a transformation of image coordinate to world coordinate is performed as the operation is usually non invertible. Therefore, determining the real world coordinate of a point from the image is a challenging problem in Image Processing. Nevertheless, the transformation process has seen a notable proportion of research in the recent years. The field is uncertain mainly due to the lack of discernment possessed by the machines as compared to the humans proving that the human intellect is more powerful than the machines as it cannot perceive the depth as easily as human eyes can. The real world coordinate of a point in the image is estimated with the help of the Homography matrix and also by constructing a 3D plane fitted through the real world points which determines the model coefficients.

1.1 Coordinate Systems

When a point or vertex is defined in the scene and is visible to the eye or to the camera, it basically appears in the image as a dot or more precisely a pixel if the image is a digital one. A Digital Image is composed of a finite number of elements, each of which elements have a particular value at a particular location and is represented as a matrix in the form of rows and columns. Every element of the matrix are referred to as picture elements, image elements and pixels. When a point is defined in the scene, the coordinates of the point are defined with respect to the global or world Cartesian coordinate system. A three dimensional Cartesian coordinate system is formed by a point called the origin and a basis consisting of three mutually perpendicular vectors namely x-, y- and z axis. The coordinates of any point in space are defined by three real numbers: X, Y, Z as shown in Figure 1.

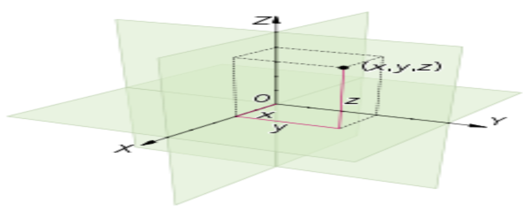


FIG-1: 3D CARTESIAN COORDINATE SYSTEM

1.2 Homography

The transformation between points expressed in the object frame and the projected points into the image plane expressed in normalized camera frame is a Homography. The Homography matrix is a 3x3 matrix but with 8 DoF (Degree of Freedom) as it is estimated up to a scale. It is generally normalized with h_{33} = 1 or h_{11}^2 + h_{12}^2 + h_{13}^2 + h_{21}^2 + h_{22}^2 + h_{23}^2 + h_{31}^2 + h_{32}^2 + h_{33}^2 = 1. Homography has many practical applications, such as image rectification or computation of camera motion- rotation and translation. The information extracted from the estimated Homography matrix may be used for navigation or to insert models of 3D objects into an image or video so that they are rendered with the correct perspective and appear to have been part of the original scene.

$$H_{3*3} = \begin{pmatrix} H11 & H12 & H13 \\ H21 & H22 & H23 \\ H31 & H32 & H33 \end{pmatrix}$$

2. PLANE MODEL CONSTRUCTION

Planes are commonly used in various vision tasks due to their abundance in man-made environments as well as to their attractive geometric properties. A plane is a flat, two dimensional surface that extends infinitely far. Planes in a three dimensional space have a natural description using a point in the plane and vector orthogonal to it which is known as normal vector. A plane is generally described by a normal vector $n= [a,b,c]^T$ and a distance d so that for a point for example, $P = [x,y,z]^T$ on the plane, the equation becomes : $(n*P) + d = 0$.

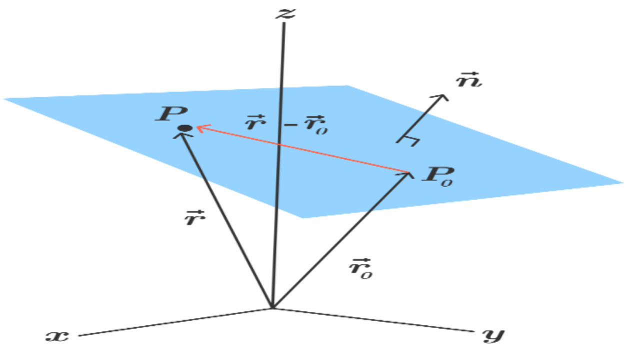


FIG-2: NORMAL VECTORS OF A PLANE

Let r_0 be the position vector of some point $P_0 = (X_0, Y_0, Z_0)$ and r be the position vector of point P as shown in Figure 2. Recalling that two vectors are perpendicular if and only if their dot product is zero, the equation can be modified as: $n . (r - r_0) = 0$. Expanding the above equation, the equation can be written as: $aX + bY + cZ + d = 0$. This is the general form of the equation of the plane.

Plane Model Segmentation is used to find all the points within the point cloud that supports a plane model. Point Cloud Library (PCL) being an extensive, open project is mainly used for point cloud processing. The library also contains numerous state-of-the art algorithms including filtering, feature estimation, model fitting and segmentation. In order to filter the outliers from noisy data, stitch 3D point clouds together, extract keypoints and compute descriptors to recognize objects in the world based on their geometric appearance and create surfaces from point clouds and visualize them, these algorithms can be used. The respective values are filled after creating the point cloud structure.

2.1 RANSAC Algorithm for Plane Detection

RANSAC is an abbreviation for "Random Sample Consensus". RANSAC is an iterative method to estimate the parameters of the mathematical model from a set of observed data that contains outliers. The basic assumption of the RANSAC algorithm is that it is a non-deterministic algorithm and that the data consists of "inliers" i.e., data whose distribution can be explained by some set of model parameters and "outliers" which are data that do not fit to the model. An assumption of the RANSAC algorithm is that given a set of inliers, there exists a procedure that estimates the parameters of a model that optimally fits the data. The blue points and the red points represent the inliers and outliers of the data as shown in Figure 3.

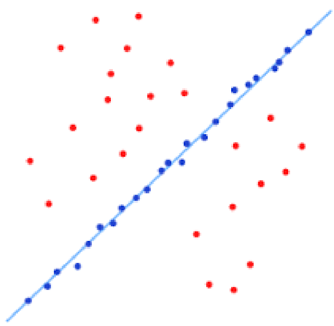


FIG-3: INLIERS AND OUTLIERS

Plane Detection is prerequisite to a variety of tasks and has been successfully employed in many diverse applications such as grouping, 3D reconstruction and scene analysis, object recognition and segmentation. The principle of RANSAC algorithm is to mainly find the best plane among the 3D point cloud. For this purpose, it randomly selects 3 points and calculates the parameters of the corresponding plane. Then it detects all the points of the original cloud belonging to the calculated plane according to a given threshold. Afterwards, it repeats the same process N times where it compares the obtained result with the last saved one. If the new result is better, it replaces the saved result by the new one.

3. RELATED WORK

A. Real World Coordinate from Image Coordinate Using Single Caliberated Camera by Analytical Geometry by Anton Sataria Prabuwono, Azizi Abdullah and Joko Siswontoro: proposed an algorithm for the determination of the real world coordinate of a point in a plane using single caliberated camera based on simple analytical geometry. This paper uniquely presents about the camera caliberation that has been done using the chessboard pattern taken from five different views for estimating the intrinsic and extrinsic camera parameters including rotation matrix, translation vector, focal length and the center of the image plane coordinate. These parameters together with the camera model and analytical geometry were used to approximate the real world coordinate of a point on a plane from its image coordinate.

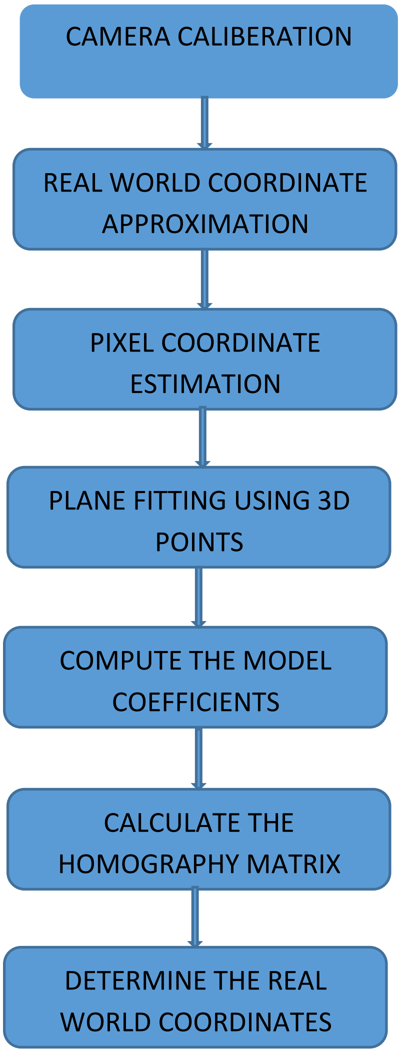
B. The Overview of 2D to 3D Conversion System by Li Sisi, Liu Wei and Wang Fei: explains the basic process of extracting the depth information from a monocular image using the depth generation algorithms which utilize different kinds of depth cues- monocular, binocular and pictorial depth cues. This paper also explains many approaches of the depth generation methods such as motion estimation, motion segmentation, object tracking algorithms, Depth Extraction, Depth Fusion machine learning algorithms in order to retrieve the depth map from different kinds of images.

C. 2D-to-3D Image Conversion by Learning Depth from Examples by Janusz Konrad, Meng Wang and Prakash Ishwar: proposes a novel approach in building a simplified and an efficient algorithm that learns the scene depth from a large repository of image+depth pairs quantitatively on a Kinect captured Image+Depth dataset using k nearest-neighbour (kNN) search, depth fusion, cross-bilateral depth filtering and stereo rendering.

4. PROPOSED SYSTEM

In the proposed system, the real world coordinates of a point in the image is estimated by considering many random points in the image as a set and approximating the real world measurements. The real world coordinates for the other points in the image will be computed in the same coordinate system as per our measurements using the Homography matrix and by constructing a 3D plane model. The 3x3 Homography matrix is responsible for transforming the 2-Dimensional image coordinates (x,y) to the 3-Dimensional world coordinates(X,Y,Z). 'N' random points are taken in the image whose image coordinates are known. For those 'N' points taken, a plane is constructed in such a way that it fits all the points for which the model coefficients (a,b,c,d) are found out. The system uses a mathematical approach in predicting the real world coordinates of any point in the image.

4.1 System Flow Diagram



4.2 Modules and Description

A. Camera Caliberation

In the first step of the implementation, camera is mainly used to capture an image of the tiled floor for proceeding with better approximation of real world coordinates in a mathematical approach. The image can either be captured by a camera present in the mobile phone or a normal camera. The camera must be projected in an angle so that the tiled floor gets captured properly as shown in Figure 4.

FIG-4: TEST IMAGE OF A TILED FLOOR

B. Real World Coordinate Approximation

In a 3-Dimensional space where there are three dimensions such as X,Y,Z, the point at which the image is captured becomes the origin point (0,0,0). Take random points in the tiled floor and approximate the real world coordinate mathematically.

C. Image Coordinate Estimation

The respective points taken for approximating the real world coordinates have to be considered once again for estimating the image coordinates which is basically represented as(x,y). The image coordinates for the points are found out by using the Paint Software.

D. Plane Fitting using 3-D Points

Since many points in the 3-Dimensional space are considered, there must be a plane that passes through or fits all these points. The process of plane fitting is mainly done in order to compute the model coefficients of the fitted plane. A point cloud of all the points is created with the Point Cloud Library. Plane fitting is implemented with the help of RANSAC algorithm which searches for the best plane that fits the 3D point cloud. At the same time, it also reduces the number of iterations if there are more number of points. A plane fitted through the real world points is shown in Figure 5.

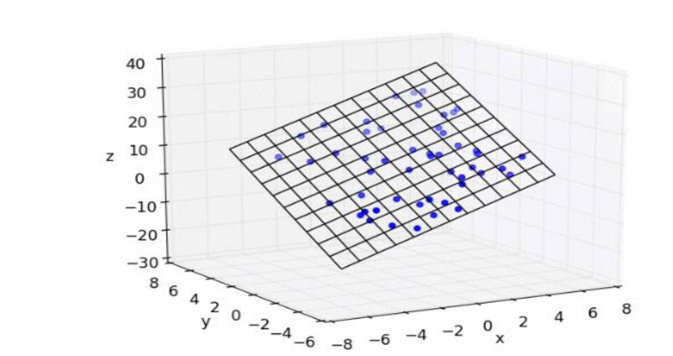


FIG-5: PLANE FITTING THE 3D POINTS

E. Compute the Model Coefficients

For any plane passing through 'N' points, the equation of a plane can be written as: aX + bY + cZ + d=0. Here, (a,b,c,d) are the model coefficients of the plane which passes through the points and (X,Y,Z) is the real world coordinate. The model coefficients are computed by using the Point Cloud Library (PCL). The RANSAC algorithm is mainly used here to estimate the model coefficients of the mathematical model whose input is a point cloud.

F. Calculate the Homography Matrix

To estimate H, we start from equation x2~Hx1. Written element by element in homogeneous coordinates, we get the following constraint:

$$(X,Y,W) = H_{3*3}\,(x,y,1) \qquad \text{……. (1)}$$

For one point,

$$\begin{pmatrix} X \\ Y \\ W \end{pmatrix} = \begin{pmatrix} H11 & H12 & H13 \\ H21 & H22 & H23 \\ H31 & H32 & H33 \end{pmatrix} * \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \qquad \text{……… (2)}$$

(X,Y,W) : Real World coordinates
(x,y,1) : Image coordinates

$$X = H11x + H12y + H13 + 0H21 + 0H22 + 0H23 + 0H31 + 0H32 + 0H33 \qquad \text{……. (3)}$$

$$Y = 0H11 + 0H12 + 0H13 + H21x + H22y + H23 + 0H31 + 0H32 + 0H33 \qquad \text{……. (4)}$$

$$W = 0H11 + 0H12 + 0H13 + 0H21 + 0H22 + 0H23 + H31x + H32y + H33 \qquad \text{……. (5)}$$

Substituting the coefficients of the Homography values from equations (3), (4), (5) in equation (2), we get:

$$\begin{pmatrix} X \\ Y \\ W \end{pmatrix} = \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x & y & 1 \end{pmatrix} * \begin{pmatrix} H11 \\ H12 \\ H13 \\ H21 \\ H22 \\ H23 \\ H31 \\ H32 \\ H33 \end{pmatrix} \qquad \text{……. (6)}$$

Similarly for 'n' points,

$$\text{Consider } [A]_{3n*1} = \begin{pmatrix} X1 \\ Y1 \\ Z1 \\ . \\ . \\ . \\ Xn \\ Yn \\ Zn \end{pmatrix}$$

$$\text{Consider } [B]_{3n*9} = \begin{pmatrix} x1 & y1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x2 & y2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x3 & y3 & 1 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ xn & yn & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & xn+1 & yn+1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & xn+2 & yn+2 & 1 \end{pmatrix}$$

Consider $[H]_{9*1} = \begin{pmatrix} H11 \\ H12 \\ H13 \\ H21 \\ H22 \\ H23 \\ H31 \\ H32 \\ H33 \end{pmatrix}$

Linear Equations: A = B x H .... (7)

Taking transpose on both the sides, $(B^T \times A) = (B^T \times B) \times H$ .... (8)

Homography matrix, $H = (B^T \times B)^{-1} \times (B^T \times A)$ .... (9)

G. Determine the Real World Coordinates

The final step of the process is to determine the real world coordinates (X, Y, Z). After finding the Homography matrix and the model coefficients of the fitted plane, the real world coordinates are computed based on the real world measurements approximated mathematically. The model coefficients are used in equation (10) to find the Z value. The calculated Homograhy matrix is plugged into equation (1) and the respective image coordinate values (x,y) are also put in equation (1) to determine the real world coordinates.

5. CONCLUSION AND FUTURE ENHANCEMENTS

Thus from the above experimentation, a simplified mathematical approach is followed in determining the real world coordinates. Admittedly, the proposed system has tried to solve the difficult problem in transforming the image coordinate to real world coordinate which purely demands human intervention. The validation was limited in determining the Z- coordinate which provides the depth value. The real world coordinates cannot be manipulated on the basis of a single image but needs multiple images from different angles and different terrains so as to give the machine a complete idea of the object that needs to be rendered making the entire process a time consuming one.

The project can be further enhanced by using many Machine Learning algorithms where the main idea is to build a machine learning model in order to calculate the Homography matrix and in turn, the real world coordinates. The model can be used to calculate various evaluation metrics such as Accuracy, Precision, F1-score and Recall. The images can be trained by the machine learning model using the supervised learning. The model must be trained again by shuffling and splitting the dataset in which the split dataset is sorted as the train and test dataset. Once the metrics reaches a value more than 90, the trained model will now be tested to predict which falls under the category called unsupervised learning where the prediction will be done by setting up a camera which captures the images and sends it to the system. The captured images are then sent to the model which will help in predicting which basically finds out the real world coordinates.

REFERENCES

[1] Anton Satra Prabuwono, Azizi Abdullah, Joko Siswantoro, "Real World Coordinate from Image Coordinate Using Single Caliberated Camera based on Analytical Geometry", S.A. Noah et al. (Eds.):M-CAIT, CCIS 378, pp. 1-11, 2013.

[2] Li Sisi, Liu Wei, Wang Fei, " The Overview of 2D to 3D Conversion System", IEEE 978-1-4244-7974-0, 2010.

[3] Janusz Konrad, Meng Wang, Prakash Ishwar, " 2D to 3D Image Conversion by Learning Depth from Examples", IEEE 978-1-4673-1612-5, 2012.