Object Dimension Extraction for Environment Mapping with Low Cost Cameras Fused with Laser Ranging

Sachini Ekanayake[#1], Nilanga Thelasingha[*2], Bavantha Udugama[#3], G.M.R.I. Godaliyadda[#4], M.P.B. Ekanayake[#5], B.G.L.T. Samaranayake[#6], J.V. Wijayakulasooriya[#7]

Dept. of Electrical and Electronic Engineering, University of Peradeniya,Peradeniya, Sri Lanka

[1]emspe9353@gmail.com,[2] nilanga901@gmail.com,[3]bavantha11@gmail.com, [4]roshangodd@ee.pdn.ac.lk

[5]mpb.ekanayake@ee.pdn.ac.lk, [6]lilantha@ee.pdn.ac.lk , [7]jan@ee.pdn.ac.lk

Abstract- **It is essential to have a method to map an unknown terrain for various applications. For places where human access is not possible, a method should be proposed to identify the environment. Exploration, disaster relief, transportation and many other purposes would be convenient if a map of the environment is available. Replicating the human vision system using stereo cameras would be an optimum solution. In this work, laser ranging based technique fused with stereo cameras has been used to extract dimension of objects for mapping. The distortions in the camera were calibrated using mathematical model of the camera. The rectified images were used to generate the disparity map by means of Semi Global Block Matching. The noise found in the disparity map was reduced using novel noise reduction method for disparity map through employing dilation. The Data from the Laser Range Finder (LRF) has been used to identify the 2D overlay of the environment. To find the missing the 3rd dimension of the objects the disparity map was analyzed through different methods like canny edge detection and pixel-wise intensity thresholding. Out of them through comparison of the results, pixel-wise approach has been selected, due its success in identification of the dimensions.**

***Keywords*- mapping, disparity, LIDAR, camera calibration, stereo vision, dimension extraction**

I. INTRODUCTION

Mapping is in general, graphical representation of environment with various objects and features. Mapping can be used to extract object information like height size, material types, object types etc. Machine vision is giving the capabilities of human vision to a machine using technology. human vision perceives as a colorful place with three dimensions.

Mapping plays a vital role since it allows us to respond to spacious geographic and social issues. Maps are useful for understanding and identifying spatial connections and explaining concepts in a visual manner that can be easily understood. There are many applications in technology of environment mapping as military applications, disaster relief, navigation and exploration etc., this can be used in many security applications in industry as well. Simply places where human cannot access this will be an ideal solution.

Explorations human inaccessible areas, finding dimensions of unknown terrains, by implementing this proposed method on a drone can be used for constructing 3D models of the down terrain. Further this can be extended to find material types and identifying object according to application. Number of sensors are available in the literature for detecting obstacles with help of SONAR [2], RADAR, LIDAR [3], vision systems and other proximity sensors [4].

But many of them are running under high computational power and with expensive equipment. On the other hand, stereo cameras are low cost and mapping has been done through stereo vision. Path planning techniques for traversing given point with minimum travel cost has been developed for some time [5]. Also, Laser Range Finder (LRF) is an accurate device for distance measurement. LRFs have been used for object dimension identification in many applications.

This paper outlines such approach step by step. through the proposed method, the LRF data has been fused with stereo cameras with low computational cost to extract dimensions of objects. While the LRF focuses on to finding 2D plane X, Y dimensions, the stereo camera is used to find Z dimension. The camera was calibrated and its properties were identified properly, so that the distortions can be corrected. The distortion corrected images were used to compute the disparity map. The disparity data has been used combined with the LRF data to successfully identify the dimensions of the objects in the environment. The methodology and the techniques used for noise filtering and dimension extraction has been discussed as well.

II. PROPOSED SOLUTION

In the proposed solution, two-low cost off the shelf web cameras has been used for stereo vision and an off the shelf neato XV – 11 LRF as sensors. Since camera calibration is a requirement it was calibrated properly up to 14 distortion coefficients without being limited to usual method.

From left and right images disparity map was obtained and the height of the object a found with it. Regarding the length and the width we use the LRF introducing this to find the planet of the objects and constructing the 2D environment as discussed in section [E].There are many researches [6] [9] which have been done using stereo cameras. But, taking multiple realization is a method with much computational cost in terms of practical implementation on a suitable platform. Hence, what has been proposed is, from stereo cameras we extract the object height information and construct the 3D object using 2D information obtained from the. In addition, the latter information will be used to construct the 3D object. For this purpose, only the two left and right images are needed with object height

information and the LRF sweep information. This is not much computationally costly or unrealizable.
The cameras used has its own distortions. Radial distortions and tangential distortions are prominent among them. Hence, it is necessary to calibrate them to avoid such distortions. Calibration involves identifying the mathematical model and details of the camera, which is done in two parts. First two cameras have been treated separately and calibrated and the error parameters were found. Also for the optimum disparity map generation rectification is necessary as the distortions are corrected and the images are aligned in a common horizontal line, so that the stereo matching can be executed.
Considering the height extraction, maximum intensity value of the disparity image was considered. Hence to remove the effect of ground plane noise values a constant distance was always maintained when taking images. So that identifying the maximum intensity with the relation of pixel – cm relation height of the object was found. There is a dimension change when taking 3D to 2D plane using stereo cameras. Hence, what we proposed is a method as discussed in section F.
Instead of taking multiple realization [6] since considering practical implementation it is not more efficient hence that we introduce novel approach to reduce noise in disparity image by dilation.
Once height is there and the plane of the object is gained using LRF our focus is to implement in a mobile platform and conclude the map reconstruction.
The dimension data obtained can be used reconstruct the objects in a virtual reality. To create a 3D model of the environment accurately. The identified dimensions can be used compute volume surface area, and they can be used for through analysis of the objects to find the density, mass and many other materialistic properties.
Today most of the areas are using depth extraction methods by stereo cameras with image processing. Proposed method is to take disparity using low cost cameras and dilation method is used to remove noise in disparity.

A. Camera model identification

The cameras have their inherent distortions. To find the depth accurately using cameras their intrinsic parameters have to be identified and distortions have to be corrected. The simple web cameras can be modelled as pinhole cameras. In a pinhole camera model, a scene viewed is developed by projecting 3D points unto the image plane using a perspective transformation as,

$$sm' = A(R|t)M'$$

or,

$$S\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where (X, Y, Z) are the coordinates of a 3D point in the world coordinate space. (u, v) are the coordinates of the projection point in pixels, and it is a principal point that is usually at the image center. A is a camera matrix and the focal lengths are expressed in pixel units.
Real lenses usually have distortions. Therefore, without modeling them, it is impossible to calculate any accurate measurement from the cameras. For the pinhole camera a. Radial distortions (Pincushion distortion & Barrel distortion.) and b. Tangential distortions are possible.
To identify these errors and to rectify the errors a calibration process is needed. The above mentioned distortions can be mathematically modeled as shown below. Here, k_1, k_2, k_3, k_4, k_5, k_6......., k_{14} are radial distortion coefficients and P_1, P_2 are tangential distortion coefficients.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t$$

$$x' = \frac{x}{z}$$

$$y' = \frac{y}{z}$$

$$x'' = x'\frac{1 + k_1r^2 + k_2r^4 + k_3r^6}{1 + k_4r^2 + k_5r^4 + k_6r^6} + 2P_1x'y' + P_2(r^2 + 2x')$$

$$y'' = y'\frac{1 + k_1r^2 + k_2r^4 + k_3r^6}{1 + k_4r^2 + k_5r^4 + k_6r^6} + 2P_2x'y' + P_1(r^2 + 2y')$$

where,

$$r^2 = x'^2 + y'^2$$

$$u = f_x x'' + C_x$$

$$v = f_y y'' + C_y$$

B. Camera calibration

Camera calibration was done in two parts. First individual cameras were calibrated and the error parameters and camera parameters were found. Then both cameras were calibrated as a pair to find the stereo parameters of the setup. For the calibration process a checkerboard with known dimensions was used as shown in figure 1.

FIGURE1: Camera calibration checker board images (left & right)

From camera calibrations following parameters were found.

- Radial distortion: Correction matrix for radial distortion
- Tangential distortion: Correction matrix for tangential distortion
- Focal length: Focal length of the camera
- Principle point: Principal point of the image plane

From stereo setup calibration following parameters were found

- Rotation of camera 2:

  Rotation of image plane of the camera 2 with respect to camera 1
- Translation of camera 2:

  Distance from the principal point of the camera 1 to principal point of the camera 2
- Fundamental matrix:

  The fundamental matrix is a relationship between any two images of the same scene that constrains where the projection of points from the scene can occur in both images.

The rotation matrix of camera two with respect to the camera one was found as,

$$\begin{matrix} 1 & 0.0027 & -0.0036 \\ -0.0027 & 1.0000 & 0.0016 \\ 0.0036 & -0.0016 & 1 \end{matrix}$$

And the translation of camera two with respect to camera one was found as,

-93.1032 1.2802 0.1104

Fundamental Matrix

-0.0000	-0.0000	0.0016
0.0000	0.0000	0.1269
-0.0018	-0.1274	0.6250

Essential Matrix

-0.0048	-0.4414	1.0334
0.1125	0.1493	93.1061
-1.2801	-93.1021	0.1459

TABLE I. CAMERA PARAMETERS

Parameter	Camera L	Camera R
Radial distortion	[0.0644 -0.2494 -0.6359]	[-0.0101 0.3192 -2.2780]
Tangential distortion	[6.9078e-04 -0.0011]	[-0.0048 -0.0035]
Focal length	[729.9077 729.4782]	[733.1340 732.8580]
Principle point	[322.5457 226.0965]	[303.3310 224.2269]

The parameters found were used to correct the distortions in the images were aligned horizontally through the process of rectification.

C. Disparity map generation

In figure 2, I represent the image plane and c denotes the principal point of image plane. (X, Y, Z) are

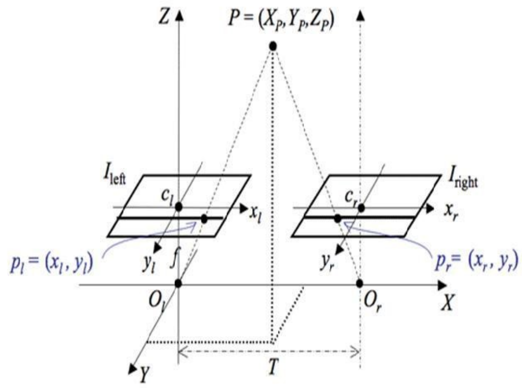

FIGURE2: Meaning of disparity

Cartesian coordinates. P denotes the real-world point and p denotes the images of that point. Subscript l and r

$$Z_p = f\frac{T}{d},$$
$$X_p = x_l\frac{T}{d},$$
$$Y_p = y_l\frac{T}{d},$$

denotes left and right. From geometry, a relationship between real world object points and image points can be derived as,

Here d stands for disparity which can be mathematically represented as,

$$d = x_l - x_r,$$

Disparity in each pixel is namely disparity map which is most of the time generated by Semi Global Block Matching (SGBM) [1]. It has been used as the basis to create a disparity map by using stereo images shown in figure 3.

FIGURE3: Stereo images of cube.

Disparity depends on the various parameters. An experiment was conducted to find those parameters as below.

For that, image pairs were captured by increasing the distance from baseline up to 1m. Also, the texture pattern of the surfaces was varied.