Distance Estimation based on Stereo Disparity Map for Autonomous Driver Less Vehicles using MatLab

1ST Mr.K.S.N.V.Someswara Rao
Dadi Institute of Engineering & Technology
JNTUGV University
Visakhapatnam ,India

2nd Alajangi Bargavi
Dadi Institute of Engineering & Technology
JNTUGV University
Visakhapatnam ,India

3rd Galla Vahini Krishna
Dadi Institute of Engineering & Technology
JNTUGV University
Visakhapatnam ,India

4th Tulagapu Vinay
Dadi Institute of Engineering & Technology
JNTUGV University
Visakhapatnam ,India

Abstract—A key area of interest in both science and industry is the possibility for a machine to replicate human vision by capturing three-dimensional (3D) information from the real environment.In order to replicate the real world and enable a machine that resembles human vision, a great deal of study has been done in the field of stereo vision. A stereo camera is used to determine an object's distance from an observer captured in an image. In order to determine the stereo characteristics of both cameras, camera calibration is carried out, and the pictures are corrected to be in the same plane. Using the disparity map and the stereo parameter to create a 3-D (3-dimensional) view, the difference between the left and right pictures is computed.By finding the centroid of the bounding box and computing the Euclidean distance between the centroid and the camera, the distance of the item is computed. The output of the object placed at a distance of between 10 and 3 metres is displayed in the results. The distance computation has an average inaccuracy of 2.08.

Index Terms: Stereo camera, parity, rectification, and camera calibration

I. INTRODUCTION

Calculating an object's distance from an observer is necessary in a number of domains, including navigation, robotics, self-driving cars, and security. As society develops more quickly, smart cars are keeping up with the changes. To prevent traffic jams and collisions, new capabilities that measure an object's distance from the car are also being developed. Active and passive methods are also available for determining an object's distance from an observer. Active measurement is the process of determining an object's distance from a source by means of delivering signals to it. Examples of such signals include radio signals, lasers, radar, infrared, and ultrasonic sound. When a distance measurement is performed without communicating with the object, this technique is known as with passive measurement, no signal is used in the information gathering process.This work focuses on the use of visual information from a stereo image captured by two cameras to calculate an object's distance from an observer. The stereo vision technology creates a three-dimensional description by combining many angles of view to capture a scene. It is regarded as passive because it does not employ additional lighting, such as a laser beam, in the scene. The primary uses for passive stereo vision are in robotics, 3D object recognition, and mobile robot navigation. In order to record the same scene from two different angles, a binocular vision formula is used in this paper's passive stereo vision technique. In automated control systems, primary data regarding object behaviour and state are crucial components.Reference offers a technique that uses an object's illumination and a fixed camera to create the illusion of perspective in an image. Reference offers a system that scans a live image captured by a group of imaginative cameras to identify movement in an object. In, a camera-based 3D picture construction method is suggested.

II. RELATED WORKS

Stereo vision is the process of creating a three-dimensional representation of a situation from two perspectives. An essential component of stereo vision is image correction. Two scenarios are generally considered for picture rectification: 1)Uncalibrated, 2)Calibrated.

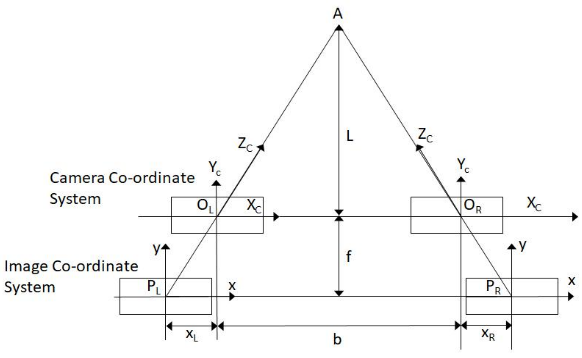
A. Uncalibrated Case

 ISSN NO : 1844-8135

An algorithm for picture rectification called Hartleys is based on stereo cameras that aren't calibrated. The Fundamental matrix in this approach is created utilizing a minimum of eight points that match between the two photos. The process of projective matrix estimation involves reducing the vertical disparity between comparable points. Many works implement the Hartleys algorithm. By breaking down F into its affine components and projective matrix, some changes were made. A method was presented in wherein the transformation matrix may be acquired without constructing the basic matrix, provided that the appropriate points were known beforehand when formulating the camera matrices was put forth; the correction is carried out in stereo picture pairs with various magnification and resolution. Epipolar lines are first approximated once the image has been reduced in size to fit the zoom.

B. Calibrated Case

To obtain the necessary transformation matrix for rectification in a calibrated stereo vision system, the projective matrix obtained from the calibration procedure is joined with the corresponding epipolar lines. This method was enhanced by using the original cameras' two projection matrices.



A summary of the image rectification processes in a calibrated stereo vision system is as follows:

(1) The stereo vision is calibrated in order to calculate the intrinsic and extrinsic parameters.
(2) For the original system, projective matrices are computed.
To produce two new projective matrices, (3) rotate the old matrix around the optical centers until the focus planes are coplanar.
(4) Calculate the matrix of transformations.
(5) Images undergo transformation.
In order to use a stereo vision system for tasks like 3D reconstruction, differences between the pairs of stereo images as well as the intrinsic and extrinsic matrices of the camera should be determined beforehand.

III.THEORY

Triangle theory is the foundation for creating a three-dimensional description of the target object utilizing stereo binocular vision.The target object and the two cameras form a triangle.The distance measurement technique using stereo binocular vision .
The point of interest is A.The two similar cameras, O L and O R, have the same focal length. The projection points of the left and right cameras, A and AR, respectively, are given. The length of the left and right abscissa, x_l and x_r, correspond to the AL and AR in the camera. Therefore, x_l and x_r represent the coordinate of the object. The actual distance between the two cameras is denoted by b, and the focal length of the cameras is given by fasshown in Fig.1, d= $x_l - x_r$ gives the difference mapping between the two images and Listhe actualdistance of the object from the camera. The value of L is given by the basic stereo binocular vision.

$$L = {}^{b*f}/_{x\,l - x\,r}$$

When modeling a stereo vision system to determine an object's distance from camera views, the following coordinate systems need to be taken into account
(1)World coordinate system(X_W,Y_W, Z_W),this is the Coordinate system defined by user in 3dimension.
(2)Camera coordinate system (X_C,Y_C,Z_C) it's the coordinate system with the camera as the centre and zaxisis taken along the optical axis.
(3)Image plane coordinate system (x,y),this planeis on the intersection of the optical center of the camera axis and the imaging point.
(4)Image coordinate system (u,v), which is in accordance With the pixel coordinate system.

The reference provides the relationship between the picture plane coordinate system and the world coordinate system, where the skew coefficient and the camera's focal length are given.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 0 \end{bmatrix} \begin{bmatrix} X_\omega \\ Y_\omega \\ Z_\omega \\ 1 \end{bmatrix}$$

Reference also provides the relationship between world Coordinate and image coordinate

$$s \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 0 \end{bmatrix} \begin{bmatrix} X_\omega \\ Y_\omega \\ Z_\omega \\ 1 \end{bmatrix}$$

Reference also provides the relationship between world coordinate and image coordinate

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_x & 0 & u_0 & 0 \\ 0 & a_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_\omega \\ Y_\omega \\ Z_\omega \\ 1 \end{bmatrix}$$

where, (u_0 ,v_0) is the coordinate when camera axis is projected to image coordinate and, a_x and a_y is u,v axis scale factors in image coordinate.

IV.PROPOSED SYSTEM

The first of the five stages in the proposed system is camera calibration. Next comes picture rectification, mapping the disparity, rebuilding the scene in three dimensions, object detection, and, finally, determining the item's distance from the camera.

A. Camera Calibration

Camera calibration is the process of estimating the camera's parameters. Camera calibration is used to estimate the intrinsic and extrinsic properties of both cameras. The extrinsic parameters account for the right camera's rotation, orientation, and translation with respect to the left camera. The intrinsic parameters give the mapping from the locations in the camera coordinates to pixel coordinates. These two parameters are utilized to restore the three dimensions of the acquired image and to correct the image.

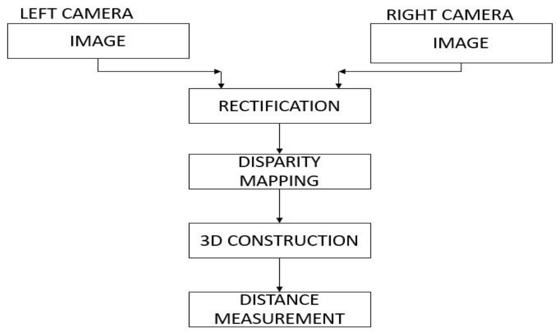


Fig.2.Flow diagram of the proposed system

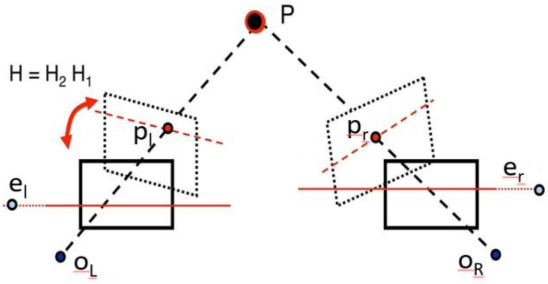


Fig.3.epipolar geometry

[c_x, c_y] is the optical center of the camera in pixels,

(f_x, f_y) is the focal length in pixels,

$f_x = p/ F_x$

$f_y = p/ F_y$

F = focal length in world units (in millimetres)

(p_x, p_y) size of pixels in world unit.

s is the skew coefficient

B.Image Rectification

Image rectification is the technique of projecting two or more images into a single image plane. Finding the exact location from one camera's perspective to the matching point on the other is a challenging undertaking. As an alternative to precise camera alignment, image rectification is carried out. In stereo vision, a scene is captured from two perspectives using two cameras that are positioned relative to one other and capture a pair of images. The associated depth of each pixel in the pair of photos is obtained by looking for matching pixels, and triangulation is then used to calculate the depth based on the matches found.

OL and *OR* are the optical centers of each camera with $M_L = K_L[I0]$ and $M_R = K_R[RT]$ are the camera matrices respectively for left and right camera and K_L and K_R are camera intrinsic matrices respectively for left and right camera, R and T are the rotation and translation matrices.

C.Mapping Disparities

The disparity is the product of two rectified images, which in turn provides the difference between the left and right images at their respective positions. A disparity map in stereo vision shows an object's depth. The disparity is inversely proportional to the object's depth in the picture. The discrepancy is going to be much greater if the object is closer to the camera.

$$d = x_l - x_r$$

D. 3D Reconstruction

Disparity and camera calibration parameters together reconstruct the three dimension of the image scene. Coordinates in three dimension which matches the pixel serve as the input of disparity map. Fundamental and essential matrices are used to find the three dimensional coordinates relative to the optical center of the left camera.

Essential matrix is given by:

$E = RS$

where R is the rotation matrix which is the extrinsic parameter of the calibrated camera and S is given by 3X3 matrix:

$$S = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$

where Tx, Ty and Tz are the three dimensions of translation matrix.

Assuming pl and pr are the pixels for a point in the left and right images then its coordinates are given by:

$\bar{p}l = K_L p_l$

and $\bar{p}r = K_R p_r$.

And p^t_r E p_l =0 which implies

$(K_R^{-1} \bar{p}r)E(K_L^{-1} \bar{p}l) = 0$

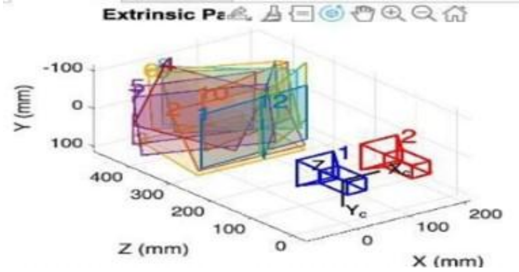


Fig.5.Stereo vision extrinsic parameter


Fig.6.Rectified image

where

$F = (K_R^{-1})^T EK_L^{-1}$

is the fundamental matrix.

After creating the 3 dimension of the scene, the object for which the distance is to be calculated is detected using a pre-trained detector for detecting upright person using the aggregate channel feature. After the upright person is detected a bounding box is drawn on the detected object.

E. Distance Calculation

Euclidean distance is used to calculate the distance between the camera and the object identified by disparity mapping in Fig. 7. Let (xc, yc, zc) be the coordinates for the centroid of the bounding box and (x, y, z) be the coordinates of the camera. The object's distance from the camera coordinates system can be determined using the reconstructed 3D scene.

$$D = \sqrt{(xc - x)^2 + (yc - y)^2 + (zc - z)^2}$$

To find the distance of the object from the baseline which connects OL and OR

V. EXPERIMENT

Two cameras that have the same specifications are placed in close proximity to one another to conduct the experiment. The object is positioned a few meters from the camera, which records a video of it. The video is then frame-by-frame examined.

A. Calibration of the Camera

Next, as depicted in Fig. 4, the stereo cameras are calibrated. Fig. 5 displays the extrinsic parameters that were determined during the stereo camera calibration. The stereo camera, which is employed to create the scene's three dimensions, generates its intrinsic and extrinsic properties. Table 1 contains the camera's specifications.

B. Image correction

As mention in the above section, the image frame is rectified to give proper alignment of the camera. The rectified image is shown

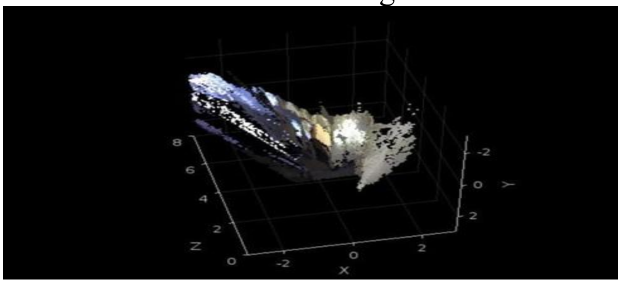


Fig. 8. Reconstructed 3D

Fig. 9. Object detection with calculated distance.

C. Mapping Disparities

One of the crucial processes in estimating the depth object in an image scene and reconstructing it in three dimensions is disparity mapping displays the mapping of disparities.

D. Reconstruction in Three Dimensions

The scene is reconstructed in three dimensions by utilizing the disparity mapping and intrinsic and extrinsic properties of the stereo camera displays the 3D scene that has been rebuilt.

E. Calculating Distance

Following the identification of the upright individual via the aggregate channel feature, the centroid of the bounding box of the detected person is estimated and the distance is calculated using Euclidean distance.

TABLE I
CAMERA PARAMETERS

Parameter	Left Camera	Right Camera
Resolution	1280 X 720	1280 X 720
Duration	4 seconds	4 seconds
Total frame	159	159
Frame rate	15 fps	15 fps
Distance between cameras	15 cm	15 cm

TABLE II
DISTANCE MEASURED

Actual distance (m)	Measured distance (m)	error(%)
5	4.82	3.6
3.6	3.5	3.3
3.4	3.35	2.4
3.3	3.28	0.6
3	2.86	0.5

VI. OUTCOME AND APPRAISAL

The distance of the person approaching the camera has been measured effectively, and the suggested method for determining an object's distance has been put into practice. The object's measured distance from the camera is displayed. Table 2 provides the actual distance, the measured distance, and the error. The following formula is used to compute the error:

$$error\ \% = |ad - md| \ / ad * 100$$

where md is the distance that the suggested system measured and ad is the actual distance.Table 2 shows that an object's distance can be measured for a master surveillance system with a small amount of inaccuracy.The primary cause of the problem is the sluggish processing of the frame-by-frame analysis and the non-simultaneous acquisition of movies from the two cameras. This suggests that the frame matching between the two images may be off, leading to an error in the distance calculation.

VII.CONCLUSION AND FUTURE SCOPE

This system uses the binocular triangle theory to create an asterovision system to determine the depth of an item in an image. It is demonstrated that distance can be estimated from the asterovision system with minimum error.The camera's estimated intrinsic and extrinsic parameters are determined by applying camera calibration.Aws in matching video frames can be minimized to further improve this strategy.The suggested system can be expanded by adding both calibrated and uncalibrated image correction algorithms, which can offer superior fundamental and spatial matrices and improve the scene's three-dimensional reconstruction.This approach can be used for robotics and autonomous cars, as well as for determining the safe driving distance.

REFERENCES

[1] A. Hou, X. Cui, Y. Geng, W. Yuan and J. Hou, “Measurement of Safe
Estimation Technique Using Laser Pointer And Single Video Camera
For Robot Vision,” International Journal of Engineering Research And
Driving Distance Based on Stereo Vision,” 2011 Sixth International
Conference on Image and Graphics, Hefei, Anhui, 2011, pp. 902-907.

[2] N. Sukhla, “A Review on Image Based Target Distance and Height
Reviews, Vol. 3, Issue 1, 2015.

[3] S. Nagar and J. Verma, “Distance Measurement Using Stereo Vision,”
International Journal of Electrical And Electronics Engineers, Vol. 07,
Issue 01, 2015.

[4] H. Kraft, J. Frey, T. Moeller, M. Albrecht, M. Grothof, B. Schink
and H. Hess, “3D-camera of high 3D-frame rate, depth-resolution and
background light elimination based on improved PMD (photonic mixer
device)-technologies,” OPTO, 2004, pp. 45-49.

[5] K. Kuhnert, M. Langer, M. Stommel and A. Kolb, “Dynamic 3DVision,”
Vision Systems: Applications, 2007, pp. 311-334.

[6] T. Gaspar and P. Oliveira, “New dynamic estimation of depth from focus
in active vision systems,” International Conference on Computer Vision
Theory and Applications, 2011, pp. 484-491.

[7] R. Hartley, “Theory and Practice of Projective Rectification,” Interna
tional Journal of Computer Vision, Vol. 35, 1999, pp. 115-127.

[8] C. Loop and Z. Zhengyou, “Computing rectifying homographies for
stereo vision,” IEEE Computer Society Conference on Computer Vision
and Pattern Recognition, Vol. 1, 1999, p. 131.