iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** X **Month of publication:** October 2017

DOI: http://doi.org/10.22214/ijraset.2017.10307

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue X, October 2017- Available at www.ijraset.com

Survey Paper on Stereo-Vision Based Object Finding Robot

Akshay Deshmukh[1], Yogeshwari Nikumbh[2], Rhishikesh Patil[3]
[1, 2, 3] Department of E&TC, SPPU

Abstract: The aim of this paper is to discuss about the survey that was conducted on stereo-vision based object tracking, recognition, and distance calculation. Previously implemented systems use two cameras to capture different images of an object. The images are processed using different algorithms to recognize objects and calculate the distance from the cameras. Distance is calculated by finding the positional differences of the objects in the images captured by the cameras. Object recognition and tracking is performed by separating the pixel values of the object from the background and assigning them values. Data is collected continuously after short periods to track objects in real time.
Keywords: Stereo-vision, two cameras, tracking, recognition, distance calculation, positional differences.

I. INTRODUCTION

In the world today, many applications require the use of robots for specialized tasks. These tasks may be difficult for humans to perform or the work environment may be too hazardous for them to work in. Robots make the jobs easier and the risk to human life can be minimized. The human visual system comprises of the two eyes that are separated by a small distance. This separation between the eyes produces two images that are processed by the brain. The images are combined to produce a single image. The advantage of such a system is that the differences in the positions of an object in these images help in determining the distance from the visual system. The stereo-vision based object finding robot uses two cameras to capture two different images of an object. The distance to the object is calculated by determining the disparity in the two images, i.e. the differences of the position and pixel values of the two images. As distance and disparity are inversely proportional, if the disparity is less, the distance is more and vice-versa.

II. METHODS USED

When a survey of previously implemented projects was conducted. It was found that some of the methods used to process the images captured by the two cameras are:

A. Jacobian Binocular Stereo-Vision

In this method, a matrix of the velocity of the cameras and feature points in the image is made. The signal to control the motion of the robot is generated based on the velocity of the feature points. An error function is generated to control the robot so that the feature points in the images reach their desired locations.

First, to make it simple, let us consider a case when the number of the feature points is one. The relation between the velocity of feature point in image ^{i}p and the velocity of camera frame ^{c}p is given as

$$^{i}p = J_c{}^{c}p \quad [1]$$

where J^c, is the Jacobian matrix which relates the two frames. Now let the translational velocity components of camera be $\boldsymbol{U_x}$, $\boldsymbol{U_y}$, and $\boldsymbol{U_z}$, and the rotational velocity components be $\boldsymbol{w_x}$, $\boldsymbol{w_y}$, $\boldsymbol{w_z}$, then we can express the camera velocity V as

$$V= [\mathbf{U_x U_y U_z w_x w_y w_z}] \quad [1]$$

Then the velocity of the feature point seen from the camera frame ^{c}p can be written

$$^{c}p = \frac{d^{c}p}{dt} \quad [1]$$

$$\frac{d}{dt}{}^{c}R_w({}^{w}p - {}^{w}p_c) \quad [1]$$

$$^{c}R_w\{-{}^{w}w_c x({}^{w}p-{}^{w}p_c)\}+{}^{c}R_w({}^{w}p-{}^{w}p_c) \quad [1]$$

When we use n feature points, image Jacobian J_1, ...,J_n are given from the coordinates of feature points in the image. By combining them, we express the image Jacobian (J_{im}) as

$$J_{im} = \{J_1.....J_n\} \quad [1]$$

Then, it is possible to express the relation of the moving velocity of the camera and the velocity of the feature points even in the case of plural feature points,

$$^{i}p = J_{im} V \quad [1]$$

B. Object Recognition and Grey Level Mapping

A control word is sent to move the camera. The image displacement is tracked and. The object can be tracked by comparing the grey level mapping of the images before and after the control word was issued. If the grey level difference is greater than a certain threshold which may be caused due to noise, it is understood as an error has occurred due to that control word.

Assumed that the object is located somewhere in the robot's work space, i.e., within the reach of the arm and the gripper and in the fields of view of the cameras. We are implicitly assuming that the work space actually exists, in other words, that the cameras are arranged in such a way that their fields of view are partly overlapping, and that part of the common field of view is accessible to the gripper. If the result of the object detector is no object to be detected in an image, a search motion is initiated. The principle of the object search is the conversion of passive "watching" into active "seeing" by rotating the cameras so that their fields of view scan the whole work space of the robot.

The number steps $\mathbf{N_{steps}}$required to search the robot's whole work space is:

$$N_{steps}=\frac{R_{limited,J1}}{\Delta_{opt-incri}}-1 \quad [2]$$

Where $R_{limited}$ is the limitation of the joint coordinate for joint

C. Pixel-to-Pixel Algorithm

In this method, operations are performed on the edge map that is generated from the obtained images. All pixels in with value 1 are considered as an edge. The algorithm takes two extreme points and compares them with a third point. The three points should not lie on the same line. The direction of the concavity is determined and plotted in a separate space.

This algorithm acts directly on the binary edge map, where each pixel values 1 if it is an edge point or 0 otherwise. In order to identify line direction, we explore the distribution of the pixels located in a window of dimensions $(2k + 1) \cdot (2k + 1)$ centred on each edge point.

Suppose the set of "1" to belong to a circular arc. Then the best estimation of radial line coincides with the perpendicular one to the chord AB passing for the mid-point C of coordinates:

$$C = ((x_A + x_B)/2 \; ; \; (y_A + y_B)/2) \quad [3]$$

To determine the two chord extremes A, B the algorithm examines all edge points in the current window and chooses those verify both:

1) Each of two must have the greatest Euclidean distance axes origin.

2) One at least of the homologue coordinates must have opposite sign. This is needed to reject some spurious points or angular points.

In order to reduce the computational burden required to compute quadratic Euclidean distance with respect to origin x2 + y2, we have defined a new distance with the same results: $d(x, y) = |x| + |y|$. So only absolute values need to be calculated. After A, B have been determined, and so the equation of perpendicular line, we have used the location of mid-point C to infer the arc concavity. In this way lines in the parameter space may be plotted only in the direction of concavity and consequently computational burden decreases.

D. Nonlinear Perspective Projection and Image Velocity Measurements

In this method, non-linear measurements are converted to linear measurements or combinations of the Cartesian coordinates. The bias and the covariance of the noise are estimated and then a robust filter is applied.

Let the velocity component in each direction be given by $[x_4, x_5, x_6]' \in R^3$ and let the acceleration in each traditionally denoted x, y and z direction be given by $[x7, x8, x_9]' \in R^3$. Hence, we can define $x = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9]' \in R^9$ such that the state evolves according to $x(k) = Ax(k - 1) + Bw(k)$ where A and B are suitably defined transition matrices given by

$$A = \begin{bmatrix} I_3 & k_s I_3 & \frac{k_s^2}{2} I_3 \\ O_3 & I_3 & k_s I_3 \\ O_3 & O_3 & I_3 \end{bmatrix} \qquad B = \begin{bmatrix} \frac{k_s^2}{2} I_3 \\ k_s I_3 \\ I_3 \end{bmatrix} \qquad [4]$$

and$w(k) \in R3$ is an uncertainty parameter.

The Riccatidfference equation is given by:

$$F(k+1) = \left[\hat{B}' S(k) \hat{B} + I\right]^{-1} \hat{B}' S(k) \hat{A} \qquad [4]$$

$$S(k+1) = \hat{A} S(k) \left[\hat{A} - \hat{B} F(k+1)\right] + CC' - K'K \qquad [4]$$

$$S(0) = N \qquad [4]$$

If we consider a set of state equations,

$$\eta(k+1) = \left[\hat{A} - F(k+1)\right]' \eta(k) + C' m(k+1) \qquad [4]$$

$$\eta(0) = Nx_0 \qquad [4]$$

$$g(k+1) = g(k) + m(k+1)' m(k+1) - \eta(k)' \hat{B} \left[\hat{B}' S(k) \hat{B} + Q(k)\right]^{-1} \hat{B}' \eta(k) \qquad [4]$$

$$g(0) = x'_0 N x_0 \qquad [4]$$

The Riccati difference equations and the state equations can be considered as the robust implementation of the standard linear Kalaman filter.

E. Perspective Projection Camera Model and Thresholding

Thresholding is used to separate the object from the background. The centroid of the object is calculated. The distance of the object from the cameras is calculated by using the coordinates of the centroid and the distance between the two cameras. hethresholding method [5, 6] to differentiate the moving object from an image scene because the target is a white light and the background is uniformly dark. The machine vision system uses the perspective projection camera model [7, 8] to determine the position of the target object. The thresholding method can be expressed by the following equation:

$$D(x,y) = \begin{cases} 255, & I(x,y) \geq T \\ 0, & otherwise \end{cases} \qquad [5]$$

where$I$ (x, y) stands for the gray-level value of (x, y) in the image plane, T is a threshold value, and $D(x, y)$ is the bilevel pixel. The thresholding method can separate the target object from the background. Then, the white pixels correspond to the target object. Because the target object is symmetrical, the centroid of the target object is taken as the location of the target object. The centroid is given by the following equation:

$$(U,V) = \left(\frac{1}{n}\sum u, \frac{1}{n}\sum v\right) \qquad [5]$$

where$n$ the total sum of the white pixel and (U, V) is the position of the target object in the image plane.

F. The advantages of these techniques are:

1) Accuracy is high
2) Easy to identify objects
3) Round object identification and segmentation is accurate

G. The disadvantages of these techniques are:

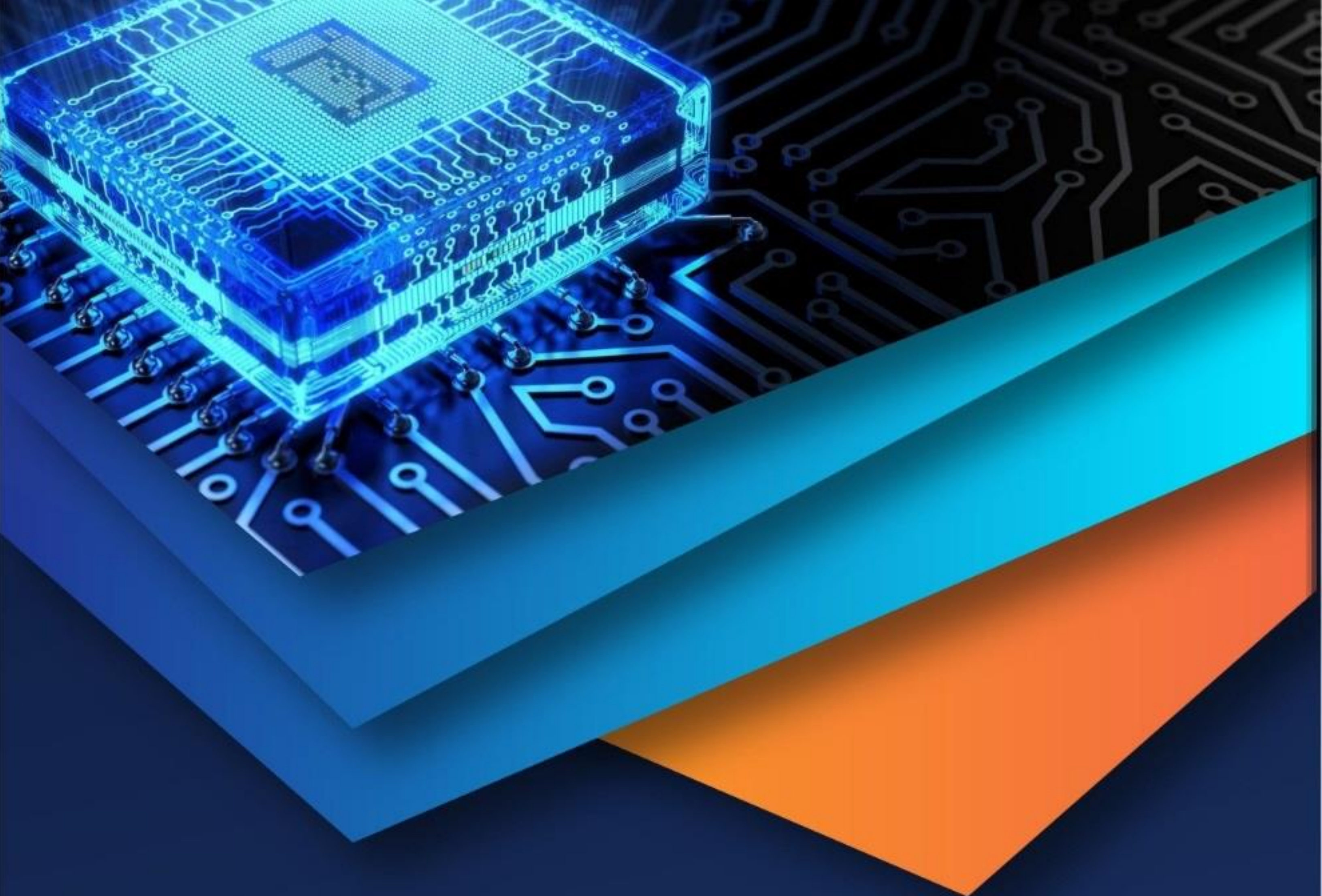
1) High complexity and computational time
2) Calibration of the cameras is not possible
3) Object tracking and the response of the system is slow

III. CONCLUSION

After reviewing several papers, it was found that the techniques used for calculating the disparity between the images captured by the two cameras and the distance to the object are time consuming, not accurate, and object identification is difficult.

REFERENCES

[1] Sung-Hyun Hank, W. H. Seo, S. Y. Lee, S. H. Lee, H. W Lee, and Higuchi Toshiro “A Study on Real-Time Implementation of Visual Feedback Control of Robot0 Manipulatorb IEEE-1999
[2] Minh-Chid Nguyen and Volker Graefe “Self-Learning Vision-Guided Robots for Searching and Grasping Objects” IEEE-200
[3] DavideScaramuzza, Stefano Pagnottelli and Paolo Valigi “Ball Detection and Predictive Ball Following Based on a Stereoscopic Vision System” IEEE-2005
[4] Pubudu N. Pathirana, Adrian N. Bishop, Andrey V. Savkin, Samitha W. Ekanayake and Timothy J. Black “A Method for Stereo-Vision based tracking for robotic applications” IEEE-2008
[5] Sho-Tsung Kao, Chung-Yi Yen, and Ming-Tzu Ho “Design and Implementation of a Stereo Vision-Guided Omnidirectional Mobile Robot for Real-Time Object Tracking” IEEE-2010
[6] middleburystereovision.edu
[7] http://en.m.wikipedia.org/wiki/Raspberry_Pi



doi crossref

10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY