Obstacle Avoidance Using Stereo Vision: A Survey

Pritesh S. Sharma[1], Dr. Nehal G. Chitaliya[2]

PG Student, Dept of E&C, SVIT, Vasad, Anand, India[1]

Associate Professor, Dept of E&C, SVIT, Vasad, Anand, India[2]

ABSTRACT*:* Stereoscopy is a technique used for recording and representing stereoscopic (3D) images. It can create an illusion of depth using two pictures taken at two or more slightly different positions. There are two possible way of taking stereoscopic pictures by using special two-lens stereo cameras or systems with two single-lens cameras joined together. Stereoscopic pictures allow us to calculate the distance from the camera(s) to the chosen object within the picture. One of the most important features for any intelligent ground vehicle is based on how is reliable and complete the perception of the environment and the capability to discriminate what an obstacle is. Stereovision system used to detect the distance from the obstacle by disparity of images. Stereo vision system provides pair of stereo images to determine distance after the detecting the object and measure distance from it and avoid the object. Avoidance done by any of the controlling device when getting the detection decision form stereo system.

KEYWORDS: Distance, stereo vision, disparity, matching, camera calibration, measurement.

I. INTRODUCTION

Obstacle avoidance is one of the main control system components in autonomous vehicles since a reliable perception of the real world is a key-feature for any obstacle detection system for dynamic environments. In last years, most of the historical approaches in literature have been readjusted in the framework of stereo vision and other 3D perception technologies (e.g. LIDAR) and important results have been provided by several experiments on autonomous ground vehicle. In order to achieve a good performance, most of the algorithms need some assumptions about the ground [21] or about the approximated free space on it. Blindness is defined as the state of being sightless.

This paper is organized as follows. Section II discusses the related work. Section III discusses the basics of stereo matching techniques. Section IV discusses the different types of applications based on stereo vision. Finally section V gives conclusion.

II. RELATED WORK

Stereoscopy is a technique used for recording and representing stereoscopic images. It can create an illusion of depth using two pictures taken at slightly different positions. In 1838, British scientist Charles Wheatstone invented stereoscopic pictures and viewing devices. Stereo vision is a technique for building a three dimensional description of a scene observed from several viewpoints. It is considered passive if no additional lighting of the scene, for instance by laser beam, is required. So defined, passive stereo vision happens to be very attractive for many applications in robotics, including 3-D object recognition and localization as well as 3-D navigation of mobile robots[18].

A. *Stereo vision*
Stereo vision is the extraction of 3D information from digital images, such as obtained by a ccd camera. By comparing information about a scene from two vantage points, 3D information can be extracted by examination of the relative positions of objects in the two panels.[5] This is similar to the biological process stereopsis. In traditional stereo vision, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene, in a manner similar to human binocular vision. The recovery of the 3D structure of a scene using two or more images of the 3D scene,

each acquired from a different viewpoint in space. The images can be obtained using multiple cameras or one moving camera. The term binocular vision is used when two cameras are employed [1].

B. *Epipolar geometry*

Epipolar geometry is the geometry of stereo vision. When two cameras view a 3D scene from two distinct positions, there are a number of geometric relations between the 3D points and their projections onto the 2D images that lead to constraints between the image points. These relations are derived based on the assumption that the cameras can be approximated by the pinhole camera model. [7].
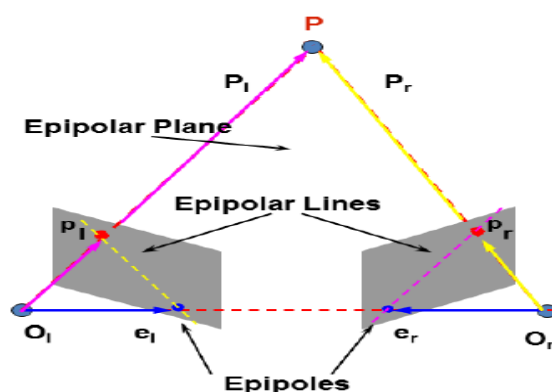

Fig. 1. Epipolar geometry

The Fig.1 depicts two pinhole cameras looking at point X. In real cameras, the image plane is actually behind the center of projection, and produces an image that is rotated 180 degrees. Here, however, the projection problem is simplified by placing a virtual image plane in front of the center of projection of each camera to produce an unrotated image. O_L and O_R represent the centers of projection of the two cameras. X represents the point of interest in both cameras. Points x_L and x_R are the projections of point X onto the image planes. Each camera captures a 2D image of the 3D world. This conversion from 3D to 2D is referred to as a perspective projection and is described by the pinhole camera model. It is common to model this projection operation by rays that emanate from the camera, passing through its centre of projection. Note that each emanating ray corresponds to a single point in the image [7].

III. STEREO MATCHING TECHNIQUES

In the past two decades, many stereo matching algorithms have been proposed. Categorizes all methods into sparse stereo and dense stereo matching. Categorizes all methods into explicit matching, hand-designed filters and network learning models. The most popular classification till now is global & local method. Calculating the distance of various points or any other primitive in a scene relative to the position of a camera is one of the important tasks of a computer vision system. The most common method for extracting depth information from intensity images is by means of a pair of synchronized camera-signals, the point-by-point matching between the two images from the stereo setup derives the depth images, or the so called disparity maps.

A. *Dense Disparity Algorithms*

Methods that produce dense disparity maps gain popularity as the computational power grows. Moreover, contemporary applications are benefited by, and consequently demand dense depth information. Dense disparity stereo matching algorithms can be divided in two general classes, according to the way they assign disparities to pixels. Firstly, there are algorithms that decide the disparity of each pixel according to the information provided by its local, neighbouring pixels [5]. There are, however, other algorithms which assign disparity values to each pixel depending on information derived from the whole image. Consequently, the former ones are called local methods while the latter ones global.

B. *Local Methods*

1) Normalized cross correlation (NCC) & Sum of absolute difference (SAD):

Muhlmann and his colleagues (2002) describe a method that uses the sum of absolute differences (SAD) correlation measure for RGB color images [6]. It achieves high speed and reasonable quality. It makes use of the left to right

consistency and uniqueness constraints and applies a fast median filter to the results. It can achieve 20 fps for 160*120 pixels image size, making this method suitable for real-time applications. Dynamic programming (DP), The similarity between pixels of the input image representation measured using various correspondence search methods such as the simple SAD-based method, the adaptive support weights method and the dynamic programming (DP) method[21].

2) Zero mean cross correlation (ZNCC):
This method integrates a neural network (NN) model, which uses the least-mean-square delta rule for training. The NN decides on the proper window shape and size for each support region [21]. The results obtained are satisfactory but the 0.024 fps running speed reported for the common image sets, on a Windows platform with a 300MHz processor.

3) Window based (fixed 2D window):
The basis for comparison of positions on different images is the result of a computation on the elements of a neighbourhood of fixed size. Windows have been very popular and are traditional within the correlation approaches. This approach has been made more robust by methods that work on a ranking of intensities of the window elements and use special metrics to compare candidate matching. An approach using more than one fixed window for each position is described by . Other window-based features can involve the output of filters or edge detectors [9]. Variable 2D window: Some approaches adaptively increase the size of an initial window, depending on a threshold on a variance measure, being more robust in large homogeneous areas of stereoscopic pairs. An advanced variable window method was proposed by that find the affine transformation that deforms the window in one of the images in such a way that a correlation measure is optimized [8].

C. *Global Methods*
Contrary to local methods, global ones produce very accurate results. Their goal is to find the optimum disparity function $d=d(x, y)$ which minimizes a global cost function E, which combines data and smoothness terms. $E(d)=E_{data}(d)+ k E_{smooth}(d)$ where E_{data} takes into consideration the (x, y) pixel's value throughout the image, E_{smooth} provides the algorithm's smoothening assumptions and k is a weight factor. The main disadvantage of the global methods is that they are more time consuming and computational demanding. The source of these characteristics is the iterative refinement approaches that they employ. They can be roughly divided in those performing a global energy minimization and those pursuing the minimum for independent scan lines using DP.

1) Color segmentation:
The algorithms that perform global optimization take into consideration the whole image in order to determine the disparity of every single pixel. Each segment is described by a planar model and assigned to a layer using a mean shift based clustering algorithm [16]. A global cost function is used that takes into account the summed up absolute differences, the discontinuities between segments and the occlusions. The assignment of segments to layers is iteratively updated until the cost function improves no more. The experimental results indicate that the percentage of unoccluded pixels whose absolute disparity error.

2) Graph cuts method (GC):
The reference image is divided in non-overlapping segments using the mean shift color segmentation algorithm [9]. Thus, a set of planes in the disparity space is generated. The goal of minimizing an energy function is faced in the segment rather than the pixel domain. A disparity plane is fitted to each segment using the graph cuts method. This algorithm presents good performance in the texture less and occluded regions as well as at disparity discontinuities.

3) Energy based:
Method which treats the two images of a stereo pair symmetrically within an energy minimization framework that can also embody color segmentation as a soft constraint. This method enforces that the occlusions in the reference image are consistent with the disparities found for the other image. Belief propagation iteratively refines the results. Moreover[10], results for the version of the algorithm that incorporates segmentation are better.

D.   *Other methods*

There are of course other methods, producing dense disparity maps, which can be placed in neither of previous categories. The below discussed methods use either wavelet-based techniques or combinations of various techniques. Such a method, based on the continuous wavelet transform (CWT) and many of feature detection and tracking filter based.

1) Feature detection method:

Feature detection technique is used to rectify and extracting the interesting points of both the images. Traditional methods of the stereo matching are SIFT and SURF. Scale-invariant feature transform is an algorithm in computer vision to detect and describe local features in images [7]. SIFT key points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation in the new image are identified to filter out good matches.

2) Speeded up robust feature detection:

SURF is a robust local feature detector that can be used in computer vision as object recognition or 3D reconstruction. This technique is inspired by the SIFT descriptor. The standard version of SURF is several times faster than SIFT and claimed by its authors to be more robust against different image transformations than SIFT[7]. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images. Both feature detector techniques are traditionally useful to find feature in stereo pair. These techniques have high execution time if used in real time application. Feature detection methods for image registration. Based on the experimental results, it is found that the SIFT has detected more number of features compared to SURF but it is suffered with speed[12]. The SURF is fast and has good performance as the same as SIFT.
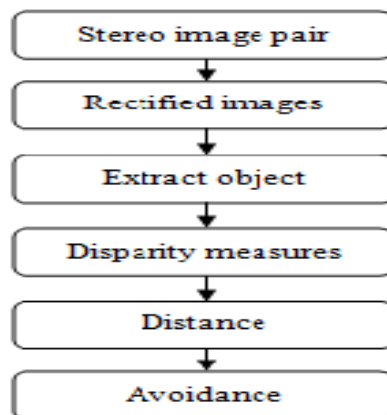

Fig.2 the Flowchart of general obstacle avoidance

 Fig. 2 shows a flow chart of the obstacle avoidance system. The input shows the pair of stereo images taken by stereo camera pair. After that rectified pair and extract the object based on color or shape of the object. Find disparity between pair and measure distance. Distance can be measure by triangulation technique or any other technique [3]. Then obstacle can be avoided.

E.   *Problems with stereo vision:*

1) Correspondence problem:

Finding pairs of matched points such, that each point in the pair is the projection of the same 3D point. Ambiguous correspondence between points in the two images may lead to several different consistent interpretation of the scene [2]. The basic problems with correlation in stereo imaging relate to the fact that objects can look significantly different from different viewpoints. It is possible for the two stereo views to be sufficiently different that corresponding areas may not be

matched correctly. Worse, in scenes with much obstruction, very important features of the scene may be present in only one view [3]. This problem is alleviated by decreasing the baseline, but the accuracy of depth determination suffers.

2) Reconstruction problem:

Given the corresponding points, we can compute the disparity map and calculate parameters on that. The disparity can be converted to a 3D map called reconstruction faces problems of the scene [3].

IV.   APPLICATION

A. 3D Tracking [11].
B. People counting (building, bus, train).
C. Monitoring trajectories (shopping, sport).
D. Safety [19].
E. Surveillance and security [21].

V.   CONCLUSION

Hence stereo vision is basically used for 3d reconstruction of image. Here, brief survey about the real-time approach for obstacle avoidance, mainly based on stereo vision. Extraction of the object from the stereo image pair is one of the main applications of stereo vision. A classification of the stereo matching methods is made in order to explain the different approaches presented in literature in the last years. Each work taken in consideration requires a good level of perception of the environment. Dense disparity map and dense scene flow map are studied. From particular distance the obstacle can be avoid by using stereo vision.

REFERENCES

1.   Hou A-Lin ,Cui Xue Geng , Ying Yuan Wen-Ju , Hou Jie, "Measurement of Safe Driving Distance Based on Stereo Vision," International Conference on Image and Graphics, IEEE,pp:209-906,2011 .
2.   Donguk Seo, Hansung Park and Kanghyun Jo, Kangik Eom, Sungmin Yang and Taeho Kim, "Omni directional Stereo Vision based Vehicle Detection and Distance Measurement for Driver Assistance System,'' IEEE ,pp:5507-5511,2013.
3.   Manaf A. Mahammed 1,Amera I. Melhum 2,Faris A. Kochery ,"Object Distance Measurement by Stereo Vision,'' International Journal of Scientific and Engineering Research (IJSAIT), Vol.2 , No.2, pp: 05-08 ,2013.
4.   Dhaval K. Patel, Pankaj A. Bachani, Nirav R. Shah,"Distance Measurement System Using Binocular Stereo Vision Approach." International journal of engineering and technology (IJERT), Vol.2, pp :2661—2664,2013.
5.   N.R.Rajaan, M.Ramkumar B.Monisha ,C.Jaiseeli ,S.Prasanna venkatesan,"Disparity Estimation from stereo images," (ICMOC-2012),Procedia engineering Science direct ,Volume 38, pp: 462–472,2012
6.   T. Surgailis, A. Valinevicius, V. Markevicius, D. Navikas, D. Andriukaitis," Avoiding Forward Car Collision using Stereo Vision syatem," Elecronika Ir Elektrotechnika, VOL. 18, pp:37-40, 2012 .
7.   Ashraf Anwar Fahmy ,"Stereo Vision Based Depth Estimation Algorithm In Uncalibrated Rectification'' International Journal of Video&Image Processing and Network Security IJVIPNS-International journals of engineering and sciences(IJENS) Vol:13,pp: 37-40, 2013.
8.   Komal D. Bhavsar, Prof. Virendrasingh ,"Analysis of Disparity Map for Stereo Matching Algorithm,'' IJSRD - International Journal for Scientific Research & Development Vol. 1, pp:2339-2347, 2014.
9.   Komal D. Bhavsar, Prof. Virendrasingh ,"Object Segmentation Based Disparity Map Determination for Stereo Images-A Review,'' IJSRD - International Journal for Scientific Research & Development Vol. 1, pp:2343-234, 2014.
10.   Prof.Ssergiu Nedevschi,Dr.Rolf Schmidt,Dr.Thorsen Graf, "High Accuracy Stereo Vision System For Far Distance Obstacle Detection,''IEEE Intelligent Vehicles Symposium,pp:292-297,June-2004.
11.   Kunsoo Hu,Jaehak Park,Junyeon Hwang, Daegun Hong,"A stereo vision-based obstacle detection system in vehicles,'' Optics and Lasers in Engineering , ,Science Direct,Elsevier,pp: 168–178 ,2008.
12.   Stefan K. Gehrig ,Fridtjof J. Stein, "Collision Avoidance for Vehicle-Following Systems,'' IEEE transactions on intelligent transportation systems, VOL. 8, pp: 233-244,2007.
13.   Filip Suligoj,Bojan Sekoranja,Marko Svaco, Bojan Jerbic,"Object Tracking with a Multiagent Robot System and a Stereo Vision Camera,'' 24th DAAAM International Symposium on Intelligent Manufacturing and Automation , ScienceDirect,pp:968 – 973, 2013
14.   L. Nalpantidis, A.Amanatiadis, G.Sirakoulis2, N. Kyriakoulis, and A. Gasteratos, "Dense Disparity Estimation Using a Hierarchical Matching Technique from Uncalibrated Stereo Vision,'' IST 2009 - International Workshop on Imaging Systems and Techniques Shenzhen, IEEE, pp:427-431,2009.
15.   Mr. Brojeshwar Bhowmick, Mr. Sambit Bhadra, Mr. Arijit Sinharay, "Stereo Vision Based Pedestrians Detection and Distance Measurement for Automotive Application,'' IEEE ,pp: 25-29 , 2011.
16.   Antonio Marín-Hernández,Michel Devy, "Application of a Stereovision Sensor for the Occupant Detection and Classification in a Car Cockpit,'' PhD student funded by the Mexican government, science direct ,pp:1-6:,2010.
17.   Young-Chul Lim, Chung-Hee Lee, Soon Kwon and Jong-hun Lee, "Tracking Distance and Velocity using a Stereo Vision System,'' 2008 IEEE, pp:-48-50,.2008.
18.   Nicola Bernini, Massimo Bertozzi, Luca Castangia, Marco Patander1 and Mario Sabbatelli, "Real-Time Obstacle Detection using Stereo Vision for Autonomous Ground Vehicles,'' Intelligent Transportation Systems (ITSC) ,pp:873-878, 2014.

19. Li Zhou, Tao Sun, Yuanzhi Zhan and Jia Wang."Software and Hardware Implementations of Stereo Matching,'' International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.7, pp:37-56,2014.
20. www.middlebury.edu

BIOGRAPHY

Mr. Pritesh S. Sharma received the Diploma degree in Electronics and Communication from M. S. University Polytechnic, Vadodara, Gujarat, India in 2010. He has received the Bachelor degree in Electronics and Communication from Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India in 2013. He is currently pursuing the Master degree in Electronics and Communication Engineering from Sardar Vallabhbhai Patel Institute of Technology, Vasad, India. His research interests include Digital Image Processing.

Dr. Nehal G. Chitaliya received B.E Electrical (1996) and M.E Electrical (2000) and Ph.D. Electrical (2013) from the Electrical Engineering department, Faculty of Technology and Engineering, The M.S. University of Baroda, Vadodara, Gujarat, India. She is Associate Professor of Electronics and Communication Engineering Department, Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India. Her research interests are in the field of Digital Image Processing, Signal Processing and Motion Analysis. She is a member of professional bodies like Indian society for Technical Education (ISTE) and International Association of Computer Science and Information Technology (IACSIT).