Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 133 (2018) 153–160

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Robotics and Smart Manufacturing (RoSMa2018)

A Hybrid Vision System for Dynamic Obstacle Detection

Kirti Shankar Sharma*, Sudeepta Ranjan Sahoo, P.V. Manivannan

Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai-600036, India.

Abstract

This paper deals with the depth analysis of a target in the dynamic state using combined techniques of stereo vision and Kanade-Lucas-Tomasi (KLT) feature tracking based method. The stereo vision system built using low-cost cameras helps in finding the depth of points on target; however, it is incapable of capturing all the depth point details on the target. Similarly, Kanade-Lucas-Tomasi (KLT) feature tracking based method provides only the direction of displacement of the target without quantifying it. Hence, we propose to develop an algorithm which fuses the techniques of stereo vision method and Kanade-Lucas-Tomasi (KLT) feature tracker to track the dynamic target with static observation point. MATLAB based point cloud generation is used frame-by-frame to map the three-dimensional environment by using the self-integrated low-cost stereo vision system. An initial frame is used to define the target to be tracked. The proposed algorithm is tested on vehicle movement with various speeds and directions. The algorithm is validated and verified for its performance and accuracy by comparing the experimental outcomes to the actual displacement of P3-DX. Results show that the experimental results are within the acceptable tolerance.

Keywords: Stereo vision; disparity; optical flow; point cloud; pupillary distance; motion trajectories; Kanade-Lucas-Tomasi(KLT) feature tracker;

1. Introduction

The development of automated vehicles has tremendously enhanced research in vision for autonomous navigation. Initially, environmental perception for the automated systems utilized multiple sensors including Ultrasonic sensors, Infrared sensors etc. [1]. The outputs of these sensors were complexly affected by environmental conditions, such as frequency of sound, vibrations, etc. Additionally, sensor integration and fusion [2] for the proper and timely perception of the obstacles has always remained a major challenge. The development of camera/ vision, greatly reduced this challenge by reducing the number of required sensors and sensor integration tasks [3]. Introduction of low-cost cameras has to a great extent, improved the perception capabilities of autonomous systems. Subsequently, inspired by nature, Roboticist began using a pair of cameras called stereo vision system [4].Similarly, with the development of high precision cameras, a lot of researchers began using monocular vision system combined with range [5] or laser

* Kirti Shankar Sharma. Tel.: +0-828-557-4339 ;
E-mail address: sharmakirtishankar15@gmail.com

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.

10.1016/j.procs.2018.07.019

sensors [6]. The accuracy of environment mapping using only a camera (or set of cameras), mainly depends on the structure of the environment, lighting conditions [7], camera alignment [8], atmospheric conditions, camera quality (nonlinearity correction), calibration [9], etc.

In early 2000, an Israeli based developer, 3DV Systems developed RGB-D imaging technique based on time-of-flight method. This technique implemented Infrared range finder along with a camera for obtaining the real-time image/video along with the depth, thereby providing three-dimensional images/videos. Shortly, Microsoft also released its first RGB-D Device Kinect version1 and version2 in 2010 and 2013 respectively. Initially, these devices were used solely for the purpose of gaming but soon became popular among Roboticist for their importance in depth perception [10]. Although these RGB-D devices are highly precise for indoor environment [11] they fail in accurately mapping the outdoor environment due to varied lighting condition[12]. This is the main reason researchers are inclined towards stereo vision [13] and monocular vision systems using readily available low-cost cameras.

Optical flow methods are primarily implemented for object tracking. Kahlouche. et al. used optical flow method for dynamic obstacle detection based on the feature point movement [14]. This method implements the selection of certain random reference points (called as feature points) from a frame and compares them with the consecutive frame, resulting in estimating the direction of object's displacement. Kuo, et al divided the total camera field view into an equally placed fixed number of points, and the movement of these optical points in each frame was shown as a vector in the direction of movement [15]. Although optical flow method is very good for the detection of dynamic obstacle and its direction of motion, it fails in the detection of the exact displacement [16]. Consequently, in this paper, we have integrated stereo vision with optical flow method (Kanade-Lucas-Tomasi-KLT feature point tracking method) [17] in order to detect the exact position of the obstacle in its movement.

The rest of the paper is organized as follows: Section 2 explains about the experimental setup developed using low-cost cameras, Section 3 includes the methodology followed and the experimental validation of the proposed algorithm, Section 4 explains about result of the proposed algorithm and the interpretation, while Section 5 and 6 includes conclusion and references respectively.

2. Experimental setup

Stereo vision is a bio-inspired vision system. In a binocularly visioned creature, both eyes are separated by a certain distance (also called Pupillary distance-PD) which varies from person to person. For better perception of the targets for the developed quadruped robot system [18], PD is chosen to be 126mm. The main objective of implementing the vision system on the robot is to better perceive the environment and to possibly avoid obstacle within the range of 1000mm. The developed stereo vision (SV) setup with two low-cost cameras is shown in Fig.1.The vision system for a robot also depends on the camera quality, focal length, frame rate, etc and to further improve the performance, two identical Logitech c310 cameras have been utilized.

Fig. 1. (a) Actual stereo camera setup; (b) Solidworks model for stereo camera stand.

The testing is done with stationary position of cameras which is connected via USB port with Intel(R) Core i7-2600 CPU @ 3.40GHz, 8.00 Gb installed RAM and x64 based processor. Logitech c310 camera used has the following specifications:

- High speed 2.0 USB.
- Plastic lens and CMOS sensor.
- Fixed focus type.
- Focal length of 4.4mm.
- Field of view = 60°

3. Methodology and Experimentation

This section deals with the methodology implemented in order to fuse both the techniques of SV and KLT to obtain the exact initial and final position of the object after every movement. The procedure followed in this work is explained through flow chart given in Fig. 2.

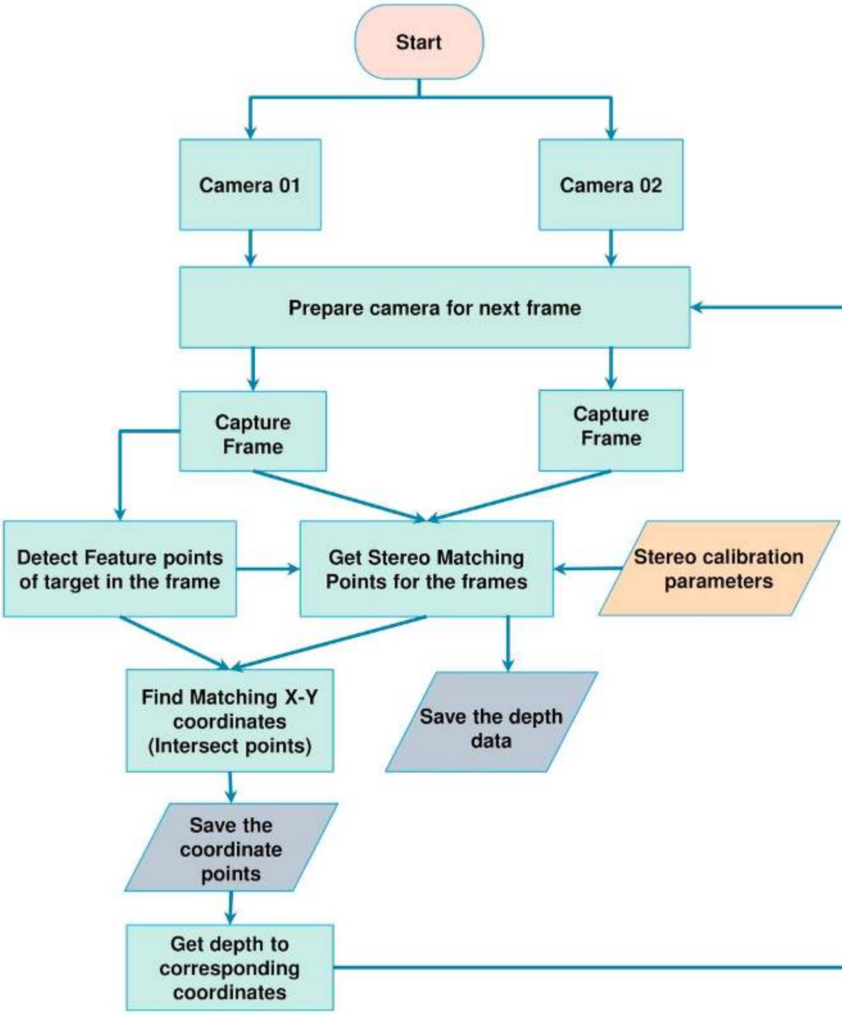


Fig. 2. flow chart showing the methodology for Algorithm.

3.1. Camera calibration and stereo vision system modeling

In the field of autonomous robotics, since the positions of the robot, obstacle/target and structure of terrain decides the subsequent steps to be taken by the autonomous robot, camera motion estimation and 3-D structure recovery from a stereo pair of images [19] are of major interest. The stereo vision system consists of two or more identical specification camera sets, pointing towards the same object and kept a certain distance apart. The two object image planes obtained from the two cameras will have some parts in common; however, there will be some parts which will be occluded from one or the other image plane. The difference of a 3D point in two image planes is referred to as disparity. The concept of disparity is used to generates a 3D perception (depth). The concept of disparity can more precisely be understood by Fig. 3. Let us consider the camera to be a linear device i.e. there is no radial and tangential

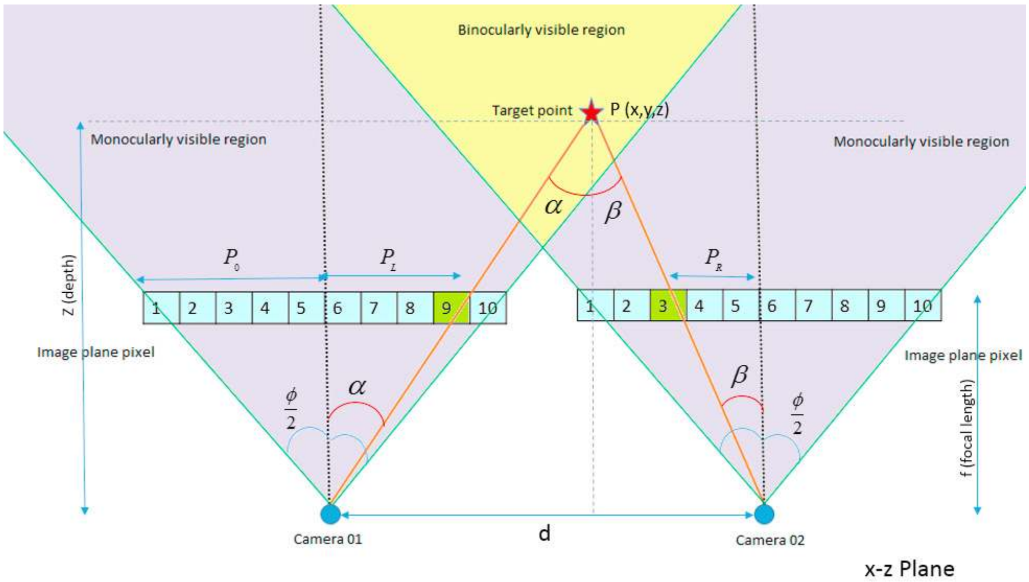


Fig. 3. Working principal of stereo vision system viewing from x-z plane

distortion, having the following variables:

- d = distance between two identical cameras (fixed)
- ϕ = Field of view for each camera. (For our case its 60°)
- f = focal length of camera (4.40mm)
- P (x, y, z) = Target point
- z = depth of target point from camera.

Disparity can be represented by difference between the pixel distances of both cameras seeing the target point. i.e. Disparity $(D) = P_L(P_0 - P_R) = 9 - 3 = 6$ Let α and β be the angle subtended by principal axis of camera 1 and camera 2 towards the target point respectively. Considering pure translation in x-z plane.

$$\text{For left camera image plane : } \frac{x}{z} = \frac{P_L}{f} \tag{1}$$

$$\text{Similarly for right camera image plane : } \frac{d-x}{z} = \frac{P_R}{f} \tag{2}$$

$$\text{After eliminating the x terms from (1) and (2), the depth point can be given by : } z = \frac{f \times d}{(P_R + P_L)} \tag{3}$$

The depth point z shows the depth for a pure translation motion along x-axis. But for a practical stereo vision system, the camera behaves nonlinear because of lens placement and manufacturing defects which in turn gives the

Intrinsic parameters (eqn. 4 and eqn. 5). Intrinsic parameters are the camera internal parameters which are necessary to find the relationship between the pixel-coordinate of image plane to the corresponding camera reference frame. The Extrinsic parameters (eqn. 6) decide the relationship between camera frame of reference and world reference frame to calibrate the stereo system. The final transformation matrix will be the multiplication of both internal and external parameters.

For the proposed stereo vision system, calibration is done with 120 sampled stereo images of a chess-board pattern using MATLAB. The intrinsic parameter matrix used for camera 1 and camera 2 is given by:

$$I_1 = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 701.19 & 0 & 312.32 \\ 0 & 702.29 & 178.14 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

$$I_2 = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 704.40 & 0 & 313.73 \\ 0 & 705.21 & 188.60 \\ 0 & 0 & 1 \end{pmatrix} \tag{5}$$

The extrinsic parameter matrix for camera 2 with respect to camera 1 is given by:

$$I_2 = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.9998 & 0.0087 & 0.0162 & -129.7029 \\ -0.0084 & 0.9998 & -0.0166 & -0.9005 \\ -0.0164 & 0.0165 & 0.9997 & 1.6368 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6}$$

Using these parameters the stereo vision system is calibrated with the Mean-Reprojection error of 0.0502. Once the camera is calibrated, the depth map is generated using point cloud mapping (Table 1 shows the movement of object in x-z plane at 100^{th}, 200^{th} and 300^{th} frame). The main drawbacks of using stereo vision system alone is: the ambiguous correspondence problem between points in the two images which may lead to several different consistent interpretations of the scene, and the reconstruction problem which in-turn depends on disparity and the correspondence points. The dynamic environment increases the correspondence problem. Consequently, KLT feature tracker based optical flow method is used to detect the dynamic surrounding and target point movement.

3.2. Kanade-Lucas-Tomasi feature tracker for motion detection

The KLT method [20] can be used to study a variety of motions such as: dynamic observer and static environment, static observer and dynamic environment, or a combination of the two, however it does not give the exact motion trajectories. Instead, it gives the motion direction which significantly helps in finding the exact displacement of target, when implemented along with SV. However, in this method the feature points defined in the initial frame is required to lie in all the frames captured consecutively thereafter. Consequently, the problems stated above motivated us to make use of stereo vision system along with KLT tracker.

The KLT motion tracker method is based on two main assumptions: the tracking objects is a rigid body, and the environment has constant visual texture. For large tracking framework, the feature point tracking method is used. The point tracker object tracks a predefined set number of points using the Kanade-Lucas-Tomasi (KLT) feature-tracking algorithm. As the point tracker algorithm progresses, certain points can be lost due to varied illumination, distinct motion, etc. To rectify these problems, re-acquirement of points is periodically carried out to track an object over a long period of time. KLT method tries to find a shift in an interest point. To search for the position that yields the best match, KLT uses spatial intensity information which leads to potentially faster matches between consecutive images. The framework is based on local optimization, usually, a squared distance criterion for a local region was optimized with respect to the transformation parameters. Using Taylor series, the approximation of the feature displacement is carried out. This framework also solves more realistic transformations (considering rotation/general affine transformations, etc.).

For a brief explanation of KLT method, two consecutive images i.e. n and (n+1) are taken. Fig. 4(a) shows the movement of a feature point in those frames. In order to do loss-less tracking, KLT method uses forward and backward error thresholding up to a margin of 3 pixels. The maximum bidirectional error is specified and if this value is less than default infinity value, the object tracks each point from the previous frame to the current frame. It then tracks the same points back to the previous frame and calculates the bidirectional error (shown in Fig. 4(b)).

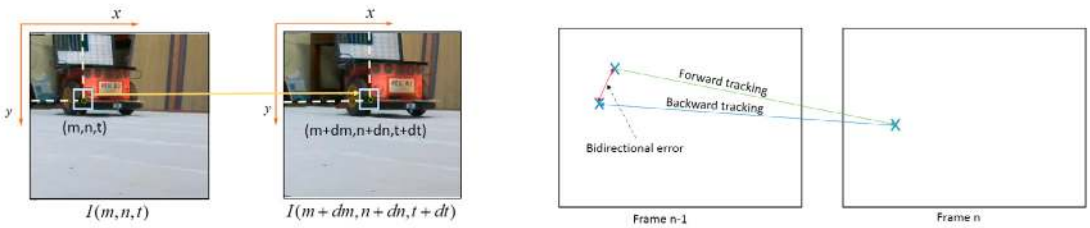
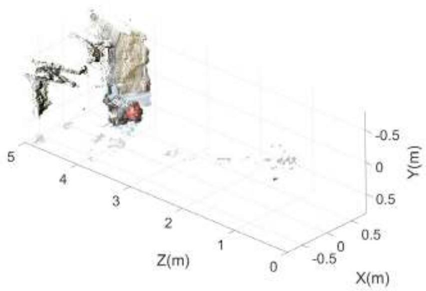
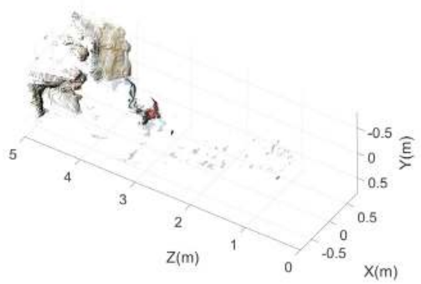
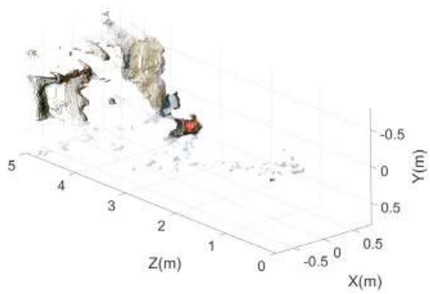


Fig. 4. (a) Frame-by-frame displacement of feature point; (b) Bidirectional error.

Table 1. Feature points and point cloud mapping of corresponding frames:

Frame sample	Feature tracking output	Depth map by stereo vision
100^{th} frame		
200^{th} frame		
300^{th} frame		

3.3. Fusion of Kanade-Lucas-Tomasi feature tracker and Point load method for motion detection

Initially, using KLT feature tracking method, the feature points are identified on the target. Subsequently, frame-by-frame, the corrected motion of those feature points are estimated. Simultaneously, the calibrated stereo vision system is used to detect the depth point at each frame. The point cloud map generated was then compared with the tracking point movement to find common co-ordinate points between them. This method is better in comparison to a normal stereo vision method because, for a particular target, there will always be certain common feature points that appear in every frame captured. The reverse depth generation of these tracked common feature points will provide the distance of vehicle at any instant. The experimental validation of this method is discussed in the following section.

4. Result and discussion

The proposed algorithm was tested with the P3-DX mobile robot as a target to be tracked. Table 1 shows the feature points tracked and the corresponding stereo points in 100th, 200th and 300th frame. 100th frame is when the vehicle starts accelerating while 200th and 300th frame shows the constantly accelerated vehicle moving at 17 cm/s. It is clear from the table that if the speed of vehicle increases there is a possibility of loss of feature points. Since the new algorithm was tested with low cost camera at 30 fps, the loss of feature points can be seen easily in 200th and 300th frames. The robot is made to move along various predefined trajectories. Fig. 5(a) shows the actual paths followed by the target. Corresponding to these actual trajectories, the estimated trajectories calculated from the algorithm are shown in Fig. 5(b). The result shows that there is a minor deviation in the trajectories of the actual path followed by the target and trajectory estimated by the proposed algorithm. The tracking path shown by purple color shows that the more be deviation from camera view point, more will be deviations from actual path, causing increased error values. Hence, the proposed algorithm can be used for obstacle detection and navigation. Fig.6 (a) shows the tracking of target feature points by the developed algorithm while the target is following path-1. Following this path makes the target move strictly in the direction of optical axis of cameras and as a result, the trajectory error (error between the estimated trajectory and actual trajectory) is minimum. As the angle between the optical axis of the cameras and the motion direction of target increases, the trajectory error increases. This is clear from Fig. 6(b).

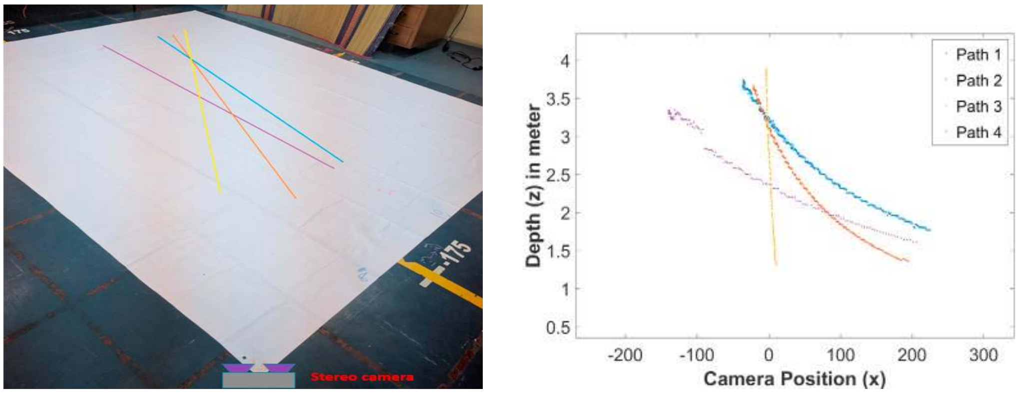


Fig. 5. (a) Original track for P3-DX movement; (b)Path followed by a tracking point.

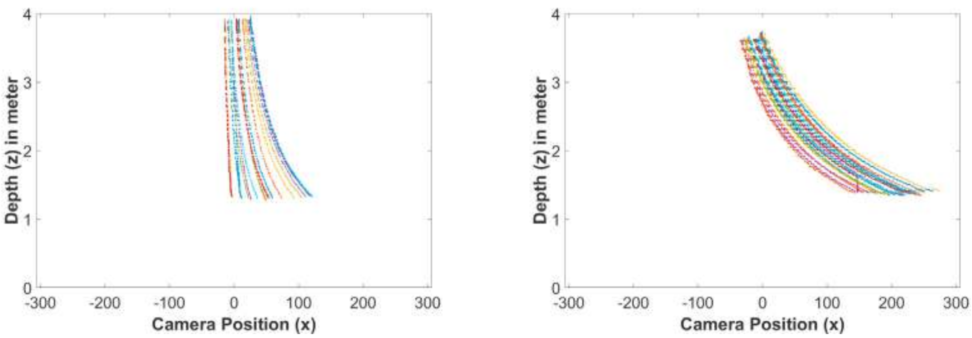


Fig. 6. Tracked feature points for vehicle (a) approaching stereo camera (b) moving alongside.

5. Conclusion

The paper describes the development and implementation of a new algorithm that integrates Stereo Vision and feature point tracking using KLT techniques, thereby increasing the robot's ability to estimate the path followed by the target. Using Stereo Vision method, we successfully found the point cloud mapping which provides the depth map of environment leaving the motion detection. By using KLT method, we are able to get the direction of motion of target vehicle without depth data. So, each method is limited by its own constraints. In this paper, we are successful in integrating both the techniques to develop a method for estimating the motion trajectory of target points. The new method uses property of both the methods and gives depth mapping along with the motion detection for a dynamic obstacle. The developed methodology is however limited to only perceive targets that are beyond one meter from the observation point. This perception distance can be reduced further by the use of cameras with larger field of view (FoV) to get more binocularly visible regions. The work in this paper is limited to tracking of dynamic obstacles, keeping the view point stationary. This method will be subsequently extended for the implementation for the case of dynamic observer and dynamic target along with the performance analysis.

References

[1] Hsia, K. H., Guo, J. H., and Su, K. L. (2013, December). "Motion guidance of mobile robot using laser range finder". In *Fuzzy Theory and Its Applications* (iFUZZY), 2013 International Conference on IEEE, pp. 293-298.

[2] Chong, T. J., Tang, X. J., Leng, C. H., Yogeswaran, M., Ng, O. E., and Chong, Y. Z. (2015). "Sensor technologies and simultaneous localization and mapping (SLAM)". *Procedia Computer Science*, 76, 174-179.

[3] Lianghong, D., Runxiao, W., Qingjuan, D., and Dongbo, W. (2012, May)."Research on visual odometry and terrain reconstruction for quadruped robot". In *Computer Science and Automation Engineering* (CSAE), 2012 IEEE International Conference on IEEE, Vol. 2, pp. 551-555.

[4] Kolter, J. Z., Kim, Y., and Ng, A. Y. (2009, May). "Stereo vision and terrain modeling for quadruped robots". In *Robotics and Automation*, 2009. ICRA'09. IEEE International Conference on IEEE, pp. 1557-1564.

[5] Lee, Y. J., Yim, B. D., and Song, J. B. (2009). "Mobile robot localization based on effective combination of vision and range sensors". *International Journal of Control*, Automation and Systems, 7(1), 97-104.

[6] Wei, L., Cappelle, C., and Ruichek, Y. (2011, October). "Unscented information filter based multi-sensor data fusion using stereo camera, laser range finder and GPS receiver for vehicle localization". In *Intelligent Transportation Systems* (ITSC), 2011 14th International IEEE Conference on IEEE, pp. 1923-1928.

[7] Nalpantidis, L., and Gasteratos, A. (2010). "Stereo vision for robotic applications in the presence of non-ideal lighting conditions". *Image and Vision Computing*, 28(6), 940-951.

[8] Saez, J. M., and Escolano, F. (2004, April). "A global 3D map-building approach using stereo vision". In *Robotics and Automation*, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on IEEE, Vol. 2, pp. 1197-1202.

[9] Li, Z., Shi, Y., Wang, C., and Wang, Y. (2008). "Accurate calibration method for a structured light system". *Optical Engineering*, 47(5), 053604

[10] Pan, H., Guan, T., Luo, Y., Duan, L., Tian, Y., Yi, L., ... and Yu, J. (2016). "Dense 3D reconstruction combining depth and RGB information". *Neurocomputing*, 175, 644-651.

[11] Khoshelham, K., and Elberink, S. O. (2012). "Accuracy and resolution of kinect depth data for indoor mapping applications". *Sensors*, 12(2), 1437-1454

[12] Yue, H., Chen, W., Wu, X., and Zhang, J. (2013, June). "Kinect based real time obstacle detection for legged robots in complex environments". In *Industrial Electronics and Applications* (ICIEA), 2013 8th IEEE Conference on IEEE, pp. 205-210.

[13] Kazmi, W., Foix, S., Aleny, G., and Andersen, H. J. (2014). "Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison". *ISPRS journal of photogrammetry and remote sensing*, 88, 128-146

[14] Souhila, K., and Karim, A. (2007). "Optical flow based robot obstacle avoidance". *International Journal of Advanced Robotic Systems*, 4(1), 2.

[15] Kuo, Y. C., Pai, N. S., and Li, Y. F. (2011). "Vision-based vehicle detection for a driver assistance system". *Computers and Mathematics with Applications*, 61(8), 2096-2100.

[16] Keshavan, J., and Humbert, J. S. (2017). "An optical flow-based solution to the problem of range identification in perspective vision systems". *Journal of Intelligent and Robotic Systems*, 85(3-4), 651-662.

[17] Dorini, L. B., and Goldenstein, S. K. (2011). "Unscented feature tracking". *Computer Vision and Image Understanding*, 115(1), 8-15.

[18] Maradkar, M. S., Chandramouli, A., and Manivannan, P. V. (2017, February). "Bio-inspired Reconfigurable Robot: Conceptual Design of an All-terrain Robot Capable of Transforming from an Erect to Sprawling Posture". In Proceedings of the 3rd International Conference on *Mechatronics and Robotics Engineering*, ACM, pp. 69-74.

[19] Geiger, A., Ziegler, J., and Stiller, C. (2011, June). "Stereoscan: Dense 3d reconstruction in real-time", In *Intelligent Vehicles Symposium* (IV), IEEE, pp. 963-968.

[20] Trummer, M., Denzler, J., and Munkelt, C. (2008). "KLT Tracking Using Intrinsic and Extrinsic Camera Parameters in Consideration of Uncertainty", In *VISAPP* (2), pp. 346-351.