A Review on Vision Simultaneous Localization and Mapping (VSLAM)

Jabulani K. Makhubela, Tranos Zuva and Olusanya Yinka Agunbiade
Faculty of Applied and Computer Sciences
Dept of Information and Communication, Vaal University of Technology
Vanderbijlpark, South Africa
kingsley.makhubela@gmail.com

***Abstract*—Simultaneous Localization and Mapping (SLAM) has seen a tremendous interest amongst research community in recent years due to its ability to make the robot truly independent in navigation. The capability of an autonomous robot to locate itself within the environment and construct a map at the same time, it is known as Simultaneous Localization and Mapping (SLAM). They are various sensors that are employed in a Simultaneous Localization and Mapping (SLAM) which characterized either as a laser, sonar and vision sensor. Visual Simultaneous Localization and mapping (VSLAM) is when autonomous robot embedded with a vision sensor such as monocular, stereo vision, omnidirectional or Red Green Blue Depth (RGBD) camera to localize and map the environment. Numerous researchers have embarked on the study of Visual Simultaneous Localization and Mapping (VSLAM) with incredible results, however many challenges stills exist. The purpose of this paper is to review the work done by some of the researchers in Visual Simultaneous Localization and Mapping (VSLAM). We conducted a literature survey on several studies and outlined the frameworks, challenges and limitation of these studies. Open issues, challenges and future research in Visual Simultaneous Localization and Mapping (VSLAM) are also discussed.**

Keywords—Navigation; Sensors; Vision; Illumination variance; Simultaneous Localization and mapping (SLAM)

I. INTRODUCTION

Simultaneous Localization and Mapping (SLAM) has captured a great deal of attention within the research community during the past recent years because of it is potential to make robot truly autonomous [1]. Visual Simultaneous localization and Mapping (VSLAM) is when a robot can independently estimate its position within environment and able to draw a map of the same environment, by utilizes vision sensor such as camera, Red Green Blue Depth (RGBD) sensor etc. [2]. Choosing a sensor for autonomous robot such as Laser Finders (LRFs), sonar, acoustic, cameras (monocular, vision stereo or omnidirectional), Red Green Blue Depth (RGBD) Sensor such as Microsoft Kinect and PrimeSense has become a critical part of the SLAM technique [3].

According to [4] vision sensors are utilized in various robotic systems like object recognition, obstacle avoidance, topological global localization. The reason for this is because vision sensor over the other sensors are potable, less expensive, compact, precise, low-priced, non-invasive and pervasive [5].

Vision sensors are able to exact more and viable information both in color and per-pixel about location than any other sensor [6]. Vision sensors are favored because people and animals seem to be navigating effectively in complicated locations using vision as prime sensor [5].

Various researchers have embarked on Visual Simultaneous Localization and Mapping (VSLAM) with exceptional results, however many challenges still exist in Visual Simultaneous Localization and Mapping (VSLAM). This paper will be reviewing the methods, achievements and the limitations on studies done on visual Simultaneous Localization and mapping (VSLAM) by some of these researchers. A reminder of this paper is organized as follows: - section II discusses a review of studies done by some of the researchers. Section III discuss open issues and challenges which still exists on studies reviewed and finally conclusion is drawn in section IV.

II. RECENT RESEARCH ON VSLAM

Visual Simultaneous Localization and Mapping (VSLAM) is when an autonomous robot use service of camera as exteroceptive sensor to navigate, map the location and localize itself. [2]. This section will focus some of the research done by other researchers under a Visual Simultaneous Localization and Mapping (VSLAM), their achievements and limitations they face when implementing their method.

A proposed method by [7] on Stereo Vision Simultaneous Localization and Mapping (VSLAM) for autonomous mobile robot navigation in an indoor location. The objective was to design a system in which an autonomous robot would exclusively utilize a vision sensor for acquiring data and navigating the environment. Their navigation system comprised of navigation and self-localization. The overall navigation hierarchy contained of localization, Perform the Region of Interest (ROI), Region of Interest (ROI) Sub Screening, grid mapping optimal path search and path planning as illustration in Figure 1. The routine activities into their Visual Simultaneous Localization and Mapping (SLAM) navigation system was to achieve 3D depth calculation of the location, scene analysis, optimal path search, real time path planning and motor speed control.

978-1-5386-6477-3/18/$31.00 ©2018 IEEE

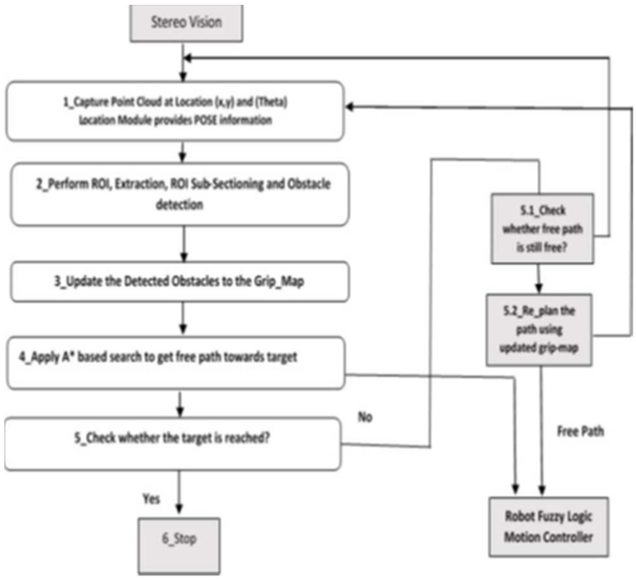

Fig 1. A proposed SLAM Navigation Strategy [7]

The experimental results showed their proposed system was capable of navigating in a pre-planned paths environment positively. However, the data association on the mobile robot became a challenge due to variations in illumination, specular reflection in environment and inconsistent point clouds which happened because of the variation in viewing angles of the camera.

In the study conducted by [8], for Visual Simultaneous Localization and Mapping (VSLAM) based Indoor/ Outdoor environment obstacle avoidance. The purpose of the study was to create an on-broad stereo-vision based mapping system that would be used for path planning and local obstacle avoidance in a search and rescue mission. Their system design consisted of three layers which was perception, mapping as well planning and control. The perception layer included dense stereo matching, visual odometry computation and sensor fusion. The mapping layer included negative edge detection which assisted in identifying negative slopes and edges in the environment, like cliffs or stair-heads; to further improve on their technique, the adaptive step and slop detection was utilized to convert the depth image into a point-cloud that was associated to the local tangent plane; Outlier Filtering was deployed to reduce the standardization, propagated and stereo correlation error in a 3D positions of the detected obstacles; Time-Based probabilistic Integration was employed to surpass the small aperture angle, and Rao-Blackwellized Particle Filters (RBPF) algorithm it was used to accurately calculate the factorization of jointly probability of the robot route and the environment model into distribution possible robot paths or route. The results of localization and model calculation would be employed by planning and control layer for robot control, local path planning and obstacle avoidance. Figure 2 gives illustration of how the Simultaneous Localization and Mapping (SLAM) system architecture was developed.

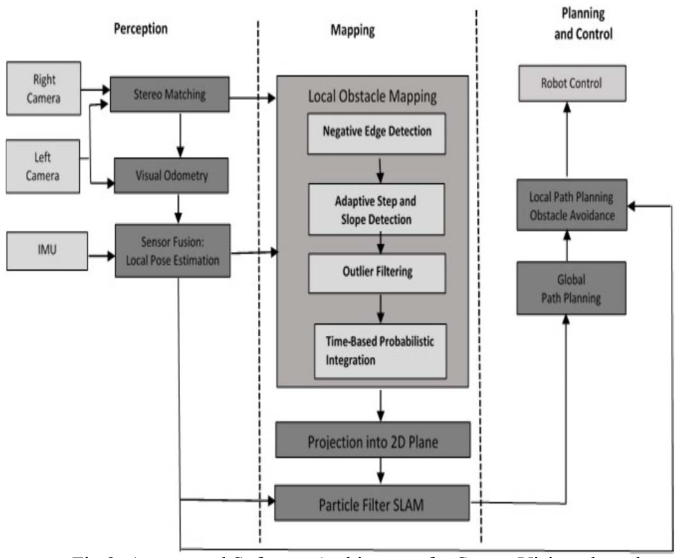

Fig 2. A proposed Software Architecture for Stereo-Vision obstacle mapping [8]

The experiments of the proposed system design were performed in an indoor, outdoor and mixed Indoor & outdoor environment. In outdoor set-up the narrow-angle and wide-angle Vision Simultaneous Localization and Mapping (VSLAM) system depicted an above precision of 0.22m mean fault, having an improved deviation of less than 0.08% and 0.06% of the full course meaning that the Vision Simultaneous Localization and Mapping (VSLAM) system was capable to handle inaccuracies, drifting fusion calculations and to rectify the drifts as well. A Vision Simultaneous Localization and Mapping (VSLAM) system also showed favorable progress in closing the large loop in an experimental set-up. The Vision Simultaneous Localization and Mapping (VSLAM) system demonstrated that it could manage with changing light mixed Indoor & Outdoor set-up. However, the system couldn't perform a quantitative evaluation of obstacle map because of lack of ground truth data, the un-textured objects like white walls, regular patterns and reflective surfaces became a problem for the system architecture in an outdoor Environment.

A proposed framework by [9] was presented on Visual Simultaneous Localization and Mapping (VSLAM) in low dynamic workspace environment using a Red Green Blue Depth (RGBD) Sensor, the aim of the framework was to update a map by keeping track on latest changes on environment as the autonomous robot was navigating the location. The proposed framework had two components which were multi-Session Visual Simultaneous Localization and Mapping (VSLAM) and graph management. The multisession Visual Simultaneous Localization and Mapping (VSLAM) module had a graph model with each landmark being a position and each edge being a restriction, by combining previous sessions with the current information and recent session to keep the model of the environment in one global coordinates. The graph management was responsible for updating the graph model and without increasing difficulties employing the out-of-dated scan identification module and redundant scan identification module. Figure 3 gives illustration of the framework.

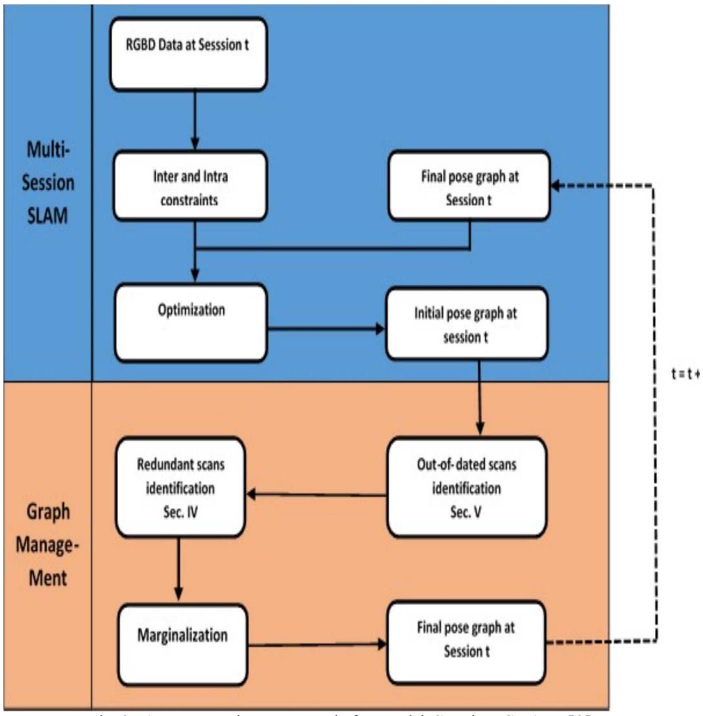

Fig 3. A proposed Framework for multi-Session SLAM [9]

The activities in each timestep was: (1) The multi-session Visual Simultaneous Localization and Mapping (SLAM) module creates model of the environment by employing the Red Green Blue Depth (RGBD) camera information; (2) The out-of-dated scans identification unit was to recognize whether the scans were consistent with previous poses were out of date, if so, the nodes would be removed since they were no longer valuable for creating model of environment and loop closure; (3) The redundant scans identification module would continue removing positions if the number of in-dated nodes was more than a threshold, which was linked to the quantity of the modelling area; (4) The graph was marginalized to reserve the information after the node was removed, establishing an combined restriction for the next session .

The experimental results showed that the proposed framework was able to update the map in a dynamic or static environment without increasing difficulties and at an acceptable error level, however they were setback on the system for out-of-dated scan identification tested in the lab and industrial workspace the small object captured by Red Green Blue Depth (RGBD) sensor with low quality, were considered as noise by the algorithm, two object were regarded as one object and when the object is removed and the new one is added in the position and such change cannot be reflected on the geometric shape.

In reference to a study done by [3] for Robust Visual Simultaneous Localization and Mapping (VSLAM) using planar point's features with the aid of a Red Green Blue Depth (RGBD) sensor. The objective of the study was to propose a method that would improve the accuracy in capturing and recording point's features into the Visual Simultaneous Localization and Mapping (VSLAM). Their proposed method utilized a planar feature to align key frames, the proposed method would start by removing planes points from cloud obtained the Red Green Blue Depth (RGBD) sensor, then attempts improve the quality of the mined planes, then depth values of planes would be modified according to the parameter of plane models and Random Sample Consensus (RANSAC) framework is then employed to sense and compare point features on individual pair of planes. According to [10] Random Sample Consensus (RANSAC) is an overall method for correcting model in the existence of outliers. Their proposed method also contained a graphic end part which executed image process ego-motion calculation and pose optimization part was for executing global pose graph optimization. The illustration of system is given in Figure 4.

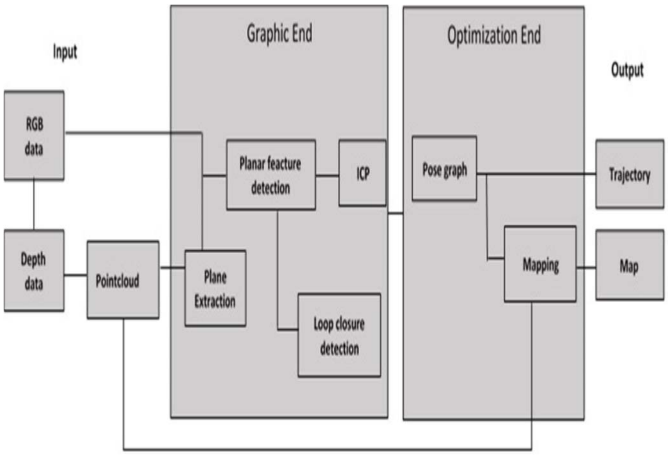

Figure 4. A proposed SLAM Planar Features System [3]

The experimental results of a proposed method on open datasets showed tracking trajectory is fairly accurate, high accuracy in motion of the two fames with less failure rate. However physical experiment results showed repetitive structure on floor, daylight from the ceiling disrupted the matching process and sunshine from window disrupts the illumination conditions.

A proposed algorithm by [11], for robust outdoor stereo Vision Simultaneous Localization and Mapping (VSLAM) for heavy machine rotation sensing. The purpose of the algorithm would be to calculate a mining rope shovel's rotation about its vertical axis. Extended Kalman Filter (EKF) was selected the main algorithm in this study because of its abilities to close loop successfully. [12] defines Extended Kalman Filter as an algorithm that is utilized in a number of nonlinear estimation and machine learning systems as well as calculating the state of a nonlinear dynamic application, calculating parameters for nonlinear system identification when system dynamic and observation models are linear. For every camera frame, based on Extended Kalman Filter (EKF) Simultaneous Localization and Mapping (SLAM), the following processes should take place in their algorithm as demonstrated in Figure 5.

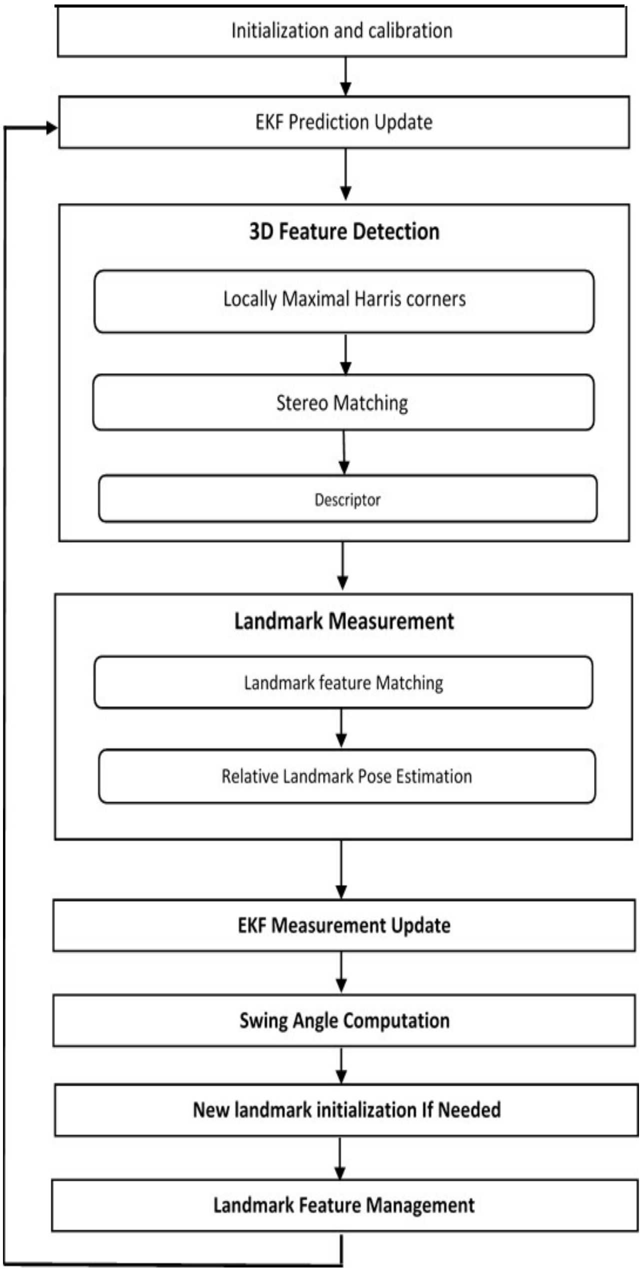

Fig 5. Algorithm flow for robust outdoor stereo vision SLAM [11]

According to the study, the proposed algorithm also employed the two methods to increase Visual Simultaneous Localization and Mapping (VSLAM) effectiveness for outdoor environment. The techniques were Harris corner (Harris & Stephen as cited by [11]), namely locally maximal feature and Feature Cluster landmark. Locally maximal feature chose features more evenly across the image and senses features that consistently in a non-uniformly illuminated scene, Whereas Feature Cluster landmark acquired each landmark in a cluster of 3D feature from a single camera frame, this will permit extremely consistence and robust landmark measurement. As corners were designated based corner threshold within an image. The study also used a percent threshold method to measure the pixel locations of edge that surpasses the corner threshold within an image.

The experiment of a vision algorithm was performed using three separate video sequence, at round 2900, 8000 and 6600 frames respectively. The maximum error for all video sequence settled within 1^o. When compared two methods used in a vision algorithm, it was a discovered that the Local maximum method was steadier than Percent Threshold method detecting frame with direct sunlight or shadows illumination.

A framework was proposed by [13] for Collaborative Visual Simultaneous Localization and Mapping (CoSLAM) in a dynamic Environment, the aim of the framework was to address the challenge of dynamic environments by employing numerous cameras into Visual Simultaneous Localization and Mapping (VSLAM) system. The Collaborative Visual Simultaneous Localization and Mapping (CoSLAM) framework would sense and detect feature points of each image from individual camera sensor inputs and provide them into Collaborative Visual Simultaneous Localization and Mapping (CoSLAM) components which are camera pose estimation, map building, point classification and camera grouping which were combined to create a global map. Collaborative Visual Simultaneous Localization and Mapping (CoSLAM) framework would at the same time calculate the position of all camera sensors continually. Figure 6 gives illustration of the framework.

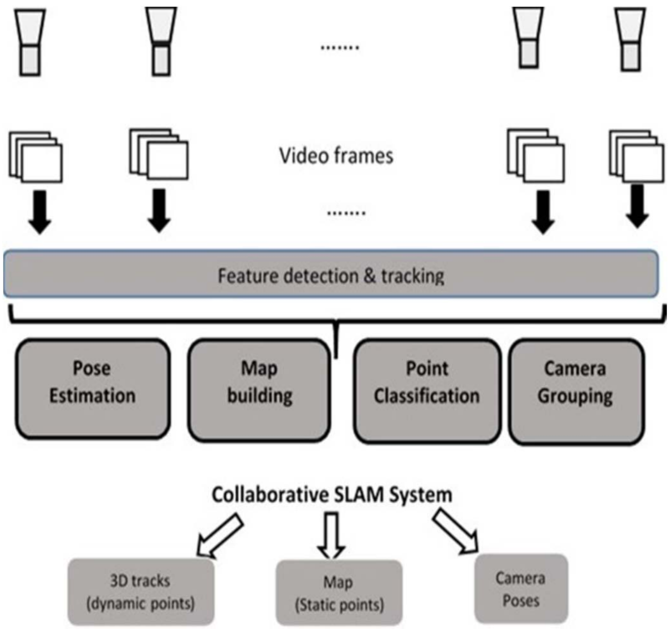

Fig 6. A Proposed COSLAM system architecture [13]

The experimental results showed that the Collaborative Visual Simultaneous Localization and Mapping (CoVSLAM) system was capable effectively calculating position in extremely dynamic and static environment. Collaborative Visual Simultaneous Localization and Mapping (CoVSLAM) was successful in handling with lighting and noise condition in an indoor location. However Collaborative Visual Simultaneous Localization and Mapping (CoVSLAM) was using twelve camera sensor map and localize environment, the performance of the system dropped because of heavy computation of sensors

III. OPEN ISSUES, CHALLENGES AND FUTURE RESEARCH IN VSLAM

Visual Simultaneous Localization and Mapping (VSLAM) over the years have receive a lot of attention and tremendous achievement has been attained. But there are still many issues limiting the full acceptance of this technology. Take for instance the issue of dynamic environment complain by many researchers has not been fully resolved [9], [13]. Dynamic features are object capable of changing their location as a result of external force acting on it. Thus, success attained in this situation is associated with how dynamic the environment is. The more dynamic the environment, the more challenging it becomes to address [9]. Another issue that has not been fully addressed is the illumination variance which mostly affect the image characteristic, with an effect that violate assumption making localization impossible. But in the static environment the effect of Illumination variance has been successful to some extent but still lacks accuracy [7]. However, it is more challenging to address in dynamic environment [9]. Furthermore, some researchers reviewed in this study complain about regular patterns [8], [3]. The effect of regular patterns on Visual Simultaneous Localization and mapping (VSLAM) algorithm can lead to mismatches and missing features which will result in obstacle either been absent or dislocated within the map created for the environment [8]. Lastly, the issue of computation complexity has related speed is often a problem complain by some researchers [13], [3], although researchers tend to reduce the Visual Simultaneous Localization and mapping (VSLAM) computational cost but result attained is successful to some extent [3]. Nevertheless, more research still needs to be done in this area. In future, SLAM technique must be able to eliminate all current challenges to attain maximum effectiveness and efficiency.

IV. CONCLUSION

Visual Simultaneous Localization and Mapping (VSLAM) has been a common method among robotic applications because of its ability to enable an autonomous robot to construct a model of its environment and estimate its position at the same time with the aid of vision sensor such as camera etc. This is an important functionality necessary for a robot to navigate in an unfamiliar area. Numerous researchers have researched in an area of Visual Simultaneous Localization and mapping (VSLAM) with remarkable results. However the review conducted in this study has discovered that challenges such as variations in illumination, specular reflection in environment, inconsistent point clouds, un-textured objects like white walls, regular patterns, reflective surfaces, direct sunlight or shadows illumination, loop closure detection and multiple camera sensors on the robot has the potential to limit or cause the Visual Simultaneous Localization and Mapping (VSLAM) technique to fail whether in a static or dynamic environment. In our effort to improve the Visual Simultaneous Localization and Mapping (VSLAM) technique, we plan to investigate the challenge of environmental noise such as shadow or light intensity/ illumination in our future work because this is often common as a result of high intensity of sunlight in our continent.

REFERENCES

[1] Ramos, F. T., Nieto, J. & Durrant-Whyte, H. F. (2007). Recognising and modelling landmarks to close loops in outdoor slam. In *Proceedings 2007 IEEE International Conference on Robotics and Automation.* p. 2036-2041. IEEE

[2] Clipp, B., Zach, C., Lim, J., Frahm, J. M. & Pollefeys, M. (2009). Adaptive, real-time visual simultaneous localization and mapping. In *Applications of Computer Vision (WACV), 2009 Workshop on.* p. 1-8. IEEE.

[3] Gao, X., & Zhang, T. (2015). Robust RGB-D simultaneous localization and mapping using planar point features. *Robotics and Autonomous Systems*, *72*, 1-14.

[4] Moreno, F. A., Blanco, J. L., & Gonzalez, J. (2009). Stereo vision specific models for particle filter-based SLAM. *Robotics and Autonomous Systems*, *57*(9), 955-970.

[5] Milford, M. J. & Wyeth, G. F. (2008). Single camera vision-only SLAM on a suburban road network. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on.* p. 3684-3689. IEEE.

[6] Park, C. & Song, J. B. (2015). Illumination change compensation and extraction of corner feature orientation for upward-looking camera-based SLAM. In *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on.* p. 224-227. IEEE.

[7] Al-Mutib, Khalid N., Et Al. "Stereo vision SLAM based indoor autonomous mobile robot navigation." *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on*. IEEE, 2014.

[8] Brand, C., Schuster, M. J., Hirschmüller, H., & Suppa, M. (2014, September). Stereo-vision based obstacle mapping for indoor/outdoor SLAM. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on* (pp. 1846-1853). IEEE.

[9] Wang, Y., Huang, S., Xiong, R., & Wu, J. (2016). A framework for multi-session RGBD SLAM in low dynamic workspace environment. *CAAI Transactions on Intelligence Technology*, *1*(1), 90-103.

[10] Raguram, R., Chum, O., Pollefeys, M., Matas, J., & Frahm, J. M. (2013). USAC: a universal framework for random sample consensus. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 2022-2038.

[11] Lin, L. H., Lawrence, P. D., & Hall, R. (2013). Robust outdoor stereo vision SLAM for heavy machine rotation sensing. *Machine vision and applications*, 1-22.

[12] Wan, E. A., & Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000* (pp. 153-158). Ieee.

[13] Zou, D., & Tan, P. (2013). Coslam: Collaborative visual slam in dynamic environments. *IEEE transactions on pattern analysis and machine intelligence*, *35*(2), 354-366.

[14] Thrun, S. (2002, August). Particle filters in robotics. In *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence* (pp. 511-518). Morgan Kaufmann Publishers Inc.

[15] Thrun, S., & Montemerlo, M. (2006). The graph SLAM algorithm with applications to large-scale mapping of urban structures. *The International Journal of Robotics Research*, *25*(5-6), 403-429.