Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 76 (2015) 174 – 179

Procedia Computer Science

2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)

Sensor Technologies and Simultaneous Localization and Mapping (SLAM)

T.J. Chong, X.J. Tang, C.H. Leng, M. Yogeswaran*, O.E. Ng, Y.Z. Chong

Universiti Tunku Abdul Rahman, UTAR Complex, Jalan Genting Kelang, 53300 Setapak, Kuala Lumpur, Malaysia.

Abstract

This paper presents a comprehensive review on sensor modalities currently in used for solving the Simultaneous Localization and Mapping (SLAM) problem. The review focuses on SLAM for mobile robots in a variety of environments. The strengths and weaknesses of acoustic modality sensors such as ultrasonic and sonar sensors, laser range finders, visual sensors such as stereo vision sensors, and RGB-D sensors like the Microsoft Kinect and the Asus Xtion Pro Live are compared based on current usage in published research papers. Based on this review, we propose that RGB-D sensors have unique advantages which make them particularly suitable for SLAM problems.

Peer-review under responsibility of organizing committee of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)

Keywords: SLAM; mapping; localization; sensors

1. Introduction

The Simultaneous Localization and Mapping (SLAM) problem can be defined as a process where a robot builds a map representing its spatial environment while keeping rack of its position within the built map. Mapping is done online with no prior knowledge of the robot's location; the built map is subsequently used by the robot for navigation. SLAM is a key component of any truly autonomous robot. Much recent research has been done tackling the computational efficiency of SLAM and the data association and landmark extraction necessary for a robust SLAM method.

Environmental mapping involves creating a mathematical model of a real environment's spatial information. SLAM extends the requirements of this mathematical model; it must also jointly represent the robot's state and the position of extracted landmarks relative to the robot's location. The robot's state includes information on the robot's position and orientation. The basic SLAM framework involves odometry, landmark prediction, landmark extraction, data association and matching, pose estimation, and map update. These processes are the backbone of every major SLAM method, and are performed in cyclic fashion.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: yogeswaran@utar.edu.my

Peer-review under responsibility of organizing committee of the 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS 2015)
doi:10.1016/j.procs.2015.12.336

SLAM algorithms must take account of a variety of parameters i.e. sensors, map representation, robot dynamics, environmental dynamics, and the integration of sensor measurements and the robot's control system over time. The integration of these diverse parameters is most often accomplished using the two major SLAM-related algorithms. Kalman filters are routinely used, with the most popular being the Extended Kalman Filter (EKF)[30,11,28]. Rao-Blackwellized Particle Filters (RBPF)[19,16] have also been used in many SLAM methods.

Mapping the spatial information of an environment requires spatial sensors to which SLAM algorithms can be applied. The two most popular sensor modalities used in SLAM are raw range scan sensors and feature (landmark) based sensors (whether extracted from scans or images)[14]. The most commonly used sensors for landmark extraction from scans are laser-based and sonar-based system. Landmark extraction from images (often referred to as visual SLAM) uses camera in a variety of configurations such as monocular configuration[4], stereo vision configuration[7,20] and multiple camera configuration[12]. Visual sensing provides information-heavy data, but with significant amounts of noise and uncertainty.

Robot localization requires sensory information regarding the position and orientation of the robot within the built map. In most simple wheeled-robot cases, this pose information can be obtained from rotary encoders which measure the rotational movement of the wheel(s). Most robots also use Global Positioning System (GPS) sensors which allow acquisition of pose information. Both of these sensor technologies are prone to uncertainty and error, either due to GPS signal strength issues or slippage and loss of traction in the robot's wheel(s).

For the past decade, SLAM methods have progressed rapidly due in part to the rapid advancement in sensor technology. Different SLAM applications require different types of sensors suitable to the unique environments the robot's will encounter. SLAM applications have been studied on ground robots, aerial robots, underwater robots, and even a combination land and aerial robot[3]. This paper reviews the most significant sensors currently in use for SLAM research and applications.

2. Sensors

One of the criteria for an autonomous robot is the ability to sense its environment. The robot's sensors transduce environmental conditions into signals suitable for processing by the robot. Proper sensor selection is crucial as it affects the quality and quantity of environmental information available to the robot and subsequently determines what SLAM approach is most suitable to be used.

2.1. Accoustic Sensors

Sonar sensors are mostly used underwater where laser and visual sensors struggle. Lower frequency sonar minimizes absorption, and sonar provides much better resolution in a subsea environment. However, the monotony of subsea regions means sonar depth information is much harder to interpret with high angular uncertainty[1].

Ultrasonic sensors are generally the cheapest available source of spatial sensing for mobile robots. They are compatible with most surface types, whether metal or non-metal, clean or opaque, as long as the surface measured has sufficient accoustic reflectivity. However, low spatial resolution and sensing range as well as sensitivity to environmental factors and slow response speeds hamper robotic use of ultrasonic sensors[2].

Nagla, KS and Uddin, Moin and Singh, Dilbag[22] pointed out the possible errors faced by ultrasonic sensor applications such as specular reflection, proposing a fuzzy logic based approach to reducing this error.. Specular reflection is a phenomenon where the sound wave emitted hits the target object at an acute surface angle and bounces away from the sensor.

2.2. Laser Range Finders

Laser-based systems are one of the most popular choices for solving the SLAM problem. Laser-based systems are able to obtain robust results in both indoor and outdoor environments. A comparison of line extraction methods for an indoor mobile robot using 2D laser range finder has been done by Nguyen et al.[23]

The high speed and high accuracy of laser range finders enable them to generate highly precise distance measurements. This contributes to the significant popularity of laser range finders in research. In Srinivasan and Lumia[26], the authors developed on low-cost laser range finder for robot application, using sinusoidally varying intensity distribution. The result of this minor modification was significantly improved information at a low incremental cost. This allowed obtaining information such as shape without the use of a separate camera system.

Surmann et al. [27] presented one of the popular method of acquiring 3D information from a 2D laser range finder. Basically, a 3D laser range scanner was built by mounting the 2D laser range finder on a vertically rotating servomotor with 120° range. The authors found that this sensor had numerous errors leading to errors and imprecisions in SLAM applications, concluding that the geometric structure of the overlapping 3D scans had to be considered in their registration problem.

Using a similar setup, [31] find that synchronization of the raw scan and the driving servo motor proves to be imperative to the scanning of the 3D environment. The authors propose a serious of patch-scan, yaw-scan, and roll-scan using a 3D laser range finder. The scanning angle is up to 180°, and the multiple scans provide the axis of restoration to a 3D laser range finder, while being slower than 2D scanning

2.3. Stereo Vision Sensors

Vision sensors can be used to estimate 3D structure (allowing for spatial information extraction), feature location, and robot pose using monocular or stereo cameras. [15] compare monocular and stereo vision systems in SLAM. Stereo cameras gain sparse distance information from disparity in textured areas of the image. Monocular cameras, on the other hand, obtain depth information of an object by repeatedly observing features to get the feature's parallax. It is worth noting that similar techniques can be applied to stereo vision cameras as well. Visual SLAM normally extracts sparse key-points from camera images using detectors and descriptors such as the Scale Invariant Feature Transform (SIFT) [17]. These key-points are more distinctive than typical geometric structures such as corners and edges.

Stereo vision systems construct 3D information from two or more 2D images. Images can be obtained from many sources, for instance two cameras located at defined relative position [21], or one moving camera with servo motor or other rotating actuator [24], or even one stereo camera which consists of at least two optical lenses [6]. Information such as features or distances (in a depth map) can be obtained from multiple images captured by cameras.

Stereo vision systems were also studied by Mustafah et al. [21] for indoor position estimation of an unmanned aerial vehicle (UAV). Their system utilized two video cameras for stereo vision capture and a set of algorithms to perform real time computation on a wireless remote image processing platform. Stereo image capture was accomplished by using two cameras placed at fixed positions and in parallel. The overlap of images for both cameras allowed accurate estimation of the UAV's position.

Another practical stereo vision system was installed on a robot [6]. The system utilized feature based SLAM as the robot was required to observe visual landmarks in the environment. Landmarks were first selected for detectability with varying distance and viewing angle, then converted to feature vectors. Unexplored areas produced new landmarks, while entering a previously explored area would lead to landmark association. The authors reported that landmarks which could not be detected from different view-poitns would lead to a failure in verifying previously explored maps, and that invariant descriptions to viewing angles and distances were complex and difficult to achieve.

2.4. RGB-D Sensors

RGB-D depth sensors project structured infrared spectrum light which is then perceived by a small baseline infrared camera. Structured light sensors are sensitive to external illumination, hence they are not usable under direct sunlight. A study on the accuracy and resolution of Kinect depth data was done by Khoshelham and Elberink [13]. The Kinect sensor integrates depth and colour data which results in a coloured point cloud that contains about 300,000 points per frame. Further observations can then increase the size of this point cloud as new features are observed in the environment. Khoshelham and Elberink [13] measured the systemic error and random error of the Kinect depth sensor and devised a calibration method to counter the observed sources of inaccuracy. In parallel, Endres et al. [5] developed the first SLAM system specifically designed for Kinect-style sensors. A detailed review of the Kinect sensor was done by Han et al. [10] which explained the working principle of the depth information capture system and demonstrated the potential of Kinect sensor.

The Asus Xtion Pro Live offers the same capability of the Kinect sensor in a more compact and lightweight package. Gonzalez-Jorge et al. [8] benchmarked the Asus Xtion Pro Live sensor and Microsoft Kinect sensor. Both of the sensors contain two CMOS sensors for RGB imaging and depth sensing. The Xtion sensor is physically smaller than the Kinect sensor. Besides that, Xtion weighs 0.225kg whereas Kinect weighs 1.36kg. The Xtion sensor is powered by a single USB power whereas the Kinect requires a separate ACDC power supply. The increased size, weight, and power requirements for the Kinect were partially attributed to the inclusion of a servomotor to allow rotation of the camera on its base.

Current RGB-D sensors use a more robust method of obtaining depth information compared to that of traditional infrared sensors, known as active stereo[5]. The passive stereo approach can be fragile for indoor depth sensing due to reliance on matching appearances and failures at smooth textureless regions, as well as a dependence on external lighting conditions which tend to be poor in indoor environments[25]. Active stereo approaches use structured light to produced a textured scene, generating fewer false positives. This approach is primarily focused on indoor usage, but could also potentially work outdoors by taking advantage of natural as well as projected texture.

In[32], the authors used an Asus Xtion Pro Live as a hand-held range sensor to reconstruct a complex real world scene. The overall idea of the project was to combine a frame-to-model registration using an offline environment framework with optimization that is able to solve loop closures and generate a globally consistent reconstruction. The operator was required to move the sensor around the target environment to obtain the image of the surrounding environment. At the same time, a smart phone which was connected to a laptop wirelessly was also carried by the operator. The smart phone was meant to show the color and depth input information and preview of the reconstruction. The laptop was mainly for offline registration and integration pipeline to produce the detailed scene model. The authors used the concept of points of interest in the scene to model the scene and preserve the detailed geometry in the scene. The major limitation of this project lay on the heart of the author's approach. The authors assumed that the errors of input could be compensated by extra careful estimation of the camera trajectory. However, the assumption was not always true, as the range images produced by consumer-grade sensors could faced problems such as substantial low-frequency distortion.

Teichman and Thrun[29]'s project implemented a solution to segment and track deformable objects by using a RGB-D sensor, with minimum assumptions made in the input stage. The main purpose of the authors' project was to segment and track non-rigid objects or moving objects. Users were required to provide some initial foreground and background labels to the system via a laptop. According to the authors, the initial segmentation labels also could be done autonomously by the system, depending on the application. The system didn't require assumptions similar to that of an ordinary segmentation and tracking system. Any object could be labeled as object of interest in the segmentation and tracking system at the initial stage of the system. The first frame of each sequence (A to H) was the seed frame. The users labeled the object of interest to be segmented and tracked in the seed frame. Then, the system started to keep track of the object of interest and showed the object of interest in every frame. The authors highlighted that their system was capable of dealing with significant non-rigid object transformations and also scenes that lack a visually distinct appearance. One of the drawbacks of the presented system was the tradeoff between stability and permissiveness, which in some cases resulted in an unoccluded, disconnected set of points near the object of interest being part of the object or background erronously. Rapid motion of the object of interest would also aggravate this problem. Additionally, more work was necessary to extend this work for objects such as the human body which contains lots of rapidly self-occluding and self-unoccluding articulated parts. Finally, thin parts or thin objects always caused problems due to the nature of the utilized edge detection methods.

Apart from the general use of RGB-D sensors in mapping, segmentation, or tracking on individual systems, both[9] and[18] involve integration of an RGB-D sensor with an autonomous robot system. The NAO Humanoid Robot[9] was studied comparing various types of sensors. Depth information was crucial to the NAO Humanoid Robot for performing tasks in many circumstances in the most effective and efficient manner. The sensors compared were the Nao Humanoid Robot's own stereo camera, sonar sensors, and an Asus Xtion Pro Live RGB-D sensor. An external NDI Polaris Spectra was used to provide ground truth as a control set. The authors concluded that the Asus Xtion Pro Live had the best fit method to achieve distance information by comparing the relative errors among the sensors. The authors further justified that Asus Xtion Pro was easy to operate and only caused slight relative errors. Sonar was not as accurate as Asus Xtion Pro Live, which could easily be clarified from the paper. Finally, the distance measurement algorithm stereo vision itself was accurate, but too many error accumulations caused high relative error.

Maier et al.[18] proposed an integrated approach for applications such as robot localization, obstacle mapping and path planning in 3D environments based on data from an onboard consumer-level depth camera. The authors demonstrated their system by installed the Asus Xtion Pro Live on the NAO Humanoid Robot. The NAO Humanoid Robot was used to navigate a multi-level environment which contained static and non-static obstacles. The front-end processed the sensor data to extract visual features and associate them to 3D points. Then, the system registered pair of images and built a pose graph. Finally, a textured voxel occupancy map using the OctoMapping approach was generated. There were different methods to be applied to compute the motion in between two scenes taken which depended on the sensor used as well. The back-end of the system was used to deal with the inherent uncertainty introduced in the front-end stage. The authors claimed that their approach could deal with highly challenging scenarios as they introduced a beam-based EEM that allows the system to evaluate the quality of a frame-to-frame estimate so that highly inaccurate estimates could be rejected in the process.

As with other spatial sensors for robots, RGB-D sensors have their own merits and demerits. When compared with traditional optical cameras, an RGB-D camera always has the advantage of being able to recover 3D world structure and 2D images simultaneously and with the depth channel being largely independent of ambient lighting[25]. Furthermore, RGB-D perception is more reliable than the results of pure depth sensors[25]. On the other hand, RGB-D sensors were found to have various over-arching limitations[32]. They cannot be operated under high intensity sunlight, do not provide reliable range data for semitransparent or highly reflective surfaces, and have a limited effective range. This would necessitate alternative methods for such important tasks as collision detection in order to compensate for this issue with RGB-D sensors.

3. Conclusion

The recent addition of affordable commercial RGB-D sensors initially seems to add a fourth piece of the puzzle, another niche sensing modality to compete with the existing trio. However, based on the works we have reviewed in this paper, we conclude that RGB-D sensors like the Microsoft Kinect and Asus Xtion Pro Live may very well supercede existing sensors by combining the affordability and information density of visual sensors with accuracy approaching that of 3D-enabled laser range finders, at a cost not significantly higher than the better acoustic sensors currently available. We believe that, despite the current surge of interest in such sensors, there is still significant room for utilizing them in various current robotic projects as more researchers discover the unique capabilities afforded by RGB-D sensors.

The primary drawbacks of RGB-D sensors have already been covered elsewhere in this paper, and mainly involve visually-reflective or translucent materials as well as computational constraints. The first drawback is not unique to this modality of sensors, as it applies both to visual sensors as well as to the laser pulses of laser range finders. The second drawback is shared with stereo vision systems, but should prove less of a concern as computational power continues to increase while costs decrease due to the natural progress of computing technology.

In conclusion, we have examined the most common sensors currently in use for solving the SLAM problem. Weighing the capabilities and drawbacks of each modality, we propose that RGB-D sensors should generally prove to be advantageous in the majority of SLAM problems, both as stand-alone solutions and in addition to currently existing solutions.

Acknowledgement(s)

This research was supported by UTARRF(6200/Y34) grant provided by Universiti Tunku Abdul Rahman for the study on 'Multi-agent 3-Dimensional SLAM using Swarm Intelligence'.

References

1. Antoni Burguera, Gabriel Oliver, and Juan D Tardos. Robust scan matching localization using ultrasonic range finders. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1367–1372. IEEE, 2005.
2. Antoni Burguera, Yolanda González, and Gabriel Oliver. Sonar sensor models and their application to mobile robot localization. *Sensors*, 9 (12):10217–10243, 2009.
3. Oscar De Silva, George KI Mann, and Raymond G Gosine. Development of a relative localization scheme for ground-aerial multi-robot systems. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 870–875. IEEE, 2012.

4. Ethan Eade and Tom Drummond. Edge landmarks in monocular {SLAM}. *Image and Vision Computing*, 27(5):588 – 596, 2009. ISSN 0262-8856. doi: http://dx.doi.org/10.1016/j.imavis.2008.04.012. URL `http://www.sciencedirect.com/science/article/pii/S0262885608000978`. The 17th British Machine Vision Conference (BMVC 2006).
5. Felix Endres, Jürgen Hess, Jürgen Sturm, Daniel Cremers, and Wolfram Burgard. 3-d mapping with an rgb-d camera. 2014.
6. Arturo Gil, Óscar Reinoso, Mónica Ballesta, and Miguel Juliá. Multi-robot visual slam using a rao-blackwellized particle filter. *Robotics and Autonomous Systems*, 58(1):68–80, 2010.
7. Arturo Gil, scar Reinoso, Mnica Ballesta, and Miguel Juli. Multi-robot visual {SLAM} using a rao-blackwellized particle filter. *Robotics and Autonomous Systems*, 58(1):68 – 80, 2010. ISSN 0921-8890. doi: http://dx.doi.org/10.1016/j.robot.2009.07.026. URL `http://www.sciencedirect.com/science/article/pii/S0921889009001158`.
8. H Gonzalez-Jorge, B Riveiro, E Vazquez-Fernandez, J Martínez-Sánchez, and P Arias. Metrological evaluation of microsoft kinect and asus xtion sensors. *Measurement*, 46(6):1800–1806, 2013.
9. Jingwei Guan and MQ-H Meng. Study on distance measurement for nao humanoid robot. In *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pages 283–286. IEEE, 2012.
10. Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton. Enhanced computer vision with microsoft kinect sensor: A review. 2013.
11. Il-Kyun Jung and Simon Lacroix. High resolution terrain mapping using low attitude aerial stereo imagery. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 946–951. IEEE, 2003.
12. Michael Kaess and Frank Dellaert. Probabilistic structure matching for visual slam with a multi-camera rig. *Computer Vision and Image Understanding*, 114(2):286–296, 2010.
13. Kourosh Khoshelham and Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12 (2):1437–1454, 2012.
14. H. Lategahn, A. Geiger, and B. Kitt. Visual slam for autonomous ground vehicles. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1732–1737, May 2011. doi: 10.1109/ICRA.2011.5979711.
15. Thomas Lemaire, Cyrille Berger, Il-Kyun Jung, and Simon Lacroix. Vision-based slam: Stereo and monocular approaches. *International Journal of Computer Vision*, 74(3):343–364, 2007.
16. Xiuzhi Li, Wei Cui, and Songmin Jia. Range scan matching and particle filter based mobile robot slam. In *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*, pages 779–784. IEEE, 2010.
17. David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
18. Daniel Maier, Armin Hornung, and Maren Bennewitz. Real-time navigation in 3d environments based on depth camera data. In *Humanoid Robots (Humanoids), 2012 12th IEEE-RAS International Conference on*, pages 692–697. IEEE, 2012.
19. Michael Montemerlo, Sebastian Thrun, Daphne Koller, Ben Wegbreit, et al. Fastslam: A factored solution to the simultaneous localization and mapping problem. In *AAAI/IAAI*, pages 593–598, 2002.
20. F.A. Moreno, J.L. Blanco, and J. Gonzalez. Stereo vision specific models for particle filter-based {SLAM}. *Robotics and Autonomous Systems*, 57(9):955 – 970, 2009. ISSN 0921-8890. doi: http://dx.doi.org/10.1016/j.robot.2009.03.002. URL `http://www.sciencedirect.com/science/article/pii/S0921889009000438`.
21. Yasir Mohd Mustafah, Amelia Wong Azman, and Fajril Akbar. Indoor uav positioning using stereo vision sensor. *Procedia Engineering*, 41: 575–579, 2012.
22. Nagla, KS and Uddin, Moin and Singh, Dilbag. Improved occupancy grid mapping in specular environment. *Robotics and Autonomous Systems*, 60(10):1245–1252, 2012.
23. Viet Nguyen, Agostino Martinelli, Nicola Tomatis, and Roland Siegwart. A comparison of line extraction algorithms using 2d laser rangefinder for indoor mobile robotics. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1929–1934. IEEE, 2005.
24. Pedro Núñez, Ricardo Vázquez-Martín, and Antonio Bandera. Visual odometry based on structural matching of local invariant features using stereo camera sensor. *Sensors*, 11(7):7262–7284, 2011.
25. Xiaofeng Ren, Dieter Fox, and Kurt Konolige. Change their perception: Rgb-d for 3-d modeling and recognition. *Robotics & Automation Magazine, IEEE*, 20(4):49–59, 2013.
26. Venugopal Srinivasan and Ronald Lumia. A pseudo-interferometric laser range finder for robot applications. *Robotics and Automation, IEEE Transactions on*, 5(1):98–105, 1989.
27. Hartmut Surmann, Andreas Nüchter, and Joachim Hertzberg. An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, 45(3):181–198, 2003.
28. Taro Suzuki, Yoshiharu Amano, and Takumi Hashizume. Development of a sift based monocular ekf-slam algorithm for a small unmanned aerial vehicle. In *SICE Annual Conference (SICE), 2011 Proceedings of*, pages 1656–1659. IEEE, 2011.
29. Alex Teichman and Sebastian Thrun. Learning to segment and track in rgbd. In *Algorithmic Foundations of Robotics X*, pages 575–590. Springer, 2013.
30. Hong-jian Wang, Jing Wang, Le Yu, and Zhen-ye Liu. A new slam method based on svm-aekf for auv. In *OCEANS 2011*, pages 1–6. IEEE, 2011.
31. Oliver Wulf and Bernardo Wagner. Fast 3d scanning methods for laser measurement systems. In *International conference on control systems and computer science (CSCS14)*, pages 2–5, 2003.
32. Qian-Yi Zhou and Vladlen Koltun. Dense scene reconstruction with points of interest. *ACM Transactions on Graphics (TOG)*, 32(4):112, 2013.